



Universidad Nacional del Sur

TESIS DE DOCTOR EN CIENCIAS DE LA COMPUTACIÓN

*Toma de decisiones individuales y colectivas
en sistemas multi-agente: un enfoque argumentativo
con explicaciones interactivas.*

Martín Ezequiel Buron Brarda

BAHÍA BLANCA

ARGENTINA

2023



Universidad Nacional del Sur

TESIS DE DOCTOR EN CIENCIAS DE LA COMPUTACIÓN

*Toma de decisiones individuales y colectivas
en sistemas multi-agente: un enfoque argumentativo
con explicaciones interactivas.*

Martín Ezequiel Buron Brarda

BAHÍA BLANCA

ARGENTINA

2023

Prefacio

Esta Tesis es presentada como parte de los requisitos para optar al grado académico de Doctor en Ciencias de la Computación, de la Universidad Nacional del Sur, y no ha sido presentada previamente para la obtención de otro título en esta Universidad u otras. La misma contiene los resultados obtenidos en investigaciones llevadas a cabo en el Departamento de Ciencias de la Computación, durante el período comprendido entre el 13 de junio de 2017 y el 14 de junio de 2023, bajo la dirección del Dr. Alejandro J. García, Profesor Titular del Departamento de Ciencias e Ingeniería de la Computación, y del Dr. Luciano H. Tamargo, Profesor Adjunto del Departamento de Ciencias e Ingeniería de la Computación.

Martín Ezequiel Buron Brarda

`martin.buron@cs.uns.edu.ar`

DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

UNIVERSIDAD NACIONAL DEL SUR

Bahía Blanca, 14 de junio de 2023.



UNIVERSIDAD NACIONAL DEL SUR
Secretaría General de Posgrado y Educación Continua

La presente tesis ha sido aprobada el / /, mereciendo
la calificación de(.....)

Agradecimientos

Esta tesis no podría haber sido escrita sin la ayuda de muchas personas e instituciones que me brindaron su apoyo a lo largo de esta etapa. En primer lugar, quiero agradecer a mis directores Luciano Tamargo y Alejandro García por haberme dado el placer de trabajar con ellos. No solo fueron mis tutores en el aspecto académico, sino que además, en varias ocasiones me han compartido sus consejos de vida, lo cual valoro, y mucho. Quiero agradecer también a la Universidad Nacional del Sur, por haberme brindado la oportunidad de realizar mis estudios de grado y posgrado de manera totalmente gratuita. Asimismo, agradezco al Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) por haberme brindado el sustento económico necesario para el desarrollo de mi doctorado.

Gracias a todos los miembros de Departamento de Ciencias e Ingeniería de la Computación por haberme abierto las puertas, y brindarme un lugar de trabajo con un ambiente excelente. Fue por mucho tiempo mi segundo hogar, y voy a estar siempre orgulloso de haber formado parte de su equipo. En especial, quiero agradecer a mis compañeros de la sala becarios, con los que hemos compartido días enteros con risas, charlas, mates, asados, etc... Hicieron que todo este camino fuese mucho más fácil de transitar, y me llevo un montón de buenas amistades. ¡Que no se corte a pesar de la distancia!

No puedo dejar de agradecer a mi familia por haberme dado las herramientas para haber llegado hasta acá, y por siempre estar cada vez que los necesito. Y sin lugar a dudas, el más especial de todos mis agradecimientos se los lleva la persona más importante en mi vida, gracias Chu. Gracias por estar siempre, y de manera incondicional. Gracias por bancarme en mis enojos, y en mis alegrías. Espero seguir compartiendo mi vida con vos por mucho tiempo más. Te amo.

Martín (Brown)

Resumen

Esta tesis presenta un marco de trabajo para toma de decisiones individuales y colectivas en sistemas multi-agentes el cual hace uso de un sistema argumentativo. Este marco de trabajo permite modelar diferentes escenarios en los que un grupo de agentes debe tomar una decisión para resolver un problema en común, teniendo en cuenta múltiples criterios de evaluación.

Como parte de las contribuciones de esta tesis, se introduce un tipo especial de reglas llamadas reglas de preferencia rebatibles. Estas reglas permiten a los agentes expresar las razones que consideran apropiadas para preferir una alternativa por sobre otra. Debido a las fortalezas y debilidades de cada alternativa, estas reglas pueden generar preferencias contradictorias que deben resolverse antes de tomar una decisión. Para abordar este problema, el marco de trabajo propuesto utiliza un sistema argumentativo donde los argumentos se generan a partir de las reglas de preferencia rebatibles y la información sobre las alternativas. Las contradicciones entre las preferencias inferidas se resuelven a través de un proceso de evaluación de argumentos que determina su aceptación y, por lo tanto, la aceptación de las preferencias que respaldan. Este proceso de evaluación tiene en cuenta los intereses de cada agente y también considera la posibilidad de priorizar los intereses de ciertos agentes. En esta tesis, no solo se define la estructura del marco de trabajo, sino que también se estudian las propiedades formales que caracterizan el comportamiento del mismo. Además, el marco de trabajo propuesto puede ser aplicado en escenarios donde la información sobre las alternativas es incompleta, y también se adapta muy bien en escenarios donde la información es dinámica.

Si bien el uso de argumentación como mecanismo de razonamiento, facilita la posibilidad de brindar explicaciones, estas no siempre son fáciles de comprender. El marco de trabajo propuesto permite obtener explicaciones interactivas para que los usuarios puedan comprender el funcionamiento del formalismo sin sentirse abrumados por la información

presentada. Además, en esta tesis se describe y estudia el funcionamiento y rendimiento de una aplicación de escritorio desarrollada para implementar este formalismo, la cual permite visualizar las explicaciones e interactuar con ellas.

Abstract

This thesis presents a framework for individual and collective decision-making in multi-agent systems that utilizes an argumentation system. This framework allows for modeling different scenarios in which a group of agents must make a decision to solve a common problem, taking into account multiple evaluation criteria.

As part of the contributions of this thesis, a special type of rules called defeasible preference rules is introduced. These rules enable agents to express the reasons they consider appropriate for preferring one alternative over another. Due to the strengths and weaknesses of each alternative, these rules can lead to contradictory preferences that need to be resolved before making a decision. To address this problem, the proposed framework utilizes an argumentative system where arguments are generated from the defeasible preference rules and information about the alternatives. Contradictions between inferred preferences are then resolved through an argument evaluation process that determines their acceptance, and therefore, the acceptance of the preferences they support. This evaluation process takes into account the interests of each agent and also considers the possibility of prioritizing the interests of certain agents. In this thesis, not only is the structure of the framework defined, but the formal properties characterizing its behavior are also studied. Moreover, the proposed framework can be applied in scenarios where information about the alternatives is incomplete and it also adapts well to scenarios with dynamic information.

While the use of argumentation as a reasoning mechanism facilitates the possibility of providing explanations, they are not always easy to comprehend. The proposed framework allows for obtaining interactive explanations so that users can understand the workings of the formalism without feeling overwhelmed by the presented information. Additionally, in this thesis, the operation and performance of a desktop application developed to im-

plement this formalism are described and studied. This application enables visualization of the explanations and interaction with them.

Índice general

1. Introducción	1
1.1. Contribuciones	5
1.2. Publicaciones	6
1.3. Organización de la tesis	7
2. Generación de argumentos	9
2.1. Escenario de elección	10
2.2. Representación de preferencias	13
2.3. Proceso de generación de argumentos	23
2.4. Resumen	26
3. Evaluación de Argumentación	29
3.1. Conflicto y derrota	31
3.2. Proceso de evaluación dialéctico	37
3.3. Resultados formales	49
3.4. Resumen	58
4. Selección de alternativas y generación de explicaciones	61
4.1. Proceso de selección de alternativas	62
4.2. Resultados formales	69
4.3. Generación de explicaciones	72

4.4.	Trabajando sin un agente de máxima prioridad	80
4.5.	Trabajando con evidencia incompleta	84
4.6.	Caracterización del conjunto de argumentos ante cambios en la evidencia o en las <i>DePref</i> -rules	91
4.7.	Resumen	99
5.	Implementación del formalismo	101
5.1.	Definiendo el problema de decisión	103
5.2.	Explicaciones interactivas	108
5.3.	Resultados experimentales	117
5.4.	Resumen	122
6.	Comparación con otros formalismos	125
6.1.	Argumentación abstracta	125
6.2.	Sistemas argumentativos para toma decisiones individuales	128
6.2.1.	Marco de trabajo para la agregación de evidencia sobre tratamientos médicos [HW12]	129
6.2.2.	El uso de argumentación para explicar las mejores decisiones [ZFLT14]	136
6.2.3.	Marco para toma de decisiones basado en sistemas de argumen- tación dinámicos [FTG ⁺ 17]	145
6.3.	Sistemas argumentativos para toma de decisiones con múltiples agentes . .	153
6.3.1.	Toma de decisiones entre dos agentes mediante diálogos ABA [FTMW14]	153
6.3.2.	Toma de decisiones basada en argumentación entre múltiples agen- tes con información privada [GTWX16]	162
6.4.	Resumen	169
7.	Conclusiones	175

Capítulo 1

Introducción

De forma general, un agente se enfrenta ante un problema de toma de decisiones cada vez que debe escoger una opción de entre un conjunto de alternativas posibles. Para resolver este problema, el agente deberá determinar qué decisión tomar (o que alternativa elegir); y si se tratara de un agente racional, se esperaría que elija aquella alternativa que más se adecue a sus preferencias, metas, o intereses ([Woo00]). Más aún, existen diversos escenarios en los que la decisión debe ser tomada en conjunto por un grupo de agentes. Algunos escenarios de este estilo podrían ser: elegir una vivienda, elegir un candidato para un puesto de trabajo, o incluso, elegir un lugar para vacacionar. La complejidad de resolver estos problemas no sólo radica en la gran cantidad de información que debe ser evaluada, sino además, en que las fortalezas y debilidades de las diferentes alternativas no suelen ser las mismas ([YH95]). Por ejemplo, si una familia quisiera buscar un nuevo lugar en donde vivir, podrían comparar las alternativas por su precio, su tamaño y por el nivel de inseguridad del barrio en el que se encuentran. En este escenario, como en muchos otros, resultaría natural que los agentes traten de buscar la alternativa que presente el mejor balance con respecto a todos los criterios de evaluación, ya que sería muy extraño que exista una alternativa que sea mejor que todas las demás con respecto a todos los criterios.

Los enfoques de la teoría clásica de decisión ([Fis70, KR93]), pretenden medir la utilidad (o conveniencia) de cada alternativa mediante una función matemática, para luego determinar la mejor alternativa en base a su utilidad. Si bien es posible definir una función de utilidad capaz de balancear la influencia de cada criterio en la decisión, en la mayoría de los casos, definir tal función de forma correcta no suele ser una tarea simple,

o peor aún, muchas veces resulta impracticable. Otro problema de estos enfoques es que a pesar de poder hallar la función que mejor se adecue a las preferencias del agente, todo el proceso de evaluación de las alternativas queda codificado dentro de la definición de esta función, la cual podría ser extremadamente compleja, por lo que el modelo resultaría difícil de interpretar y/o entender. Por otro lado, existen enfoques declarativos [dSCLS94, BBHP99, BD02, BLW10, VAHJ12] (en [DHKP11, Kac11, PTV16] se puede encontrar un excelente resumen de varios de estos enfoques) que simplifican la tarea de la representación de las preferencias, como así también la interpretación del modelo. El problema con algunos de estos modelos es que, por su simplicidad, se vuelve difícil (y a veces imposible) establecer preferencias que tengan en cuenta las fortalezas de las alternativas para compensar sus debilidades, es decir, preferencias con las que se pueda balancear la importancia que tiene cada criterio en la decisión. Dicho problema puede observarse, por ejemplo, al tratar de definir las preferencias por medio de un orden lexicográfico entre los criterios. Retomando el ejemplo de la vivienda, si se define que el precio es más importante que el tamaño y que la seguridad, entonces, sin lugar a dudas la vivienda más económica va a ser la preferida, sin importar ninguno de los otros criterios. Aunque según el modelo, esto sería lo correcto, es probable que las preferencias de muchos agentes no puedan ser representadas adecuadamente por dicha aproximación. Más allá de todo esto, cabe mencionar, que tanto los modelos clásicos que usan funciones de utilidad como los enfoques declarativos citados anteriormente, están pensados para representar las preferencias de un solo agente, pero no resuelven (o al menos no de manera trivial) cómo integrar las preferencias de varios agentes.

En los últimos años, la argumentación ha demostrado ser una herramienta de especial utilidad para el tratamiento computacional del proceso de toma de decisiones, ya sea considerando un solo agente ([BG96, KM03, ABP05, ABCM06, FECS08, Amg09, CMG09, ODLM11, HW12, VHJ12, VHJ13, BBD⁺14, FECS14, TGGS15, Wak17, ZFM⁺18, ZFLT19, CO19, RCBT20, TGGS20]), o varios ([LM11, DG14, FTMW14, GTWX16, BCHS18, AGG19, BCKK19]). Esto se debe a que la capacidad de los modelos argumentativos para trabajar con información contradictoria, los hace propicios para contrastar las fortalezas y debilidades de cada alternativa. Como será analizado en esta tesis, una gran parte de los formalismos para toma de decisiones basados en argumentos que han sido propuestos hasta el momento, no permiten definir preferencias de manera relativamente sencilla o intuitiva y que a la vez logren establecer algún tipo de balance entre fortalezas y debilidades. Adicionalmente, el uso de formalismos argumentativos, por

lo general, facilita la comprensión de los resultados obtenidos (vea [CRA⁺21, EDP⁺22] para más información acerca del uso de argumentación como herramienta para la generación de explicaciones). Esto se debe a que su funcionamiento se basa en la exposición de las razones a favor y en contra asociadas a cada decisión, lo que suele resultar intuitivo para los seres humanos ([MTB⁺13]). Esta característica es de suma importancia ya que es más que claro que los sistemas de recomendación o asistencia durante el proceso de toma de decisiones, y más aún, aquellos sistemas que son diseñados para funcionar de forma autónoma pueden afectar las vidas de las personas. De hecho, en algunos países las explicaciones se han vuelto un requerimiento. Por ejemplo, el Reglamento General de Protección de Datos (o GDPR, por sus siglas en inglés), que es aplicable en toda la Unión Europea desde mayo del 2018, establece en su Considerando 71 que los ciudadanos tendrán derecho a solicitar una explicación de cualquier decisión que haya resultado de un proceso automatizado ([Eur16, GF17]).

En esta tesis, se presenta un marco de trabajo para toma de decisiones individuales y colectivas en sistemas multi-agentes que hace uso de un sistema argumentativo. Esta propuesta, permite modelar escenarios como los que fueron mencionados anteriormente. Como parte de las contribuciones de esta tesis, se presenta un tipo especial de reglas, denominadas *reglas de preferencia rebatibles* (*DePref*-rules por su abreviación en inglés) que permiten a los agentes expresar las condiciones por las que consideran adecuado que una alternativa sea preferida a otra. Por la forma en la que están definidas, estas reglas brindan la posibilidad de definir hasta que punto los agentes están dispuestos a tolerar que una alternativa sea peor que otra en cierto/s criterio/s, sabiendo cuanto mejor es con respecto a otro/s criterio/s. De esta manera, los agentes pueden balancear el peso que tiene cada criterio en la decisión. Asimismo, el formalismo aquí presentado, considera la posibilidad de que los agentes puedan no compartir los mismos intereses. Más aún, como suele suceder en muchos escenarios, este formalismo también permite establecer prioridades entre los agentes, y por ende, entre sus intereses, siendo posible modelar esquemas jerárquicos entre los agentes.

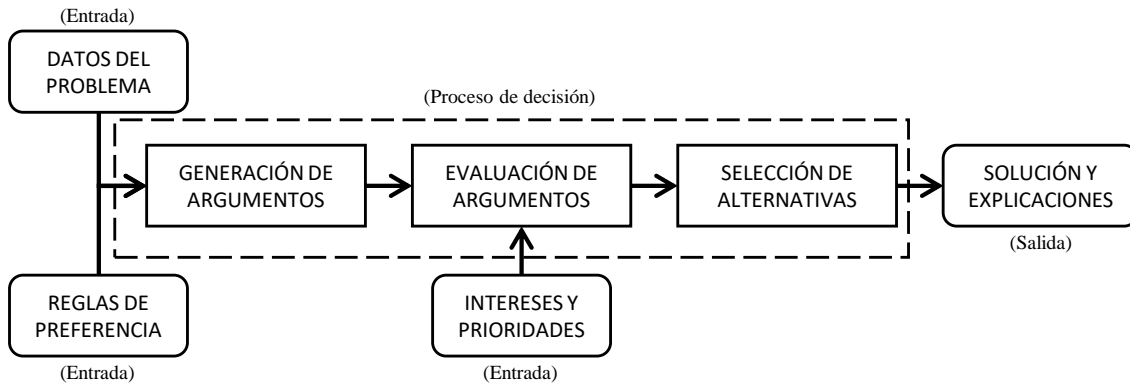


Figura 1.1: Etapas del proceso de decisión.

La Figura 1.1 muestra las distintas etapas que componen el proceso de decisión que se ejecuta dentro del marco de trabajo que se presenta en esta tesis, y el flujo de datos entre cada una de estas etapas. En la etapa de generación de argumentos se reciben los datos del problema (compuestos básicamente de las alternativas y su descripción) junto con las reglas de preferencia de los agentes. A partir de estos elementos, se generan los argumentos para comparar las alternativas utilizando las reglas de preferencia (es decir, las *DePref*-rules). Luego, estos argumentos ingresan en la etapa de evaluación de argumentos, donde se mide y compara la fuerza de los argumentos en base a los intereses de los agentes y el orden de prioridad entre los mismos. A raíz de esto, se determina la relación de derrota entre los argumentos, y finalmente se evalúa la aceptabilidad de los mismos. A continuación, las conclusiones de los argumentos aceptados, es decir, las preferencias que estos sustentan, son enviadas a la etapa de selección de alternativas. En esta etapa se analizan las conclusiones obtenidas, y a partir de ellas, se genera un orden de preferencia entre las alternativas. Teniendo en cuenta este orden, se seleccionan las mejores alternativas, lo que se brinda como solución al problema de decisión. Junto con la solución, el proceso de decisión retornará una explicación en forma de grafo, mediante la cual los usuarios podrán interactuar e indagar acerca del proceso de razonamiento que fue llevado a cabo.

En lo que resta de este capítulo, se detallarán las principales contribuciones de esta tesis, las publicaciones que surgieron de ella, y finalmente se detallará la organización de la misma.

1.1. Contribuciones

A continuación, se resumen las principales contribuciones de esta tesis:

Marco de trabajo para toma de decisiones grupales basado en argumentos

Se presenta un marco de trabajo para toma de decisiones grupales basado en argumentos. A partir de este formalismo es posible modelar diferentes escenarios en donde un grupo compuesto por uno o varios agentes debe tomar una decisión con el fin de darle solución a un problema en común. En esta tesis, no solo se define la estructura del marco de trabajo sino que también se estudian las propiedades formales que caracterizan el comportamiento del mismo. Además, se define formalmente la manera en la que el marco de trabajo brinda explicaciones acerca de sus resultados. Esto tiene particular importancia ya que es poco común que las explicaciones sean consideradas parte de los formalismos. Por último, se estudia formalmente como optimizar el proceso de generación de argumentos ante la presencia de cambios en la información que se tiene de las alternativas, o cambios en el conjunto de reglas de preferencia.

Reglas de preferencia rebatibles

Se presenta la sintaxis y semántica de las reglas de preferencia rebatibles (*DePref*-rules). Estas reglas establecen un esquema de comparación entre las alternativas que permite inferir preferencias entre ellas considerando uno o varios criterios a la vez. Las *DePref*-rules brindan la posibilidad de definir hasta que punto los agentes están dispuestos a tolerar que una alternativa sea peor que otra en cierto/s criterio/s, sabiendo cuanto mejor es con respecto a otro/s criterio/s. De esta manera, los agentes pueden balancear el peso que tiene cada criterio en la decisión. En particular, en este trabajo se presenta una nueva versión de las *DePref*-rules que busca sacar provecho de las versiones presentadas en trabajos anteriores ([BBTG19, BBTG20]), obteniendo lo mejor de cada una de ellas. Finalmente, también caracterizan las preferencias inferidas a partir de las *DePref*-rules.

Implementación del marco de trabajo

Se ha desarrollado una aplicación de escritorio que implementa el marco de trabajo presentado en esta tesis. En este documento, se describe su funcionamiento y un análisis de rendimiento de la misma. Esta aplicación no es sólo una prueba de conceptos, sino que también habilita a otras personas a utilizar el marco de trabajo, ya sea con propósitos académicos o personales. Más aún, el desarrollo de esta aplicación permite sacar más provecho de las explicaciones que brinda el sistema mediante una interfaz gráfica amigable.

1.2. Publicaciones

Parte de los resultados obtenidos durante el desarrollo de esta tesis fueron publicados en revistas de la especialidad. A continuación, se detallan dichos trabajos:

1. En el trabajo “An approach to enhance argument-based multi-criteria decision systems with conditional preferences and explainable answers” [BBTG19] publicado en la revista *Expert Systems with Application* vol. 126 (2019) se definió un marco de trabajo para toma de decisiones individuales basado en argumentos, una primera versión de las reglas de preferencia rebatibles, y se introdujo la noción de grafo de explicaciones y una interfaz gráfica prototípica con la que se detallan algunas posibles interacciones con este grafo. Además, se presentó un estudio empírico que demuestra la factibilidad de implementación y uso del formalismo. Este artículo establece las bases de lo desarrollado en los Capítulos 2, 3, 4 y parte del Capítulo 5

2. En el trabajo “Using argumentation to obtain and explain results in a decision support system” [BBTG20] publicado en la revista *IEEE Intelligent Systems* vol. 36 (2020) se describió una aplicación de escritorio que implementa el marco de trabajo presentado en el artículo anterior. En dicho artículo se explican los pasos que los usuarios del sistema deben efectuar para modelar un problema de decisión y obtener una solución, y luego, se detallan las distintas interacciones que los usuarios pueden realizar para obtener explicaciones acerca del proceso que se llevó a cabo para brindar dicha solución. Todo esto es desarrollado en el Capítulo 5, aunque para el caso de este documento, la aplicación descrita corresponde a la última versión de la aplicación que permite modelar problemas de decisión multi-agente.

3. En el trabajo “Multi-agent multi-criteria decision framework with interactive explanation tools” [BBTG23] publicado en la revista *Expert Systems with Application* vol. 228 (2023) se presentó la versión multi-agente del marco de trabajo presentado en los artículos anteriores, el cual será explicado en detalle en los Capítulos 2, 3 y 4. En particular, se incorpora formalmente la noción de distancia para las premisas de las *DePref*-rules, lo que será explicado en el Capítulo 2. Además, se introducen los Marcos de Evaluación de Preferencias Multi-agente (o MPEF, por sus siglas en inglés) que permiten evaluar los argumentos generados a partir de un conjunto de *DePref*-rules, teniendo en cuenta los intereses de varios agentes y la posibilidad de establecer prioridades sobre los mismos. Este formalismo, junto con todas las pruebas formales que verifican el correcto funcionamiento

del sistema argumentativo fueron incorporadas al Capítulo 3. En dicho artículo, también se estudió el comportamiento de la propuesta para trabajar con conjuntos de evidencia incompletos y en entornos dinámicos, lo cual está detallado en las Secciones 4.5 y 4.6. Por último, se presentaron algunas mejoras de la interfaz gráfica de la aplicación frente a lo presentado en [BBTG20]. Como se mencionó anteriormente, en el Capítulo 5 se muestran todos los detalles de esta aplicación.

1.3. Organización de la tesis

El resto de esta tesis se encuentra organizado de la siguiente manera:

Capítulo 2. En este capítulo se comienza a describir el marco de trabajo. En particular, se describe el formato de los elementos de entrada (alternativas, reglas de preferencia, intereses de los agentes y prioridades) y cómo estos deben ser interpretados para generar argumentos (ver Figura 1.1).

Capítulo 3. En este capítulo se describen los procesos que se ejecutan dentro de la etapa de evaluación de argumentos (ver Figura 1.1). Principalmente, se explica cómo se comparan los argumentos que soportan preferencias contradictorias, para determinar las derrotas entre los mismos, y finalmente, decidir cuáles de estos deben ser aceptados.

Capítulo 4. En este capítulo se describen los procesos que se ejecutan dentro de la etapa de selección de alternativas (ver Figura 1.1). Se explica cómo se obtiene la solución al problema de decisión, es decir, cómo se determinan las mejores alternativas, y se describen formalmente las explicaciones que justifican los resultados. Como apartados de este capítulo, se estudia el comportamiento del marco de trabajo en escenarios en los que la información acerca de las alternativas es incompleta, y además, se caracterizan los cambios que sufre el conjunto de argumentos ante cambios en la evidencia o en el conjunto de *DePref*-rules.

Capítulo 5. En este capítulo se describe una aplicación de escritorio que implementa el marco de trabajo presentado en los capítulos anteriores. Para finalizar el capítulo, se presenta un análisis empírico del rendimiento de la aplicación.

Capítulo 6. En este capítulo se resumen algunos trabajos relacionados en lo que respecta al área de toma de decisiones individuales y grupales basados en argumentación, y luego se los compara con la propuesta presentada en esta tesis.

Capítulo 7. En este capítulo se detallan los resultados obtenidos en el presente trabajo y se presentan algunas conclusiones. Finalmente, se plantean nuevas líneas de investigación para el futuro.

Capítulo 2

Generación de argumentos

En el capítulo anterior, se describió de manera general la propuesta de esta tesis, explicando a grandes rasgos las distintas etapas que involucra. En particular, la Figura 1.1 enseñaba un esquema en el que aparecen estas etapas. En este capítulo, se explicará en profundidad la primera de estas etapas.

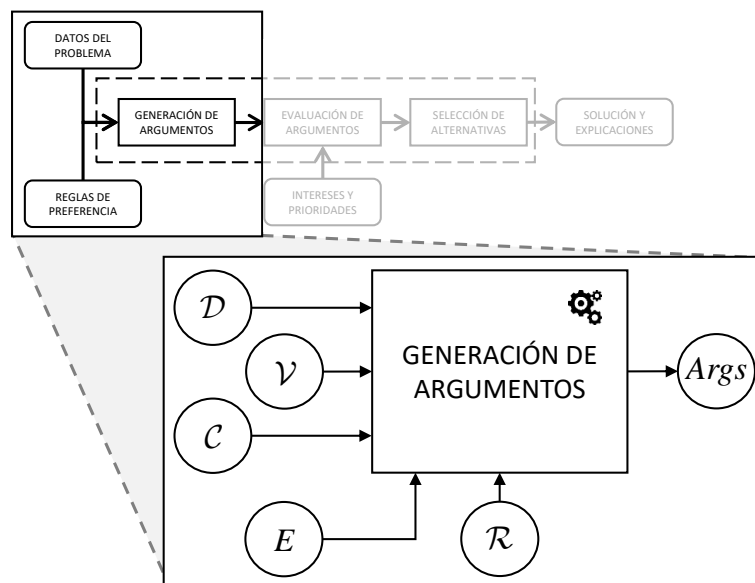


Figura 2.1: Detalle de la etapa de generación de argumentos.

La Figura 2.1 muestra en detalle las entradas y salidas de esta etapa. Como se mostrará en detalle a continuación, las entradas \mathcal{D} , \mathcal{C} , \mathcal{V} y \mathcal{E} componen el escenario de elección,

mediante el cual se caracterizan las alternativas, mientras que la entrada \mathcal{R} contiene las reglas de preferencia rebatibles con las que los agentes representan sus preferencias. Todos estos elementos serán utilizados durante el proceso de generación de argumentos del cual se obtendrán los argumentos de preferencia que serán utilizados para comparar las alternativas, representados por la salida $Args$.

2.1. Escenario de elección

El escenario de elección contendrá los datos involucrados en el problema de decisión que se quiere modelar. Este incluye: el conjunto de alternativas, los criterios mediante los cuales serán comparadas estas alternativas, los rangos de valores con los que pueden ser evaluados los criterios, y las valoraciones o características de las alternativas.

Definición 2.1.1 (Escenario de elección). *Un escenario de elección Π es una tupla $\langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$, donde:*

- \mathcal{D} es un conjunto finito de constantes que representan a las alternativas.
- \mathcal{C} es un conjunto finito de constantes que representan a los criterios por los que se comparan a las alternativas.
- \mathcal{V} es una función que a cada criterio $c \in \mathcal{C}$ le asigna un rango de valores totalmente ordenados con respecto a una relación “es mejor que”, denotada como $>_c$.
- E es un conjunto finito de predicados totalmente instanciados (sin variables) de la forma $c(d, v)$, llamados **piezas de evidencia**, donde $c \in \mathcal{C}$, $d \in \mathcal{D}$ y $v \in \mathcal{V}(c)$. Una pieza de evidencia $c(d, v)$ representa que, según el criterio c , la alternativa d es valorada v . Se asume que todo conjunto de evidencia E es **consistente**, es decir, que no existen dos piezas de evidencia $c(d, v)$ y $c(d, u)$ en E para la misma alternativa y con distinta valoración ($v \neq u$).

Como puede observarse en la Definición 2.1.1 los conjuntos \mathcal{D} y \mathcal{C} , junto con los valores de los rangos de la imagen de \mathcal{V} , establecen el vocabulario que se utiliza para definir el conjunto de evidencia E .

A continuación, se introduce un ejemplo de un escenario de elección. El mismo será utilizado a lo largo de este documento con el objetivo de ejemplificar distintos aspectos del marco de trabajo aquí propuesto.

Ejemplo 2.1.1. *La familia Jota está compuesta por cuatro miembros: Mamá, Papá, Carolina y Tomás. Para celebrar los 15 años de Carolina, los Jota planean ir de vacaciones, y para esto, deben elegir una, entre 5 posibles alternativas. Las alternativas son: ir a un pueblo entre medio de las montañas, ir a una moderna ciudad en la playa, ir a un pequeño pueblo en la costa de un lago, ir a un complejo turístico en el medio de la jungla, o tomar un crucero por el caribe. Los Jota concuerdan en que para comparar las alternativas utilizarán los siguientes criterios: la cantidad de días de paseo, el nivel de entretenimiento y la calidad del servicio contratado. Además, los miembros de la familia, tienen un presupuesto limitado y el precio por día de cada alternativa es diferente, sólo podrán pagar 14 días en las montañas, 11 días en la playa, 20 días en el lago, 14 días en la jungla, o bien, 10 días en crucero. Como se muestra en la Tabla 2.1, ellos también han acordado las valoraciones de las alternativas con respecto al nivel de entretenimiento y a la calidad del servicio.*

Alternativas	Criterios y valoraciones		
	Días	Entretenimiento	Servicio
montañas	14	regular	pésimo
playa	11	bueno	excelente
lago	20	malo	bueno
jungla	14	regular	excelente
crucero	10	bueno	bueno

Tabla 2.1: Valoraciones de las alternativas que tienen los Jota para irse de vacaciones.

De este modo, el escenario de elección para los Jota queda definido como $\langle \mathcal{D}_J, \mathcal{C}_J, \mathcal{V}_J, E_J \rangle$, donde $\mathcal{D}_J = \{mntñs, playa, lago, jngla, crcro\}$ ¹ representa las alternativas y $\mathcal{C}_J = \{días, entmto, srvcio\}$ representa los criterios de comparación. En cuanto a la función \mathcal{V}_J , $\mathcal{V}_J(días)$ es un rango de enteros de 1 a 30 ordenados de mayor a menor (cuantos más días de estadía, mejor), $\mathcal{V}_J(entmto) = \mathcal{V}_J(srvcio)$ siendo el rango de valores [pésimo, malo, regular, bueno, excelente] ordenados de izquierda a derecha de peor a mejor

¹Como puede observarse, se usaron algunas abreviaturas para representar las constantes y predicados del escenario de elección

valor. Siguiendo lo acordado por la familia Jota en la Tabla 2.1, el conjunto de evidencia queda entonces definido de la siguiente manera:

$$E_J = \left\{ \begin{array}{lll} \text{días}(\text{mntñs}, 14), & \text{entmto}(\text{mntñs}, \text{regular}), & \text{srvcio}(\text{mntñs}, \text{pésimo}), \\ \text{días}(\text{playa}, 11), & \text{entmto}(\text{playa}, \text{bueno}), & \text{srvcio}(\text{playa}, \text{excelente}), \\ \text{días}(\text{lago}, 20), & \text{entmto}(\text{lago}, \text{malo}), & \text{srvcio}(\text{lago}, \text{bueno}), \\ \text{días}(\text{jngla}, 14), & \text{entmto}(\text{jngla}, \text{regular}), & \text{srvcio}(\text{jngla}, \text{excelente}), \\ \text{días}(\text{crcro}, 10), & \text{entmto}(\text{crcro}, \text{bueno}), & \text{srvcio}(\text{crcro}, \text{bueno}) \end{array} \right\}$$

Como será explicado en la próxima sección, conociendo los criterios que definen el escenario de elección, los agentes tomadores de decisiones podrán expresar sus preferencias de forma general, usando un tipo especial de reglas denominadas *reglas de preferencia rebatibles*. Utilizando luego esas reglas, los agentes serán capaces de especificar las condiciones que deben cumplirse para preferir una alternativa por sobre otra.

Retomando el Ejemplo 2.1.1, Carolina podría querer expresar que entre dos alternativas, en general, prefiere la que provea mayor entretenimiento, aunque en ciertas ocasiones, ella prefiere tener más días de vacaciones a tener más entretenimiento. Una de estas ocasiones es la siguiente: si entre dos alternativas X e Y , si por una mínima diferencia, X es peor que Y con respecto al entretenimiento, pero X ofrece al menos 3 días más de paseo que Y , entonces X sería preferida a Y . En otras palabras, para Carolina sería preferible una alternativa menos entretenida, si la pérdida de entretenimiento no es muy grande, siempre y cuando exista una buena ganancia con respecto a los días de paseo. De esta forma, Carolina preferiría ir al jungla en lugar de ir a la playa ya que la diferencia de entretenimiento es mínima, pero al ir a la jungla tendría 3 días más que si fueran a la playa. Sin embargo, esta regla no sería aplicable al tratar de comparar la playa con el lago porque a pesar de existir una diferencia de 9 días a favor del lago, la pérdida de entretenimiento no sería tolerable.

Del ejemplo anterior, se puede notar que la diferencia entre valoraciones de alternativas será un elemento de importancia para esta propuesta. La siguiente definición establece cómo calcular la diferencia que existe entre dos valoraciones, y será utilizada para definir las reglas de preferencia rebatibles.

Definición 2.1.2 (Distancia). *Sea $c \in \mathcal{C}$ un criterio, u y v dos posibles valoraciones pertenecientes a $\mathcal{V}(c)$. La distancia entre u y v es:*

$$\mathbf{dist}_c(u, v) = \begin{cases} |u - v| & \text{si } \mathcal{V}(c) \subseteq \mathbb{R}, \\ |\mathbf{ord}_c(u) - \mathbf{ord}_c(v)| & \text{en caso contrario.} \end{cases}$$

donde $\mathbf{ord}_c : \mathcal{V}(c) \rightarrow \mathbb{N}$ es la función que retorna el ordinal de una valoración con respecto al orden de $\mathcal{V}(c)$ establecido por $>_c$.

Como puede verse, la función $\mathbf{dist}_c(u, v)$ indicará cuan diferente son las valoraciones de dos alternativas (con respecto a un criterio c), retornando la diferencia absoluta que existe entre u y v en caso de que estos pertenezcan a un sub-rango de número reales, como es el caso de $\mathcal{V}_{\text{días}}$, o bien, la diferencia absoluta entre el orden de estos valores en caso de que estos pertenezcan a un rango de valores simbólicos y discretos, como es el caso de $\mathcal{V}_{\text{entmto}}$ o $\mathcal{V}_{\text{srvicio}}$. Por ejemplo, considerando los criterios del Ejemplo 2.1.1 y los valores de las alternativas *playa* y *jnsla*, se tiene que $\mathbf{dist}_{\text{días}}(11, 14) = |11 - 14| = 3$; $\mathbf{dist}_{\text{entmto}}(\text{bueno}, \text{regular}) = |\mathbf{ord}_{\text{entmto}}(\text{bueno}) - \mathbf{ord}_{\text{entmto}}(\text{regular})| = |4 - 3| = 1$; y que $\mathbf{dist}_{\text{srvicio}}(\text{excelente}, \text{excelente}) = |\mathbf{ord}_{\text{srvicio}}(\text{excelente}) - \mathbf{ord}_{\text{srvicio}}(\text{excelente})| = |5 - 5| = 0$. Esto significa que la distancia en días de paseo entre ir a la jungla e ir a la playa es 3, la distancia en entretenimiento es 1 y la distancia en la calidad del servicio es 0.

2.2. Representación de preferencias

Como se dijo anteriormente, los agentes podrán expresar bajo qué condiciones prefieren una alternativa por sobre otra utilizando reglas de preferencia rebatibles (o *DePref*-rules por su abreviación en inglés). Mediante estas reglas, los agentes podrán definir los distintos criterios que deberán ser tomados en cuenta para comparar las alternativas, así como también, la influencia que tendrá cada criterio en la comparación. Los agentes podrán balancear dicha influencia siendo capaces de establecer hasta cuánto están dispuestos a perder en algunos criterios, por obtener una ganancia con respecto a otros criterios.

Definición 2.2.1 (Sintaxis de las *DePref*-rules). *Sea $\langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$ un escenario de elección, sean X e Y dos variables que representan dos alternativas en \mathcal{D} . Una *DePref*-rule es una regla esquemática de la forma “cuerpo $\Rightarrow \text{pref}(X, Y)$ ” que indica en su cuerpo las condiciones que deben satisfacerse sobre las valoraciones de las alternativas, para tener una razón a favor de que X sea preferida a Y , y debe ser especificada a partir de la siguiente gramática:*

$$\langle \text{regla} \rangle ::= \{ \langle \text{cuerpo} \rangle \} \Rightarrow \text{pref}(X, Y)$$

$$\langle \text{cuerpo} \rangle ::= \langle \text{premisa} \rangle \mid \langle \text{cuerpo} \rangle , \langle \text{premisa} \rangle$$

$$\langle \text{premisa} \rangle ::= \langle \text{b_premisa} \rangle \mid \langle \text{e_premisa} \rangle \mid \langle \text{w_premisa} \rangle$$

$$\langle \text{b_premisa} \rangle ::= \text{better}(X, Y, c) \ [, \text{min_dist}(X, Y, c, \rho)] \ [, \text{min}(X, c, v)] \ [, \text{max}(Y, c, u)]$$

$$\langle \text{e_premisa} \rangle ::= \text{equal}(X, Y, c) \ [, \text{min}(X, c, v) \mid , \text{max}(Y, c, u)]$$

$$\langle \text{w_premisa} \rangle ::= \text{worse}(X, Y, c) \ [, \text{max_dist}(X, Y, c, \rho)] \ [, \text{min}(X, c, v)] \ [, \text{max}(Y, c, u)]$$

donde $c \in \mathcal{C}$, $\rho \in \mathbb{R}^+$, y $\{u, v\} \subseteq \mathcal{V}(c)$. Además, toda *DePref-rule* debe ser **coherente**, esto es, debe haber sido construida a partir de al menos una **b_premisa** y no pueden existir dos premisas que evalúen el mismo criterio.

El significado intuitivo de $\text{better}(X, Y, c)$ es que la alternativa X es mejor que la alternativa Y con respecto al criterio c . Análogamente, el de $\text{equal}(X, Y, c)$ y el $\text{worse}(X, Y, c)$ es que X es igual (resp. peor) que Y con respecto a c . Por otro lado, el significado intuitivo de $\text{min_dist}(X, Y, c, \rho)$ es que la diferencia entre las valoraciones de X y Y con respecto al criterio c es al menos tan grande como ρ , mientras que el de $\text{max_dist}(X, Y, c, \rho)$ es que esta diferencia no es superior ρ . Finalmente, el significado intuitivo de $\text{min}(X, c, v)$ y de $\text{max}(Y, c, u)$ es que el valor de X con respecto c es al menos tan buena como v y que el valor de Y es a lo sumo tan bueno como u , respectivamente.

La Definición 2.2.1, introduce la sintaxis de las *DePref-rules*. La semántica será introducida más adelante en la definición 2.2.3, antes de de ello se explicarán algunos detalles y se mostrarán algunas *DePref-rules* a modo de ejemplo. Como se puede observar de la gramática, toda *DePref-rule* debe tener al menos una premisa, aunque también puede tener más. Note que dentro de una misma premisa, c hace referencia a un único criterio, pero dos premisas diferentes no pueden compartir el mismo criterio. Dentro del cuerpo de una *DePref-rule* se pueden encontrar tres tipos de premisas: **b_premisa**, **e_premisa** y **w_premisa**. Además, de los predicados específicos de cada tipo de premisa, es decir, *better*, *equal* y *worse* (respectivamente), éstas pueden incluir condiciones sobre las valoraciones de las alternativas usando *min_dist*, *max_dist*, *min* y *max*. En cuanto a *min_dist* y *max_dist*, éstas permiten establecer la mínima y máxima distancia (ver Definición 2.1.2) que debe existir entre las valoraciones de X e Y con respecto al criterio que está siendo evaluado por la premisa a la que pertenecen. Cabe mencionar que *min_dist* sólo puede pertenecer a una **b_premisa**, mientras que *max_dist* sólo puede pertenecer a una **w_premisa**.

Esto se debe a que carece de sentido exigir una separación máxima entre los valores de X e Y sabiendo que X es mejor que Y (cuánto más separación exista mejor es para X), ni tampoco exigir una separación mínima sabiendo que X es peor que Y (cuánto menos separación exista mejor es para X). Las otras dos condiciones (*min* y *max*) permiten expresar los valores mínimos y máximos que tanto X como Y deben respetar, respectivamente. A continuación, mostraremos un ejemplo en el que pueden observarse distintas *DePref*-rules, y más adelante, se explicará cómo evaluar dichas reglas.

Ejemplo 2.2.1. *Considere nuevamente el escenario de elección $\langle \mathcal{D}_J, \mathcal{C}_J, \mathcal{V}_J, E_J \rangle$ del Ejemplo 2.1.1. A continuación, se mostrarán algunas *DePref*-rules que pueden ser construidas teniendo en cuenta los elementos de este escenario.*

$$\left\{ \text{better}(X, Y, \text{entmto}) \right\} \Rightarrow \text{pref}(X, Y)$$

$$\left\{ \begin{array}{l} \text{worse}(X, Y, \text{entmto}), \text{max_dist}(X, Y, \text{entmto}, 1), \\ \text{better}(X, Y, \text{días}), \text{min_dist}(X, Y, \text{días}, 3) \end{array} \right\} \Rightarrow \text{pref}(X, Y)$$

$$\left\{ \begin{array}{l} \text{worse}(X, Y, \text{días}), \text{min}(X, \text{días}, 10), \text{max_dist}(X, Y, \text{días}, 5), \\ \text{equal}(X, Y, \text{srcio}), \text{min}(X, \text{srcio}, \text{bueno}), \\ \text{better}(X, Y, \text{entmto}), \text{min}(X, \text{entmto}, \text{regular}), \text{min_dist}(X, Y, \text{entmto}, 2) \end{array} \right\} \Rightarrow \text{pref}(X, Y)$$

$$\left\{ \begin{array}{l} \text{better}(X, Y, \text{días}), \text{min_dist}(X, Y, \text{días}, 5), \\ \text{better}(X, Y, \text{srcio}), \text{min}(X, \text{srcio}, \text{excelente}), \\ \text{worse}(X, Y, \text{entmto}), \text{min}(X, \text{entmto}, \text{malo}), \text{max}(Y, \text{entmto}, \text{bueno}) \end{array} \right\} \Rightarrow \text{pref}(X, Y)$$

La primera regla puede leerse así: “Si X es mejor que Y en cuanto al entretenimiento, entonces existe una razón para preferir a X sobre Y ”.

Por otro lado, la segunda regla puede leerse así: “A pesar de que X sea peor que Y en cuanto al entretenimiento, mientras la distancia entre las valoraciones con respecto a este criterio no sea mayor a 1, si X es mejor que Y en cuanto a los días de paseo y X ofrece al menos 3 días más que Y , entonces existe una razón para preferir a X sobre Y ”.

La tercer regla puede leerse como “A pesar de que X ofrezca menos días de paseo que Y , mientras que X ofrezca al menos 10 días y la diferencia de días no sea mayor a 5, si ambas alternativas ofrecen un servicio de igual calidad que al menos sea bueno, y además, X ofrece un mejor entretenimiento que Y siendo como mínimo regular y con al menos una distancia de 2 en lo que respecta a este criterio, entonces existe una razón para preferir a X sobre Y ”.

Por último, la cuarta regla se puede leer de la siguiente forma: “Si X ofrece al menos 5 días más de paseo que los que ofrece Y , y el servicio que ofrece X es excelente y mejor que el de Y , a pesar de que Y ofrezca un mejor entretenimiento, mientras que el entretenimiento de X no sea pésimo (es decir, sea al menos malo) y mientras que el de Y no sea excelente (es decir, sea a lo sumo bueno), entonces existe una razón para preferir a X sobre Y ”.

Siguiendo el estilo de lo que es usual en área de la programación lógica y argumentación estructurada (ver por ejemplo, [Lif96, GDS09, Pra10, Ton14]), las *DePref*-rules definen reglas de preferencia esquemáticas en términos de dos variables X e Y . Por lo tanto, para poder establecer una preferencia concreta, es necesario instanciar estas variables con las alternativas que pertenecen al escenario del elección del problema que se está modelando.

Definición 2.2.2 (*DePref*-rule instanciada). Sea $\langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$ un escenario de elección, con $d_x, d_y \in \mathcal{D}$, y sea r una *DePref*-rule. La versión de r instanciada con d_x y d_y , denotada $r(d_x, d_y)$, se obtiene reemplazando en r cada aparición de X por d_x y cada aparición de Y por d_y .

Ejemplo 2.2.2. Considere las *DePref*-rules del Ejemplo 2.2.1, y sea r_i la segunda de las reglas allí mostradas. Es decir,

$$r_i = \left\{ \begin{array}{l} \text{worse}(X, Y, \text{entmto}), \text{max_dist}(X, Y, \text{entmto}, 1), \\ \text{better}(X, Y, \text{días}), \text{min_dist}(X, Y, \text{días}, 3) \end{array} \right\} \Rightarrow \text{pref}(X, Y)$$

Al instanciar esta regla con las alternativas $\{\text{playa}, \text{jngla}\} \subset \mathcal{D}_J$, se obtiene

$$r_i(\text{playa}, \text{jngla}) = \left\{ \begin{array}{l} \text{worse}(\text{playa}, \text{jngla}, \text{entmto}), \text{max_dist}(\text{playa}, \text{jngla}, \text{entmto}, 1), \\ \text{better}(\text{playa}, \text{jngla}, \text{días}), \text{min_dist}(\text{playa}, \text{jngla}, \text{días}, 3) \end{array} \right\} \Rightarrow \text{pref}(\text{playa}, \text{jngla})$$

Note que al invertir el orden de las alternativas se obtiene una instancia distinta de la misma regla.

$$r_i(\text{jngla}, \text{playa}) = \left\{ \begin{array}{l} \text{worse}(\text{jngla}, \text{playa}, \text{entmto}), \text{max_dist}(\text{jngla}, \text{playa}, \text{entmto}, 1), \\ \text{better}(\text{jngla}, \text{playa}, \text{días}), \text{min_dist}(\text{jngla}, \text{playa}, \text{días}, 3) \end{array} \right\} \Rightarrow \text{pref}(\text{jngla}, \text{playa})$$

Una *DePref*-rule instanciada especifica las condiciones que permiten establecer una preferencia entre dos alternativas particulares. Para determinar cuáles de estas condiciones

se cumplen, es necesario analizar la evidencia del escenario de elección. La Definición 2.2.3 formaliza este concepto, al establecer la semántica de los predicados que pueden aparecer en una *DePref*-rule.

Definición 2.2.3 (Semántica de las *DePref*-rules). *Sea $\langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$ un escenario de elección, $d_x, d_y \in \mathcal{D}$ dos alternativas, $c \in \mathcal{C}$ un criterio, v y u dos valoraciones posibles pertenecientes a $\mathcal{V}(c)$, $\rho \in [0, +\infty)$ y dist la función de la Definición 2.1.2.*

- *E satisface $\text{better}(d_x, d_y, c)$ si y sólo si $\{c(d_x, u), c(d_y, v)\} \subseteq E$ y $u >_c v$,*
- *E satisface $\text{equal}(d_x, d_y, c)$ si y sólo si $\{c(d_x, u), c(d_y, v)\} \subseteq E$ y $u =_c v$,*
- *E satisface $\text{worse}(d_x, d_y, c)$ si y sólo si $\{c(d_x, u), c(d_y, v)\} \subseteq E$ y $v >_c u$,*
- *E satisface $\text{min_dist}(d_x, d_y, c, \rho)$ si y sólo si $\{c(d_x, u), c(d_y, v)\} \subseteq E$ y $\text{dist}_c(u, v) \geq \rho$,*
- *E satisface $\text{max_dist}(d_x, d_y, c, \rho)$ si y sólo si $\{c(d_x, u), c(d_y, v)\} \subseteq E$ y $\text{dist}_c(u, v) \leq \rho$,*
- *E satisface $\text{min}(d_x, c, v)$ si y sólo si $\{c(d_x, u)\} \subseteq E$ y $u \geq_c v$*
- *E satisface $\text{max}(d_x, c, v)$ si y sólo si $\{c(d_x, u)\} \subseteq E$ y $v \geq_c u$*

*Luego, se dice que un conjunto de piezas de evidencia E satisface una *DePref*-rule instanciada $r(d_x, d_y)$ si y sólo si E satisface cada uno de los predicados en el cuerpo de $r(d_x, d_y)$.*

Ejemplo 2.2.3. *Considere las *DePref*-rules instanciadas $r_i(\text{playa}, \text{jngla})$ y $r_i(\text{jngla}, \text{playa})$ del Ejemplo 2.2.2. Como estas reglas evalúan a las alternativas $\{\text{playa}, \text{jngla}\} \subset \mathcal{D}_J$, para ver si E_J satisface a estas reglas basta con analizar sólo las piezas de evidencia correspondientes a las mismas, las cuales serían:*

$$\left\{ \begin{array}{l} \text{días}(\text{playa}, 11), \quad \text{entmto}(\text{playa}, \text{bueno}), \quad \text{srvcio}(\text{playa}, \text{excelente}), \\ \text{días}(\text{jngla}, 14), \quad \text{entmto}(\text{jngla}, \text{regular}), \quad \text{srvcio}(\text{jngla}, \text{excelente}) \end{array} \right\} \subset E_J$$

Ahora se puede ver, uno a uno, si E_J satisface los predicados de $r_i(\text{playa}, \text{jngla})$ y $r_i(\text{jngla}, \text{playa})$. Comenzando por $r_i(\text{playa}, \text{jngla})$, se tiene que E_J no satisface a $\text{worse}(\text{playa}, \text{jngla}, \text{entmto})$ ya que las vacaciones en la playa ofrecen un nivel de entretenimiento mejor que el de las vacaciones en la jungla ($\{\text{entmto}(\text{playa}, \text{bueno}), \text{entmto}(\text{jngla}, \text{regular})\} \subset E_J$ y $\text{bueno} >_{\text{entmto}} \text{regular}$). Por lo tanto, E_J no satisface a $r_i(\text{playa}, \text{jngla})$.

Por otro lado, en cuanto a $r_i(jngla, playa)$, $worse(jngla, playa, entmto)$ sí es satisfecho por E_J ya que como se mencionó anteriormente, las vacaciones en la jungla ofrecen un nivel de entretenimiento peor que las vacaciones en la playa. Además, también se cumple que la diferencia de entretenimiento entre $jngla$ y $playa$ es igual a 1, por lo que $\mathbf{dist}_{entmto}(jngla, playa) \leq 1$. Esto significa que E_J satisface a $\mathbf{max_dist}(jngla, playa, entmto, 1)$. Observe también, que $jngla$ es mejor que $playa$ en cuanto a los días de paseo ya que las vacaciones en la jungla duran 14 días mientras que las vacaciones en la playa duran 11, por lo que, $jngla$ es mejor que $playa$ por 3 días ($\{días(playa, 11), días(jngla, 14)\} \subset E_J$, $14 >_{días} 11$, y $\mathbf{dist}_{días}(jngla, playa) \geq 3$). Por lo tanto, E_J satisface a $\mathbf{better}(jngla, playa, días)$ y a $\mathbf{min_dist}(jngla, playa, días, 3)$. Por lo tanto, E_J satisface a $r_i(jngla, playa)$.

Como puede verse en el ejemplo anterior, en esta propuesta, las preferencias que puedan inferirse usando *DePref*-rules, dependerán tanto de que exista una regla que concluya ese preferencia, como de la evidencia que se tenga del escenario de elección. Las *DePref*-rules son las encargadas de representar las razones generales por las que un agente especifica sus preferencias, y solamente las *DePref*-rules instanciadas que sean satisfechas por la evidencia disponible serán las que se usarán para inferir las preferencias para el escenario concreto en el cual hay que tomar la decisión. A partir de un mismo conjunto de evidencia, se pueden inferir conclusiones contradictorias si se tienen dos *DePref*-rules que sean satisfechas por la evidencia y sus conclusiones sean opuestas.

Definición 2.2.4 (Opuesto). *El opuesto de una conclusión $\alpha = \mathbf{pref}(d_x, d_y)$ es $\bar{\alpha} = \mathbf{pref}(d_y, d_x)$.*

Note que los opuestos son simétricos. Esto es, $\mathbf{pref}(d_x, d_y)$ es el opuesto de $\mathbf{pref}(d_y, d_x)$, si y sólo si $\mathbf{pref}(d_y, d_x)$ es el opuesto de $\mathbf{pref}(d_x, d_y)$. Por lo tanto, también es válido decir que $\mathbf{pref}(d_x, d_y)$ y $\mathbf{pref}(d_y, d_x)$ son opuestos.

En el ejemplo anterior se mostró que a partir de r_i se puede inferir que existe una razón para preferir las vacaciones en la jungla sobre las vacaciones en la playa ya que E_J satisface a $r_i(jngla, playa)$, cuya conclusión es $\mathbf{pref}(jngla, playa)$. No obstante, a partir del mismo conjunto de evidencia E_J , es posible inferir lo opuesto, es decir, $\mathbf{pref}(playa, jngla)$, ya que la *DePref*-rule

$$r_j = \left\{ \mathbf{equal}(X, Y, srvcio), \mathbf{better}(X, Y, entmto) \right\} \Rightarrow \mathbf{pref}(X, Y)$$

establece que hay una razón para preferir a una alternativa X sobre otra alternativa Y , siempre y cuando ambas ofrezcan un servicio de igual calidad, y X ofrezca un mejor nivel de entretenimiento que Y . Luego, como *playa* y *jngla* (ambas) ofrecen servicios excelentes ($\{srvcio(playa, excelente), srvcio(jngla, excelente)\} \subset E_J$), y como *playa* ofrece mejor nivel de entretenimiento que *jngla* ($\{entmto(playa, bueno), entmto(jngla, regular)\} \subset E_J$), se tiene que E_J satisface a $r_j(playa, jngla)$, cuya conclusión es $pref(playa, jngla)$. Note que,

$$r_j(playa, jngla) = \left\{ equal(playa, jngla, srvcio), better(playa, jngla, entmto) \right\} \Rightarrow pref(playa, jngla)$$

Por lo tanto, se infiere que existe una razón para preferir las vacaciones en la playa sobre las vacaciones en la jungla. Claramente, esto contradice la preferencia inferida a partir de $r_i(jngla, playa)$.

No obstante, dada la asimetría que existe entre los predicados *better* y *worse* ya que una alternativa no puede ser mejor, y a la vez, peor que otra con respecto a un mismo criterio (recuerde que la evidencia debe ser consistente, ver Definición 2.1.1), y dado que toda *DePref*-rule debe estar construida a partir de al menos una **b**-premisa para ser coherente (ver Definición 2.2.1), no es posible inferir a partir de una misma *DePref*-rule preferencias que sean contradictorias. Note lo que sucede, por ejemplo, con

$$r_j(jngla, playa) = \left\{ equal(jngla, playa, srvcio), better(jngla, playa, entmto) \right\} \Rightarrow pref(jngla, playa)$$

Como ya se dijo, *playa* ofrece un mejor entretenimiento que *jngla*, y por ende, *jngla* no ofrece un mejor entretenimiento que *playa*. Es por eso, que $r_j(jngla, playa)$ no es satisfecha por E_J , y por ende, no es posible inferir (a partir de r_j) que existe una razón para preferir las vacaciones en la jungla sobre las vacaciones en la playa.

Estando formalmente definida la semántica de las *DePref*-rules, es posible caracterizar las preferencias que pueden ser inferidas a partir de ellas, independientemente del conjunto de evidencia que conforma el escenario elección. La siguiente proposición muestra formalmente que dado un conjunto de evidencia, a partir de una misma *DePref*-rule no es posible inferir dos conclusiones opuestas. Es decir, si la evidencia satisface una regla r instanciada como $r(d_x, d_y)$, entonces no es posible que E satisfaga a $r(d_y, d_x)$.

Proposición 2.2.1. *Sea $\langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$ un escenario de elección tal que $d_x, d_y \in \mathcal{D}$, y sea r una *DePref*-rule. Si E satisface a $r(d_x, d_y)$, entonces E no satisface a $r(d_y, d_x)$.*

Demostración. Sea r una *DePref*-rule, y $\langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$ un escenario de elección tal que E satisface a $r(d_y, d_x)$, donde $d_x, d_y \in \mathcal{D}$. Como r es coherente (ver Definición 2.2.1), se puede asegurar que r tiene al menos una **b_premisa** P . Además, como E satisface a $r(d_x, d_y)$, debe satisfacer a los predicados de P . Siendo que P es una **b_premisa**, y por ende, tiene un predicado de la forma $better(d_x, d_y, c)$ con $c \in \mathcal{C}$. De acuerdo con la Definición 2.2.3, para que E satisfaga a ese predicado, E debe contener al conjunto de evidencia $\{c(d_x, u), c(d_y, v)\}$, donde $u >_c v$. Por el absurdo, suponga que E también satisface a $r(d_y, d_x)$. Entonces E debe satisfacer la **b_premisa** P' de $r(d_y, d_x)$ que evalúa el criterio c , y tiene un predicado de la forma $better(d_y, d_x, c)$. Sin embargo, para esto E debe contener al menos una pieza de evidencia $c(Z, w)$ en la que $Z = d_x$ y $v >_c w$, o bien, $Z = d_y$ y $w >_c u$. En ambos casos, E no sería consistente (ver Definición 2.1.1) ya que para una misma alternativa existirían dos piezas de evidencia diferentes con el mismo criterio. Absurdo que provino de suponer que E satisface a $r(d_y, d_x)$. \square

Además, como consecuencia de la restricción de coherencia de las *DePref*-rules, es posible asegurar que al satisfacerse una *DePref*-rule, realmente existe una razón para preferir una alternativa X por sobre otra Y ya que, para al menos uno de los criterios, X debe ser mejor que Y . Si esto no se cumple, es decir, si Y es mejor que X con respecto a todos los criterios en los que X tiene una valoración, entonces sería imposible inferir a partir de una *DePref*-rule que existe una razón para preferir a X sobre Y .

Proposición 2.2.2. *Sea $\langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$ un escenario de elección, y sean $d_x, d_y \in \mathcal{D}$ dos alternativas. Si para todo $c(d_x, v) \in E$, existe $c(d_y, u) \in E$ tal que $u \geq_c v$, no existe una *DePref*-rule r , tal que E satisface $r(d_x, d_y)$.*

Demostración. Sea $\langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$ un escenario de elección, y sean $d_x, d_y \in \mathcal{D}$ dos alternativas tales que para todo $c(d_x, v) \in E$, existe $c(d_y, u) \in E$ tal que $u \geq_c v$. Por el absurdo, suponga que existe una *DePref*-rule r tal que E satisface a $r(d_x, d_y)$. Entonces, según la Definición 2.2.3, E debe satisfacer cada uno de los predicados que aparecen en el cuerpo de $r(d_x, d_y)$. Como r debe ser coherente (ver Definición 2.2.1), $r(d_x, d_y)$ debe haber sido construida a partir de una **b_premisa**, por lo que debe contener un predicado de la forma $better(d_x, d_y, c)$ con $c \in \mathcal{C}$. Por lo tanto, como E satisface a dicho predicado, $\{c(d_x, v), c(d_y, u)\} \subseteq E$, con $v >_c u$ (ver Definición 2.2.3). Sin embargo, esto no es posible ya que por hipótesis $u \geq_c v$. Absurdo que provino de suponer que existía r tal que E satisface a $r(d_y, d_x)$. \square

En consecuencia, si las valoraciones de una alternativa con respecto a cada uno de los criterios en los que está evaluada, son iguales a las valoraciones de otra alternativa con respecto a esos mismos criterios, entonces no será posible inferir una preferencia entre estas dos.

Corolario 2.2.1. *Sea $\langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$ un escenario de elección, y sean $d_x, d_y \in \mathcal{D}$ dos alternativas. Si para todo $c(d_x, v) \in E$, existe $c(d_y, u) \in E$ tal que $v =_c u$, entonces no existe una *DePref-rule* r , tal que E satisfice a $r(d_x, d_y)$, ni tampoco satisfice a $r(d_y, d_x)$.*

Demostración. Trivial por la Proposición 2.2.2 ya que para todo $c(d_x, v) \in E$ existe $c(d_y, u) \in E$ con $v =_c u$, y por lo tanto, $v \geq_c u$. \square

Como caso especial de esta propiedad indicada por el Corolario 2.2.1, se tiene que es imposible satisfacer una *DePref-rule* r cuyas variables fueron instanciadas con la misma alternativa d_x , es decir, es imposible satisfacer la regla $r(d_x, d_x)$.

Corolario 2.2.2. *Sea $\langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$ un escenario de elección, y sea $d_x \in \mathcal{D}$ una alternativa. No existe una *DePref-rule* r , tal que E satisfaga a $r(d_x, d_x)$.*

Demostración. Trivial por el Corolario 2.2.1. \square

Para completar el ejemplo de la familia Jota, cuyo escenario de elección fue introducido en el Ejemplo 2.1.1, a continuación, se presentarán las *DePref-rules* que representan las preferencias de los integrantes de esta familia. Para el formalismo que está siendo presentado en esta tesis, se asumirá que todos los agentes involucrados en el problema de decisión aceptarán las *DePref-rules* de los demás. Note que esta consideración no implica una restricción al formalismo, ya que los agentes deben comunicar sus preferencias por medio de las *DePref-rules*. No aceptarlas, implicaría no aceptar las preferencias de los demás. La idea en la que se basa esta propuesta es que cada agente indicará sus preferencias mediante un conjunto de *DePref-rules*, pero como se quiere tomar una decisión grupal, cada agente debe aceptar las preferencias de los demás.

Observación 2.2.1. *Se asume que cada agente acepta las *DePref-rules* de los demás.*

Ejemplo 2.2.4. *Siguiendo con el Ejemplo 2.1.1, los cuatro integrantes de la familia Jota están de acuerdo en que el hecho de que una alternativa sea mejor que otra con respecto a*

cualquiera de los criterios (*días*, *entmtoo srvcio*), es suficiente para tener una razón para preferirla. Esto es representado por las siguientes DePref-rules.

$$r_1 = \left\{ \text{better}(X, Y, \text{días}) \right\} \Rightarrow \text{pref}(X, Y)$$

$$r_2 = \left\{ \text{better}(X, Y, \text{entmtto}) \right\} \Rightarrow \text{pref}(X, Y)$$

$$r_3 = \left\{ \text{better}(X, Y, \text{srvcio}) \right\} \Rightarrow \text{pref}(X, Y)$$

Por su parte, Carolina, considera que aunque el entretenimiento es importante, bajo algunas condiciones especiales, elegiría una alternativa que es peor en cuanto al entretenimiento. En particular, cuando la diferencia de entretenimiento no sea mayor a 1, si la alternativa menos entretenida ofrece al menos 3 días más de paseo que la otra (r_4), o si la calidad del servicio es al menos 2 puntos mejor (r_5), entonces la alternativa menos entretenida, sería la preferida. Esto es representado de la siguiente manera:

$$r_4 = \left\{ \begin{array}{l} \text{worse}(X, Y, \text{entmtto}), \text{max_dist}(X, Y, \text{entmtto}, 1), \\ \text{better}(X, Y, \text{días}), \text{min_dist}(X, Y, \text{días}, 3) \end{array} \right\} \Rightarrow \text{pref}(X, Y)$$

$$r_5 = \left\{ \begin{array}{l} \text{worse}(X, Y, \text{entmtto}), \text{max_dist}(X, Y, \text{entmtto}, 1), \\ \text{better}(X, Y, \text{srvcio}), \text{min_dist}(X, Y, \text{srvcio}, 2) \end{array} \right\} \Rightarrow \text{pref}(X, Y)$$

También siendo más específica, la mamá de Carolina, piensa que entre dos alternativas que ofrecen la misma calidad de servicio, la que sea más entretenida debería ser preferida (r_6). Sin embargo, su padre, piensa que en dicho caso la alternativa que debería ser preferida es aquella que ofrezca más de días de paseo (r_7).

$$r_6 = \left\{ \text{equal}(X, Y, \text{srvcio}), \text{better}(X, Y, \text{entmtto}) \right\} \Rightarrow \text{pref}(X, Y)$$

$$r_7 = \left\{ \text{equal}(X, Y, \text{srvcio}), \text{better}(X, Y, \text{días}) \right\} \Rightarrow \text{pref}(X, Y)$$

Por último, Tomás elegiría una alternativa con menos días de paseo solo cuando pueda asegurarse de tener al menos 14 días, un mejor servicio y un mejor entretenimiento, que además, sea al menos regular. Esto es representado por la siguiente DePref-rule:

$$r_8 = \left\{ \begin{array}{l} \text{worse}(X, Y, \text{días}), \text{min}(X, \text{días}, 14), \\ \text{better}(X, Y, \text{srvcio}), \\ \text{better}(X, Y, \text{entmtto}), \text{min}(X, \text{entmtto}, \text{regular}) \end{array} \right\} \Rightarrow \text{pref}(X, Y)$$

A pesar de que las preferencias de los Jota no coinciden totalmente, para tomar una decisión grupal, cada uno de ellos debe aceptar las preferencias de los demás. Por lo tanto, el conjunto de DePref-rules que representa las preferencias de los Jota queda definido como $\mathcal{R}_J = \{r_1, r_2, r_3, r_4, r_5, r_6, r_7, r_8\}$.

2.3. Proceso de generación de argumentos

Las *DePref*-rules permiten a los agentes expresar sus razones para preferir una alternativa por sobre otra. Por tal motivo, las *DePref*-rules pueden ser utilizadas como esquemas de razonamiento para generar argumentos. Las conclusiones de estos argumentos darán lugar a una relación de preferencia entre las alternativas que será utilizada para establecer cuáles de estas serán seleccionadas.

Definición 2.3.1 (Argumento de preferencia). *Un argumento de preferencia construido a partir de un escenario de elección $\langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$ y una *DePref*-rule r , es una tupla $\mathcal{A} = \langle r(d_x, d_y), \text{pref}(d_x, d_y) \rangle$ tal que E satisface a $r(d_x, d_y)$ y $\text{pref}(d_x, d_y)$ es la conclusión de $r(d_x, d_y)$.*

A continuación, se muestra un algoritmo que esquematiza el proceso de generación de argumentos a partir de un escenario de elección y un conjunto de *DePref*-rules (ver Figura 2.1). Los argumentos generados (*Args*) dependerán tanto del escenario de elección ($\langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$), como de las *DePref*-rules de los agentes (\mathcal{R}).

Algoritmo 1 Proceso de generación de argumentos

Entrada: Un escenario de elección $\langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$ y un conjunto de *DePref*-rules \mathcal{R} .

Salida: Un conjunto de argumentos de preferencia *Args*.

```

Args  $\leftarrow \emptyset$ 
for each  $r \in \mathcal{R}$  do
  for each  $\langle d_x, d_y \rangle \in \mathcal{D} \times \mathcal{D}$  tal que  $d_x \neq d_y$  do
    if  $E$  satisface a  $r(d_x, d_y)$  then
       $\mathcal{A} \leftarrow \langle r(d_x, d_y), \text{pref}(d_x, d_y) \rangle$ 
       $\text{Args} \leftarrow \text{Args} \cup \{\mathcal{A}\}$ 
    end if
  end for
end for
return Args

```

El Algoritmo 1, muestra como se construye el conjunto de argumentos *Args* instanciando cada una de las *DePref*-rules \mathcal{R} con todos los distintos pares de alternativas en $\mathcal{D} \times \mathcal{D}$. Por cada *DePref*-rule instanciada de esta manera, se verifica si E la satisface, y

si ese es el caso, se genera un nuevo argumento de preferencia. Cabe mencionar, que este algoritmo podría optimizarse haciendo uso de la Proposición 2.2.1 que indica que si una *DePref*-rule instanciada $r(d_x, d_y)$ es satisfecha por un conjunto de evidencia E , entonces E no satisface a $r(d_y, d_x)$. Hacer uso de esta propiedad, en el mejor de los casos, podría reducir a la mitad los chequeos de satisfacción de las reglas. Sin embargo, en el peor de los casos, no ofrece ningún beneficio, siendo que el hecho de que $r(d_x, d_y)$ no sea satisfecha por E , no indica nada con respecto a la satisfacción de $r(d_y, d_x)$, por lo que no puede saltarse este chequeo. Es por tal motivo, que no se aprovechó esta propiedad para optimizar el proceso de generación de argumentos. Note que, esto no es así en lo que respecta a la propiedad del Corolario 2.2.2, que indica que una *DePref*-rule no puede satisfacerse si se instancia con dos alternativas iguales. Es por esto, que en el algoritmo se considera que $d_x \neq d_y$.

Ejemplo 2.3.1. *A continuación, se muestra el conjunto de argumentos de preferencia $Args_J$ que pueden ser construidos a partir del escenario de elección $\langle \mathcal{D}_J, \mathcal{C}_J, \mathcal{V}_J, E_J \rangle$ del Ejemplo 2.1.1 y el conjunto de DePref-rules \mathcal{R}_J del Ejemplo 2.2.4.*

$$\begin{aligned}
 Args_J = \{ & \\
 & \langle r_1(jngla, playa), pref(jngla, playa) \rangle, & \langle r_2(playa, jngla), pref(playa, jngla) \rangle, \\
 & \langle r_4(jngla, playa), pref(jngla, playa) \rangle, & \langle r_6(playa, jngla), pref(playa, jngla) \rangle, \\
 & \langle r_7(jngla, playa), pref(jngla, playa) \rangle, & \langle r_1(jngla, crcro), pref(jngla, crcro) \rangle, \\
 & \langle r_2(crcro, jngla), pref(crcro, jngla) \rangle, & \langle r_3(jngla, crcro), pref(jngla, crcro) \rangle, \\
 & \langle r_4(jngla, crcro), pref(jngla, crcro) \rangle, & \langle r_1(lago, jngla), pref(lago, jngla) \rangle, \\
 & \langle r_2(jngla, lago), pref(jngla, lago) \rangle, & \langle r_3(jngla, lago), pref(jngla, lago) \rangle, \\
 & \langle r_4(jngla, lago), pref(jngla, lago) \rangle, & \langle r_8(jngla, lago), pref(jngla, lago) \rangle, \\
 & \langle r_3(jngla, mntñs), pref(jngla, mntñs) \rangle, & \langle r_1(playa, crcro), pref(playa, crcro) \rangle, \\
 & \langle r_3(playa, crcro), pref(playa, crcro) \rangle, & \langle r_1(lago, playa), pref(lago, playa) \rangle, \\
 & \langle r_2(playa, lago), pref(playa, lago) \rangle, & \langle r_3(playa, lago), pref(playa, lago) \rangle, \\
 & \langle r_1(mntñs, playa), pref(mntñs, playa) \rangle, & \langle r_2(playa, mntñs), pref(playa, mntñs) \rangle, \\
 & \langle r_3(playa, mntñs), pref(playa, mntñs) \rangle, & \langle r_4(mntñs, playa), pref(mntñs, playa) \rangle, \\
 & \langle r_1(lago, crcro), pref(lago, crcro) \rangle, & \langle r_2(crcro, lago), pref(crcro, lago) \rangle, \\
 & \langle r_6(crcro, lago), pref(crcro, lago) \rangle, & \langle r_7(lago, crcro), pref(lago, crcro) \rangle, \\
 & \langle r_1(mntñs, crcro), pref(mntñs, crcro) \rangle, & \langle r_2(crcro, mntñs), pref(crcro, mntñs) \rangle, \\
 & \langle r_3(crcro, mntñs), pref(crcro, mntñs) \rangle, & \langle r_4(mntñs, crcro), pref(mntñs, crcro) \rangle, \\
 & \langle r_1(lago, mntñs), pref(lago, mntñs) \rangle, & \langle r_2(mntñs, lago), pref(mntñs, lago) \rangle, \\
 & \langle r_3(lago, mntñs), pref(lago, mntñs) \rangle, & \langle r_4(lago, mntñs), pref(lago, mntñs) \rangle, \\
 & \langle r_5(lago, mntñs), pref(lago, mntñs) \rangle \\
 & \}
 \end{aligned}$$

En el Ejemplo 2.2.3, ya se podía observar que a partir de distintas DePref-rules era posible inferir preferencias opuestas. Esto queda en evidencia en el Ejemplo 2.3.1. A partir de las reglas r_1 , r_4 y r_7 , se infiere $pref(jngla, playa)$, mientras que a partir de las reglas r_2 y r_6 se infiere $pref(playa, jngla)$. Es por este motivo, que en nuestra propuesta se pueden generar argumentos de preferencia que sustentan conclusiones contradictorias. En pos de determinar aquellas preferencias que deben prevalecer, la aceptabilidad de estos

argumentos debe ser evaluada; tarea que será llevada a cabo durante la etapa de evaluación de argumentos. Esto será explicado en detalle en el capítulo siguiente.

2.4. Resumen

En este capítulo se explicó lo que sucede durante la etapa de generación de argumentos. En primer lugar, se dio a conocer la manera en la que serán representadas las alternativas y sus características, lo que se logra al especificar el escenario de elección. El escenario de elección, permite a los agentes del sistema establecer cuáles van a ser las alternativas disponibles (\mathcal{D}), mediante qué criterios van a ser evaluadas estas alternativas (\mathcal{C}), cuáles son los posibles valores que pueden ser utilizados para valorar una alternativa con respecto a un dado criterio (\mathcal{V}), y cuál es la evidencia (E) que se tiene con respecto a las alternativas, es decir, cuáles son las valoraciones que tiene cada alternativa asociadas a los distintos criterios.

Luego, se presentó la sintaxis y semántica de las reglas de preferencia rebatibles (o *DePref*-rules), mediante las cuales los agentes pueden representar sus preferencias. Una *DePref*-rule es una regla esquemática en término de dos variables X e Y , que indica las condiciones que deben satisfacerse sobre las valoraciones de dos alternativas cualquiera (X e Y) para tener una razón para creer que una de ellas es preferida a la otra. Algo que hace a las *DePref*-rules diferentes a otros enfoques de representación de preferencias, es su capacidad para indicar de forma explícita cómo se valoran las virtudes y las falencias de las alternativa. Estas reglas pueden ser utilizadas para representar preferencias muy simples como: “Si X es mejor que Y en cuanto al entretenimiento, entonces existe una razón para preferir a X sobre Y ”; o también más complejas como: “Si X ofrece al menos 5 días más de paseo que los que ofrece Y , y el servicio que ofrece X es excelente y mejor que el de Y , a pesar de que Y ofrezca un mejor entretenimiento, mientras que el entretenimiento de X no sea pésimo (es decir, sea al menos malo) y mientras que el de Y no sea excelente (es decir, sea a lo sumo bueno), entonces existe una razón para preferir a X sobre Y ”.

Al instanciar las *DePref*-rules con alternativas concretas, es posible inferir una preferencia entre dos alternativas específicas, siempre y cuando, la evidencia determine que las condiciones establecidas en el cuerpo de las *DePref*-rules se cumplen. Esto da lugar a que las *DePref*-rules puedan ser usadas como esquemas de razonamiento para generar argumentos que sustenten las preferencias inferidas. Las preferencias inferidas a partir

de un conjunto de *DePref*-rules pueden ser contradictorias, por lo que, contar con un modelo argumentativo para resolver estas contradicciones es una buena forma para determinar cuáles de estas preferencias realmente son válidas. En línea con esto, se presentó el Algoritmo 1, que implementa el proceso de generación de argumentos, y mediante el cual es posible obtener todos los argumentos de preferencia que pueden ser construidos a partir de un escenario de elección y un conjunto de *DePref*-rules. Como será explicado en el siguiente capítulo, estos argumentos serán evaluados durante la segunda etapa para determinar cuáles son las preferencias válidas, es decir, aquellas que están garantizadas por los resultados del proceso de evaluación de argumentos.

Adicionalmente, en este capítulo, se presentaron algunas propiedades formales que permiten caracterizar las preferencias que pueden ser inferidas a partir las *DePref*-rules, independientemente de cuál sea el escenario de elección involucrado. Además, la Proposición 2.2.2 será utilizada más adelante para la prueba de la Proposición 4.2.4, la cual permite caracterizar el conjunto de alternativas seleccionadas.

Capítulo 3

Evaluación de Argumentación

Como se mostró en el capítulo anterior, al considerar más de una *DePref*-rule es muy probable que se puedan construir argumentos que sustenten preferencias contradictorias. El objetivo de la etapa de evaluación de argumentos es determinar cuáles de estos deben ser aceptados, y de esta forma, obtener un conjunto de conclusiones libre de contradicciones.

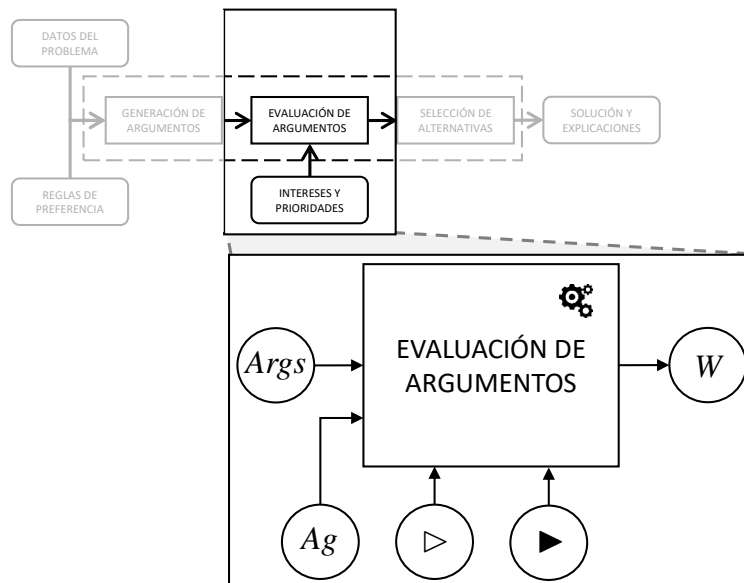


Figura 3.1: Detalle de la etapa de evaluación de argumentos. Esta etapa recibe el conjunto de argumentos ($Args$) de la etapa anterior, el conjunto de agentes (Ag), una lista de ordenes parciales (\triangleright) con la que se representan los intereses de los agentes, y un orden parcial (\blacktriangleright) que representa la prioridad entre los agentes.

La Figura 3.1 muestra en detalle las entradas y salidas de la etapa de evaluación de argumentos. El primer paso de esta etapa, es identificar los conflictos que surgen a raíz de las contradicciones entre los argumentos de preferencia (*Args*). Dado que los argumentos fueron construidos a partir de las *DePref*-rules que modelan las preferencias de los agentes, para esta propuesta se ha considerado que la fortaleza de los argumentos dependa de los intereses del grupo de agentes (*Ag*). Siguiendo este razonamiento, se incorpora una lista de ordenes parciales (\triangleright) para representar los intereses individuales de cada agente, y otro orden parcial (\blacktriangleright) para representar la prioridad que se le debe dar a los intereses de cada uno de ellos. Así, durante esta etapa será posible comparar la fortaleza de los argumentos en conflicto para determinar las derrotas entre los mismos. Cuando un argumento derrota a otro, este último es invalidado a menos que exista un tercer argumento que derrote al primero (y defienda al segundo), produciendo que el segundo vuelva a ser válido. A su vez, cada uno de estos argumentos pueden tener otros derrotadores, y así sucesivamente. Como la relación de derrota entre los argumentos puede ser muy compleja, para determinar cuándo un argumento debe ser aceptado se lleva a cabo un proceso de evaluación dialéctico que tiene en cuenta todos los argumentos que lo atacan y defienden de forma directa o indirecta. Finalmente, las conclusiones de los argumentos aceptados dan lugar al conjunto de conclusiones garantizadas (*W*).

En lo que sigue de este capítulo, se explicará cómo se identifican los conflictos entre los argumentos y la manera en la que se comparan para determinar las derrotas entre ellos, de acuerdo con los intereses de los agentes y la prioridad entre los mismos. Finalmente, se estudiarán algunas propiedades acerca del comportamiento del proceso de evaluación dialéctico con respecto a los argumentos aceptados y el conjunto de conclusiones garantizadas. Además, este capítulo cuenta con una sección de resultados formales en la que se demuestra la consistencia del conjunto de conclusiones garantizadas. Esto significa, que el proceso de evaluación dialéctico no puede dar por aceptados dos argumentos de preferencia cuyas conclusiones representen preferencias opuestas, es decir, dos argumentos que estén en conflicto, cuyas conclusiones sean $pref(d_x, d_y)$ y $pref(d_y, d_x)$. Esta propiedad, asegura que el proceso de evaluación dialéctico funciona correctamente.

3.1. Conflicto y derrota

Como fue introducido en el capítulo anterior (ver Definición 2.3.1), un argumento de preferencia construido a partir de un escenario de elección $\langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$ y una *DePref*-rule r es un tupla $\langle r(d_x, d_y), \text{pref}(d_x, d_y) \rangle$ tal que $d_x, d_y \in \mathcal{D}$, E satisface a $r(d_x, d_y)$ y $\text{pref}(d_x, d_y)$ es la conclusión de $r(d_x, d_y)$. Dado un argumento de preferencia $\mathcal{A} = \langle r(d_x, d_y), \text{pref}(d_x, d_y) \rangle$, $\text{rule}(\mathcal{A}) = r$ identifica la *DePref*-rule con la que fue construido \mathcal{A} , y $\text{claim}(\mathcal{A}) = \text{pref}(d_x, d_y)$ identifica la conclusión de \mathcal{A} . En esta propuesta, los conflictos surgen por la existencia de argumentos cuyas conclusiones son opuestas (ver Definición 2.2.4). Es decir, que la preferencia que sustenta uno, contradice la preferencia que sustenta el otro.

Definición 3.1.1 (Conflicto). *Dos argumentos \mathcal{A} y \mathcal{B} están en conflicto si y sólo si $\text{claim}(\mathcal{A})$ y $\text{claim}(\mathcal{B})$ son opuestos.*

Ejemplo 3.1.1. *Considere los argumentos de preferencia del Ejemplo 2.3.1, y sean*

- $\mathcal{H}_1 = \langle r_1(\text{jngla}, \text{playa}), \text{pref}(\text{jngla}, \text{playa}) \rangle$,
- $\mathcal{H}_2 = \langle r_2(\text{playa}, \text{jngla}), \text{pref}(\text{playa}, \text{jngla}) \rangle$,
- $\mathcal{H}_3 = \langle r_4(\text{jngla}, \text{playa}), \text{pref}(\text{jngla}, \text{playa}) \rangle$,
- $\mathcal{H}_4 = \langle r_6(\text{playa}, \text{jngla}), \text{pref}(\text{playa}, \text{jngla}) \rangle$,
- $\mathcal{H}_5 = \langle r_7(\text{jngla}, \text{playa}), \text{pref}(\text{jngla}, \text{playa}) \rangle$.

Se puede observar que \mathcal{H}_1 , \mathcal{H}_3 y \mathcal{H}_5 , los cuales poseen la conclusión $\text{pref}(\text{jngla}, \text{playa})$, están en conflicto con \mathcal{H}_2 y \mathcal{H}_4 , los cuales poseen la conclusión $\text{pref}(\text{playa}, \text{jngla})$.

Anteriormente, se dijo que todos los agentes deben aceptar el mismo conjunto de *DePref*-rules con el cual las alternativas van a ser comparadas (Observación 2.2.1). A pesar de esta restricción, sería válido pensar que los agentes podrían no estar completamente de acuerdo en cuáles son los intereses más importantes para el grupo y/o para sí mismos. Si bien un agente puede aceptar las razones de otro agente para preferir una alternativa por sobre otra, es decir, puede aceptar una *DePref*-rule de otro agente, también podría considerar que existen otras *DePref*-rules que son mas importantes, siendo que estas representen mejor sus intereses o (según este agente) los intereses del grupo. Con el objetivo de modelar este tipo de situaciones, el marco de trabajo aquí propuesto permite que cada uno de los agentes establezca su propio orden de importancia entre las *DePref*-rules.

Definición 3.1.2 (Órdenes de importancia). *Sea \mathcal{R} un conjunto de DePref-rules, y Ag un conjunto de agentes. Una lista de órdenes de importancia \triangleright para Ag es tal que para cada agente $a \in Ag$ existe un orden de importancia $\triangleright_a \in \triangleright$ que define un orden parcial estricto sobre los elementos de \mathcal{R} . Dadas dos reglas $r, r' \in \mathcal{R}$, $r \triangleright_a r'$ significa que r es más importante que r' para el agente a .*

En muchos escenarios pueden existir agentes cuyos intereses deban estar por encima de los intereses de otros agentes, por ejemplo, porque son agentes de mayor jerarquía, o porque la decisión los afecta en mayor medida que a los demás. Este marco de trabajo permite establecer un orden de prioridad entre los agentes, siendo posible así, darle más peso a los intereses de ciertos agentes.

Definición 3.1.3 (Orden de prioridad). *Sea Ag un conjunto de agentes. Un orden de prioridad \blacktriangleright es una relación que define un orden parcial estricto sobre los elementos de Ag . Dados dos agentes $a, a' \in Ag$, $a \blacktriangleright a'$ significa que a tiene prioridad sobre a' .*

Ejemplo 3.1.2. *De ahora en adelante, los miembros de la familia Jota también serán identificados como el conjunto de agentes $Ag_J = \{c, m, p, t\}$, siendo $c = Carolina$, $m = Mamá$, $p = Papá$ y $t = Tomás$. Como se explicó en el Ejemplo 2.2.4, el conjunto de DePref-rules que modelan las preferencias de Ag_J es $\mathcal{R}_J = \{r_1, r_2, r_3, r_4, r_5, r_6, r_7, r_8\}$. Si bien todos los miembros de la familia aceptan las reglas del conjunto \mathcal{R}_J , cada uno tiene sus propio orden de importancia: $\triangleright_J = [\triangleright_c, \triangleright_m, \triangleright_p, \triangleright_t]$.*

El orden de importancia de Carolina es $\triangleright_c = \{r_2 \triangleright_c r_1, r_2 \triangleright_c r_3, r_2 \triangleright_c r_7, r_4 \triangleright_c r_2, r_5 \triangleright_c r_2\}$. Esto es, Carolina considera que el entretenimiento (r_2) es más importante que la calidad de servicio (r_3) y que la cantidad de días de paseo (r_1). Incluso, considera que r_2 es más importante que r_7 ya que en general prefiere más entretenimiento a tener más días de paseo, independientemente de la calidad de servicio. Sin embargo, piensa que r_4 y r_5 son más importantes que r_2 ya que esas dos DePref-rules especifican situaciones excepcionales en las que el entretenimiento deja de ser el criterio más relevante.

Los ordenes de importancia de la mamá y el papá son $\triangleright_m = \{r_3 \triangleright_m r_1, r_3 \triangleright_m r_2, r_6 \triangleright_m r_1, r_6 \triangleright_m r_7, r_2 \triangleright_m r_7\}$ y $\triangleright_p = \{r_3 \triangleright_p r_1, r_3 \triangleright_p r_2, r_7 \triangleright_p r_2, r_7 \triangleright_p r_6, r_1 \triangleright_p r_6\}$. Tanto la mamá como el papá, prefieren la calidad de servicio sobre los otros dos criterios. Sin embargo, la mamá considera que r_2 es más importante que r_7 , mientras que el papá considera que r_7 es más importante que r_2 . Esto es porque ante iguales condiciones en cuanto a la calidad de servicio, para la mamá, es más importante el entretenimiento, mientras que el papá

prefiere tener más días de paseo. Estos dos agentes tampoco coinciden en cuanto a la importancia entre las DePref-rules r_1 , r_6 y r_7 .

El orden de importancia de Tomás es $\triangleright_t = \{r_1 \triangleright_t r_2, r_1 \triangleright_t r_3, r_2 \triangleright_t r_3, r_8 \triangleright_t r_1\}$. Tomás prefiere tener vacaciones más largas a tener unas más entretenidas o con mejor servicio, mientras que entre entretenimiento y servicio, prefiere mayor entretenimiento. Además, Tomás considera que la regla r_8 es más importante que la r_1 ya que r_8 codifica la única situación en la que, para Tomás, los días vacaciones dejarían de ser el criterio más importante.

En cuanto al orden de prioridad entre los Jota, debido a que ellos están planeando el viaje para celebrar el cumpleaños de Carolina, los intereses de ella tendrán prioridad por sobre los del resto. No habrá prioridad entre los intereses de Mamá y Papá, aunque ellos tendrán prioridad por sobre los intereses de Tomás. Por lo tanto, el orden de prioridad entre los Jota queda definido como $\blacktriangleright_J = \{c \blacktriangleright m, c \blacktriangleright p, c \blacktriangleright t, m \blacktriangleright t, p \blacktriangleright t\}$.

Para poder evaluar los argumentos, y determinar cuales son aceptados, se utilizará un marco argumentativo. Este marco utilizará los órdenes de importancia de los agentes y la prioridad entre ellos para determinar si un argumento derrota a otro (o no), cuando están en conflicto.

Definición 3.1.4 (Marco de evaluación de preferencias multi-agente). *Un marco de evaluación de preferencias multi-agente (abreviado MEPM) es una tupla $\langle Ag, Args, \triangleright, \blacktriangleright \rangle$ donde Ag es un conjunto de agentes, $Args$ es un conjunto de argumentos de preferencia, \triangleright es una lista de ordenes de importancia para Ag , y \blacktriangleright es un orden de prioridad sobre Ag .*

En un MEPM los argumentos serán comparados por su fuerza. Para este trabajo, se propone que la fortaleza de los argumentos dependa de los intereses individuales de cada agente, y de la prioridad de los agentes dentro del grupo.

Definición 3.1.5 (Fortaleza). *Sea $\mathbf{F} = \langle Ag, Args, \triangleright, \blacktriangleright \rangle$ un MEPM tal que $\mathcal{A}, \mathcal{B} \in Args$. Se dice que \mathcal{A} es más fuerte que \mathcal{B} en \mathbf{F} si y sólo si existe $a \in Ag$ tal que $rule(\mathcal{A}) \triangleright_a rule(\mathcal{B})$, y para todo $b \in Ag$ tal que $rule(\mathcal{B}) \triangleright_b rule(\mathcal{A})$, se cumple que $a \blacktriangleright b$.*

Ejemplo 3.1.3. *Considere los argumentos del Ejemplo 3.1.1, y los ordenes de importancia y de prioridad entre los agentes presentados en el Ejemplo 3.1.2. Al comparar los argumentos que están en conflicto se tiene que:*

- \mathcal{H}_2 es más fuerte que \mathcal{H}_1 ya que Carolina considera que r_2 es más importante que r_1 ($r_2 \triangleright_c r_1$), y aunque Tomás considere lo contrario ($r_1 \triangleright_t r_2$), Carolina tiene prioridad sobre Tomás ($c \blacktriangleright_J t$).

- \mathcal{H}_2 es más fuerte que \mathcal{H}_5 ya que Carolina considera que r_2 es más importante que r_7 ($r_2 \triangleright_c r_7$), y aunque su papá considere lo contrario ($r_7 \triangleright_p r_2$), Carolina tiene prioridad sobre su papá ($c \blacktriangleright_J p$).

- \mathcal{H}_3 es más fuerte que \mathcal{H}_2 ya que Carolina considera que r_4 es más importante que r_2 ($r_4 \triangleright_c r_2$), y no existe ningún otro agente que considere lo contrario, es decir, $\nexists x \in Ag_J$ tal que $r_2 \triangleright_x r_4$.

Note que, a pesar de que $r_1 \triangleright_p r_6$, \mathcal{H}_1 no es más fuerte que \mathcal{H}_4 , ni vice-versa, ya que $r_6 \triangleright_m r_1$ y no existe prioridad entre Mamá y Papá. Lo mismo ocurre entre \mathcal{H}_1 y \mathcal{H}_5 . Tampoco \mathcal{H}_3 es más fuerte que \mathcal{H}_4 , ni vice-versa, ya que ningún agente considera que r_4 sea más importante que r_6 , o vice-versa.

Teniendo una herramienta con la que comparar los argumentos que forman parte de un MEPM, es posible determinar si un argumento derrota a otro cuando están en conflicto. Esto permitirá, en parte, decidir si un argumento debe ser aceptado, o no.

Definición 3.1.6 (Derrota). Sea $\mathbf{F} = \langle Ag, Args, \triangleright, \blacktriangleright \rangle$ un MEPM y sean \mathcal{A} y \mathcal{B} dos argumentos en conflicto pertenecientes a $Args$. Se dice que \mathcal{B} derrota a \mathcal{A} en \mathbf{F} si y sólo si \mathcal{A} no es más fuerte que \mathcal{B} en \mathbf{F} .

Ejemplo 3.1.4. Sea $\mathbf{F}_{J'} = \{Ag_J, Args_{J'}, \triangleright_J, \blacktriangleright_J\}$ un MEPM, donde $Args_{J'} = \{\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3, \mathcal{H}_4, \mathcal{H}_5\}$ ($Args_{J'} \subset Args_J$) es el conjunto de argumentos de preferencia del Ejemplo 3.1.1. De acuerdo con los resultados mostrados en el Ejemplo 3.1.3, las derrotas que ocurren en $\mathbf{F}_{J'}$ son las siguientes:

- \mathcal{H}_1 derrota a \mathcal{H}_4 ,
- \mathcal{H}_2 derrota a \mathcal{H}_1 y a \mathcal{H}_5 ,
- \mathcal{H}_3 derrota a \mathcal{H}_2 y a \mathcal{H}_4 ,
- \mathcal{H}_4 derrota a \mathcal{H}_1 , a \mathcal{H}_3 y a \mathcal{H}_5 ,
- \mathcal{H}_5 derrota a \mathcal{H}_4 .

Dados dos argumentos en conflicto \mathcal{A} y \mathcal{B} , puede darse una de tres situaciones posibles:

- i) que \mathcal{A} derrote a \mathcal{B} y que \mathcal{B} no derrote a \mathcal{A} ya que \mathcal{A} es más fuerte que \mathcal{B} ;
- ii) que \mathcal{B} derrote a \mathcal{A} y que \mathcal{A} no derrote a \mathcal{B} ya que \mathcal{B} es más fuerte que \mathcal{A} ; o bien,
- iii) que \mathcal{A} y \mathcal{B} se derroten mutuamente ya que ninguno es más fuerte que el otro.

No puede suceder, que entre dos argumentos en conflicto no existan derrotas.

En este trabajo se propone utilizar un grafo de argumentos para visualizar, no sólo las derrotas entre los argumentos, sino también, la relación de conflicto.

Definición 3.1.7 (Grafo de argumentos). *Un grafo de argumentos para un MEPM $\mathbf{F} = \langle Ag, Args, \triangleright, \blacktriangleright \rangle$ es un grafo dirigido $\langle Args, R \rangle$, donde $Args$ son los nodos del grafo, y $R = R_d \cup R_c$ ($R_d \cap R_c = \emptyset$) son los arcos del grafo tal que todo arco $(\mathcal{A}, \mathcal{B}) \in R$ está etiquetado con el conjunto $\{a \in Ag \mid rule(\mathcal{A}) \triangleright_a rule(\mathcal{B})\}$, y*

- $(\mathcal{A}, \mathcal{B}) \in R_d$ si y sólo si \mathcal{A} derrota a \mathcal{B} en \mathbf{F} ,
- $(\mathcal{A}, \mathcal{B}) \in R_c$ si y sólo si \mathcal{A} y \mathcal{B} están en conflicto y \mathcal{A} no derrota a \mathcal{B} en \mathbf{F} .

Un grafo de argumentos posee dos tipos de arcos: aquellos perteneciente a R_d , y aquellos pertenecientes a R_c , observe que $R_d \cap R_c = \emptyset$. Los arcos en R_d representan las derrotas entre los argumentos, y los arcos en R_c representan que un argumento está en conflicto con otro, pero que no pudo derrotarlo. Además, todo arco $(\mathcal{A}, \mathcal{B}) \in R$ está etiquetado con el conjunto de agentes que considera que $rule(\mathcal{A})$ es más importante que $rule(\mathcal{B})$.

La Figura 3.2 muestra el grafo de argumentos para el MEPM \mathbf{F}_J del Ejemplo 3.1.4. Cada argumento (nodo) está representado por un triángulo que contiene el identificador del argumento y la *DePref*-rule a partir de la cual fue construido. En el vértice superior del triángulo se encuentra la conclusión del argumento. Por ejemplo, de la Figura 3.2, se puede observar que \mathcal{H}_1 fue construido a partir de r_1 y que su conclusión es $pref(jngla, playa)$, así como también que \mathcal{H}_4 fue construido a partir de r_6 y que su conclusión es $pref(playa, jngla)$. Note que en esta figura los literales *jngla* y *playa* fueron abreviados como *j* y *p*, respectivamente.

otro \mathcal{B} , \mathcal{B} no debe más fuerte que \mathcal{A} , para lo que se tiene que dar que ninguno de los agentes que consideran que la $\text{rule}(\mathcal{B})$ es más importante que $\text{rule}(\mathcal{A})$ tenga más prioridad que todos los agentes que consideran que $\text{rule}(\mathcal{A})$ es más importante que $\text{rule}(\mathcal{B})$.

3.2. Proceso de evaluación dialéctico

Como se puede observar en la Figura 3.2, un argumento puede derrotar a varios argumentos a la vez, como sucede con \mathcal{H}_2 que derrota a \mathcal{H}_1 y a \mathcal{H}_5 , y también puede pasar que un argumento sea derrotado por varios argumentos a la vez, como es el caso de \mathcal{H}_4 que es derrotado por \mathcal{H}_1 , \mathcal{H}_5 y \mathcal{H}_3 . Además, para el caso de \mathcal{H}_4 , se tiene que este derrota a todos sus derrotadores, entre los que se encuentra \mathcal{H}_3 , que a su vez también derrota a \mathcal{H}_2 . Queda claro, entonces, que un argumento puede tener varios derrotadores, y que a su vez estos derrotadores pueden tener otros derrotadores (que defienden al primero), y así siguiendo. Incluso podrían aparecer ciclos de argumentos que se derrotan entre sí. Por lo tanto, para determinar qué argumentos deben prevalecer, no alcanza con sólo establecer las derrotas entre los argumentos, sino que hace falta considerar y analizar, de forma exhaustiva, toda secuencia de derrotadores. El proceso con el que se evaluarán los argumentos, está inspirado en el proceso dialéctico usado en DeLP ([GS04]), y está basado en las nociones de línea de argumentación aceptable y árboles de dialéctica que serán introducidas a continuación.

Definición 3.2.1 (Línea de argumentación aceptable). *Sea $\mathbf{F} = \langle Ag, Args, \triangleright, \blacktriangleright \rangle$ un MEPM tal que $\mathcal{A}_1 \in Args$. Una línea de argumentación aceptable para \mathcal{A}_1 en \mathbf{F} es una secuencia finita de argumentos $\lambda = [\mathcal{A}_1 \dots \mathcal{A}_n]$, donde $\mathcal{A}_i \in Args$, \mathcal{A}_i derrota a \mathcal{A}_{i-1} en \mathbf{F} ($1 < i \leq n$) y $\mathcal{A}_j \neq \mathcal{A}_k$ ($1 \leq j, k \leq n$).*

Esto es, en toda línea de argumentación aceptable, excepto por el primer argumento, cada elemento de la secuencia debe derrotar a su predecesor, y para evitar ciclos, ningún argumento puede aparecer dos veces. Por ejemplo, $[\mathcal{H}_1]$, $[\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3]$ y $[\mathcal{H}_1, \mathcal{H}_4, \mathcal{H}_5, \mathcal{H}_2, \mathcal{H}_3]$ son tres líneas de argumentación aceptables para el argumento \mathcal{H}_1 del MEPM de la Figura 3.2. En cambio, la secuencia $[\mathcal{H}_1, \mathcal{H}_4, \mathcal{H}_1]$ no es una línea de argumentación aceptable ya que \mathcal{H}_1 aparece dos veces, como tampoco lo es la secuencia $[\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_5]$ ya que \mathcal{H}_5 no derrota a \mathcal{H}_2 .

Vale la pena aclarar, que la noción de línea de argumentación aceptable utilizada en esta propuesta difiere de aquella definida para DeLP. En DeLP, se dice que un argumento derrota a otro por bloqueo si ambos argumentos se derrotan mutuamente, tal como pasa entre \mathcal{H}_1 y \mathcal{H}_4 , o como entre este último y \mathcal{H}_5 . En dicha propuesta, para que una línea de argumentación sea aceptable no pueden ocurrir dos bloqueos consecutivos, es decir, si aparece un argumento que derrota al anterior por bloque, entonces el próximo debe ser un derrotador propio (no por bloqueo). Por lo tanto, en DeLP, $[\mathcal{H}_1, \mathcal{H}_4, \mathcal{H}_5]$ no sería una línea de argumentación aceptable. Esta restricción se decidió eliminar en base a una de las observaciones de [GPS20]. Allí, se muestra que al considerar esta restricción, puede ocurrir que un argumento podría verse favorecido en un escenario en el sus derrotadores sean propios, en lugar de que lo derroten por bloqueo. Es decir, un argumento puede ser favorecido en un escenario en el que los derrotadores son más fuertes, lo cual resulta un poco contradictorio.

Definición 3.2.2 (Segmento). *Dada una línea de argumentación aceptable $\lambda = [\mathcal{A}_1, \dots, \mathcal{A}_n]$, una secuencia de argumentos no vacía $[\mathcal{B}_1, \dots, \mathcal{B}_m]$ es un segmento de λ si y sólo si existe un argumento \mathcal{A}_i en λ ($1 \leq i < n$) tal que $\mathcal{A}_{i+k} = \mathcal{B}_{1+k}$ para todo k con $0 \leq k < m$.*

Considere nuevamente la línea de argumentación aceptable $[\mathcal{H}_1, \mathcal{H}_4, \mathcal{H}_5, \mathcal{H}_2, \mathcal{H}_3]$. Las secuencias $[\mathcal{H}_1, \mathcal{H}_4, \mathcal{H}_5]$ y $[\mathcal{H}_5, \mathcal{H}_2]$ son segmentos de esta línea, pero $[\mathcal{H}_5, \mathcal{H}_3]$ no lo es. Como caso particular, se dice que un segmento σ de una línea de argumentación aceptable λ es un **segmento inicial** de λ si y sólo si σ y λ comienzan con el mismo argumento. Por ejemplo, $[\mathcal{H}_1]$ y $[\mathcal{H}_1, \mathcal{H}_4, \mathcal{H}_5]$ son segmentos iniciales de $[\mathcal{H}_1, \mathcal{H}_4, \mathcal{H}_5, \mathcal{H}_2, \mathcal{H}_3]$. Observe que la secuencia vacía no es un segmento de ninguna línea, que toda línea de argumentación aceptable es un segmento de sí misma, y que una secuencia con un único argumento \mathcal{A} es un segmento de cualquier línea de argumentación aceptable que contenga a \mathcal{A} .

Proposición 3.2.1. *Sea λ una línea de argumentación aceptable, cualquier segmento de λ es una línea de argumentación aceptable.*

Demostración. Sea $\lambda = [\mathcal{A}_1, \dots, \mathcal{A}_n]$ una línea de argumentación aceptable, y $\sigma = [\mathcal{B}_1, \dots, \mathcal{B}_m]$ un segmento de λ . Por la Definición 3.2.2, debe existir un argumento \mathcal{A}_i en λ ($1 \leq i < n$) tal que $\mathcal{A}_{i+k} = \mathcal{B}_{1+k}$ para todo k con $0 \leq k < m$. Por el absurdo, suponga que σ no es una línea de argumentación aceptable. Luego, de acuerdo con la Definición 3.2.1, para que esto suceda, debe existir un argumento \mathcal{B}_j en σ con $j > 1$ tal

que \mathcal{B}_j no derrote a \mathcal{B}_{j-1} , o bien, $\mathcal{B}_j = \mathcal{B}_{j-k}$, para algún k con $1 \leq k < j$. Si esto fuese así, existiría un argumento \mathcal{A}_j en λ ($\mathcal{A}_j = \mathcal{B}_j$) que no derrota a \mathcal{A}_{j-1} ($\mathcal{A}_{j-1} = \mathcal{B}_{j-1}$), o bien, $\mathcal{A}_j = \mathcal{A}_{j-k}$ ($\mathcal{A}_{k-j} = \mathcal{B}_{k-j}$), para algún k con $1 \leq k < j$. Sin embargo, esto contradice la hipótesis de que λ es una línea de argumentación aceptable. Por lo tanto, σ es una línea de argumentación aceptable. \square

En una línea de argumentación aceptable $\lambda = [\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_4, \dots, \mathcal{A}_n]$, se sabe que \mathcal{A}_2 derrota a \mathcal{A}_1 , y como \mathcal{A}_3 derrota a \mathcal{A}_2 , entonces (indirectamente) defiende a \mathcal{A}_1 . De igual forma, como \mathcal{A}_4 derrota a \mathcal{A}_3 , (indirectamente) defiende a \mathcal{A}_2 . Por lo tanto, dada una línea de argumentación aceptable λ , es posible identificar si un argumento está a favor o en contra del primer argumento de λ , a partir de la posición en la que se encuentra.

Definición 3.2.3 (Conjuntos de soporte e interferencia). *Sea $\lambda = [\mathcal{A}_1, \dots, \mathcal{A}_n]$ ($1 \leq n$), el conjunto de argumentos de soporte de λ es $\mathbf{pros}(\lambda) = \{\mathcal{A}_i \text{ en } \lambda \mid i \in \text{impares}(\{1, \dots, n\})\}$, y el conjunto de argumentos de interferencia de λ es $\mathbf{cons}(\lambda) = \{\mathcal{A}_i \text{ en } \lambda \mid i \in \text{pares}(\{1, \dots, n\})\}$.*

Los argumentos de soporte son aquellos que se encuentran en posiciones impares, y los argumentos de interferencia son aquellos que se encuentran en posiciones pares.

Observación 3.2.1. *Sea λ una línea de argumentación aceptable para un argumento \mathcal{A} de un MEPM tal que $\mathbf{claim}(\mathcal{A}) = \alpha$. Todos los argumentos en $\mathbf{pros}(\lambda)$ tendrán como conclusión α , y todos los argumentos en $\mathbf{cons}(\lambda)$ tendrán como conclusión $\bar{\alpha}$.*

Como se indica en la Observación 3.2.1, en el contexto del marco de trabajo aquí propuesto, las conclusiones de todos los argumentos en $\mathbf{pros}(\lambda)$ serán las mismas, al igual que sucede con la de los argumentos en $\mathbf{cons}(\lambda)$. Más aún, la conclusión de los argumentos en $\mathbf{pros}(\lambda)$ será la conclusión opuesta a la de los argumentos en $\mathbf{cons}(\lambda)$. Por ejemplo, para la línea de argumentación $[\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3, \mathcal{H}_4, \mathcal{H}_5]$, se tiene que la conclusión de los argumentos \mathcal{H}_1 , \mathcal{H}_3 y \mathcal{H}_5 es $\mathbf{pref}(\mathbf{jngla}, \mathbf{playa})$, mientras que \mathcal{H}_2 y \mathcal{H}_4 tienen la conclusión opuesta, $\mathbf{pref}(\mathbf{playa}, \mathbf{jngla})$.

Como se muestra en la Figura 3.2, \mathcal{H}_1 es derrotado por \mathcal{H}_2 y \mathcal{H}_4 , por lo que de cada uno de ellos se desprenden líneas de argumentación diferentes: aquellas que comienzan con la secuencia $[\mathcal{H}_1, \mathcal{H}_2]$, y aquellas que comienzan con $[\mathcal{H}_1, \mathcal{H}_4]$. A su vez, como \mathcal{H}_4 es derrotado por \mathcal{H}_3 y \mathcal{H}_5 , de $[\mathcal{H}_1, \mathcal{H}_4]$ se desprenden otras dos líneas más: $[\mathcal{H}_1, \mathcal{H}_4, \mathcal{H}_3]$ y $[\mathcal{H}_1, \mathcal{H}_4, \mathcal{H}_5]$.

Note que si bien \mathcal{H}_4 es derrotado por \mathcal{H}_1 , este último no se puede introducir en ninguna secuencia porque ya aparece al principio de todas ellas. De esta forma, partiendo de un único argumento, se puede construir un árbol para analizar, en conjunto, todos los derrotadores que están relacionados con ese argumento.

Definición 3.2.4 (Árbol de argumentación). Sea $\mathbf{F} = \langle Ag, Args, \triangleright, \blacktriangleright \rangle$ un MEPM y $\mathcal{A} \in Args$. Un árbol de argumentación T para \mathcal{A} es un árbol tal que:

(i) La raíz está etiquetada con \mathcal{A} , y todo nodo interno está etiquetado con un derrotador para la etiqueta de su padre.

(ii) No existen dos hermanos en T que estén etiquetados con el mismo argumento; y

(iii) La secuencia de etiquetas de cada camino desde la raíz a una de las hojas de T debe ser una línea de argumentación aceptable para \mathcal{A} en \mathbf{F} .

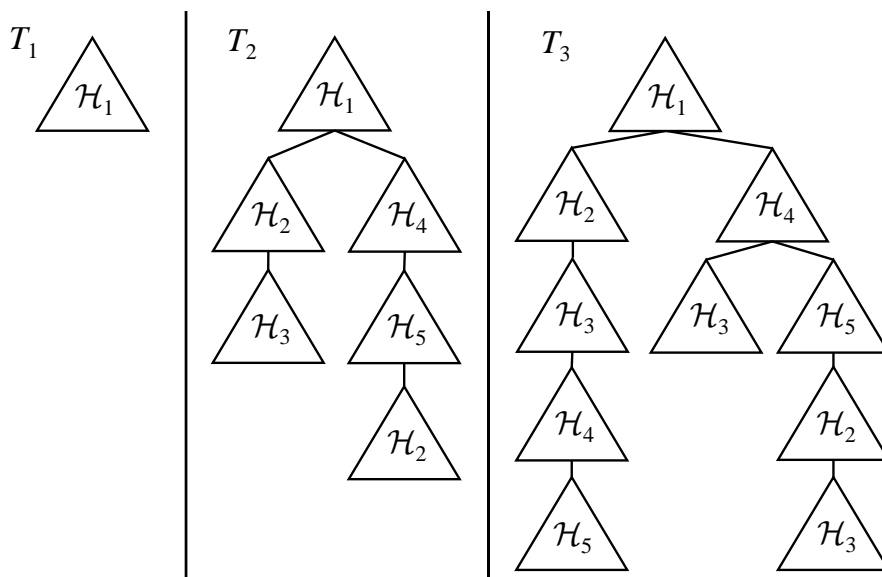


Figura 3.3: Algunos árboles de argumentación para el argumento \mathcal{H}_1 del MEPM de la Figura 3.2.

Un árbol de argumentación puede verse como un conjunto de líneas de argumentación aceptables organizadas en forma de árbol. Dado un árbol de argumentación T , si λ es una secuencia de etiquetas que se obtiene siguiendo un camino desde la raíz de T hasta una de sus hojas, entonces se dice que λ es una línea de argumentación aceptable de T , o que λ fue usada para construir T . Tres árboles de argumentación para el argumento \mathcal{H}_1

del MEPM \mathbf{F}_J de la Figura 3.2 son mostrados en la Figura 3.3. Los nodos de los árboles están representados con triángulos, dentro de los cuales se encuentra el identificador del argumento con el que está etiquetado cada nodo. Las líneas que unen a los nodos representan la relación *padre-hijo*. Observe que, la única línea de T_1 es $[\mathcal{H}_1]$, las líneas de T_2 son $[\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3]$ y $[\mathcal{H}_1, \mathcal{A}_4, \mathcal{H}_5, \mathcal{H}_2]$, y las líneas de T_3 son $[\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3, \mathcal{H}_4, \mathcal{H}_5]$, $[\mathcal{H}_1, \mathcal{H}_4, \mathcal{H}_3]$ y $[\mathcal{H}_1, \mathcal{H}_4, \mathcal{H}_5, \mathcal{H}_2, \mathcal{H}_3]$.

La razón detrás de hacer explícita la distinción entre los nodos de un árbol de argumentación y su etiqueta es porque un mismo argumento puede ser la etiqueta de varios nodos del mismo árbol. Por ejemplo, \mathcal{H}_3 aparece tres veces en T_3 . La restricción de que en un árbol de argumentación no pueda haber dos hermanos con la misma etiqueta, apunta a que no existan dos nodos cuyas secuencias de etiquetas desde la raíz sean las mismas. De esta forma, se evita que existan dos caminos desde la raíz hasta una de las hojas cuya secuencia de etiquetas sean las mismas. Se puede ver como las líneas $[\mathcal{H}_1, \mathcal{H}_4, \mathcal{H}_3]$ y $[\mathcal{H}_1, \mathcal{H}_4, \mathcal{H}_5, \mathcal{H}_2, \mathcal{H}_3]$ de T_3 que comparten el segmento $[\mathcal{H}_1, \mathcal{H}_4]$ también comparten los dos primeros nodos en el camino cuya secuencia de etiquetas es igual a estas líneas.

Definición 3.2.5 (Línea exhaustiva). *Sea $\mathbf{F} = \langle Ag, Args, \triangleright, \blacktriangleright \rangle$ un MEPM y $\lambda = [\mathcal{A}_1, \dots, \mathcal{A}_n]$ una línea de argumentación aceptable para un argumento \mathcal{A}_1 en \mathbf{F} ($\mathcal{A}_1 \in Args$). Se dice que λ es exhaustiva si y sólo si no existe $\mathcal{A}_{n+1} \in Args$ tal que \mathcal{A}_{n+1} derrote a \mathcal{A}_n y \mathcal{A}_{n+1} no aparezca en λ .*

En otras palabras, una línea de argumentación aceptable para un argumento es exhaustiva si y sólo si no es posible agregar un nuevo argumento al final de esa línea, y así obtener una línea de argumentación aceptable más larga que la anterior. Considerando las líneas de argumentación con las que se construyeron los árboles de la Figura 3.3, se puede observar que solo las líneas de T_3 ($[\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3, \mathcal{H}_4, \mathcal{H}_5]$, $[\mathcal{H}_1, \mathcal{H}_4, \mathcal{H}_3]$ y $[\mathcal{H}_1, \mathcal{H}_4, \mathcal{H}_5, \mathcal{H}_2, \mathcal{H}_3]$) son exhaustivas. Más aún, estas son todas las líneas de argumentación aceptable exhaustivas para el argumento \mathcal{H}_1 del MEPM de la Figura 3.2. Las líneas de los otros árboles ($[\mathcal{H}_1]$, $[\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3]$ y $[\mathcal{H}_1, \mathcal{H}_4, \mathcal{H}_5, \mathcal{H}_2]$) no son exhaustivas ya que todas son segmentos de alguna línea de T_3 , lo que significa que es posible agregarles más argumentos al final y obtener una nueva línea de argumentación aceptable.

Definición 3.2.6 (Árbol de dialéctica). *Sea $\mathbf{F} = \langle Ag, Args, \triangleright, \blacktriangleright \rangle$ un MEPM tal que $\mathcal{A} \in Args$, T un árbol de argumentación para \mathcal{A} , y Λ el conjunto de todas las líneas de argumentación aceptables exhaustivas para \mathcal{A} en \mathbf{F} . Se dice que T es el árbol de dialéctica*

para \mathcal{A} , denotado $\mathcal{T}(\mathcal{A})$, si y sólo si el conjunto de todas las líneas de argumentación de T es igual a Λ .

Todo árbol de dialéctica es un árbol de argumentación, pero la inversa no vale. Note que a pesar de que pueden existir múltiples árboles de argumentación para un dado argumento \mathcal{A} , el árbol de dialéctica para \mathcal{A} es único ya que este incluye todas las líneas de argumentación aceptables exhaustivas para \mathcal{A} . De los árboles de argumentación mostrados en la Figura 3.3, el único árbol de dialéctica es $T_3 = \mathcal{T}(\mathcal{H}_1)$.

Dado un argumento \mathcal{A} , el árbol de dialéctica $\mathcal{T}(\mathcal{A})$ será usado para determinar la aceptabilidad de \mathcal{A} , ya que esta estructura relaciona todos los argumentos que atacan y/o defienden (directa o indirectamente) a \mathcal{A} . Para esto, se utilizará el procedimiento de marcado $\mathbf{m}(\cdot)$ definido a continuación.

Definición 3.2.7 (Procedimiento de marcado). *El procedimiento de marcado \mathbf{m} toma un árbol de argumentación T , y lo marca de la siguiente forma:*

- i) Un nodo es marcado como \mathbf{U} (undefeated) si y sólo si no tiene hijos, o bien, ninguno de sus hijos está marcado \mathbf{U} .*
- ii) Un nodo es marcado como \mathbf{D} (defeated) si y sólo si alguno de sus hijos está marcado \mathbf{U} .*

Dado un árbol de argumentación T , el procedimiento de marcado \mathbf{m} comenzará marcando las hojas de T como \mathbf{U} . Note que \mathbf{m} no podría comenzar de ningún otro punto ya que las marcas de todos los demás nodos dependen de las marcas de sus hijos. Luego, los padres de toda hoja quedarán marcados como \mathbf{D} , y así, el proceso seguirá avanzando desde abajo hacia arriba hasta llegar a la raíz. La Figura 3.4 muestra tres instantes diferentes que ocurren durante la ejecución del procedimiento de marcado aplicado sobre el árbol de dialéctica para el argumento \mathcal{H}_1 del MEPM de la Figura 3.2. Comenzando por la rama de la izquierda (Figura 3.4(1)), la hoja etiquetada con \mathcal{H}_5 es marcada \mathbf{U} ya que no posee ningún hijo. Luego, el padre de esta hoja (etiquetado con \mathcal{H}_4) es marcado como \mathbf{D} ya que tiene al menos un hijo marcado como \mathbf{U} , su padre (etiquetado con \mathcal{H}_3) es marcado como \mathbf{U} ya que todos sus hijos están marcados \mathbf{D} , y el padre de este último (etiquetado con \mathcal{H}_2) es marcado como \mathbf{D} . Para determinar la marca de la raíz, primero se debe determinar la marca de su hijo derecho etiquetado con \mathcal{H}_4 (Figura 3.4(2)). Observe, que el hijo izquierdo de este nodo (etiquetado con \mathcal{H}_3) es una hoja, por lo que es marcada como \mathbf{U} . Por lo

tanto, la marca de su padre debe ser **D**, independientemente de la marca que tenga su hijo derecho. Luego, la raíz es marcada como **U**. Note que, si el objetivo con el que se efectúa el procedimiento de marcado es sólo el de conocer la marca de la raíz, entonces el proceso podría terminar en este punto. No obstante, como será explicado más adelante, conocer la marca de todos los nodos del árbol podría resultar útil para poder brindar una explicación más detallada acerca del comportamiento del sistema. Finalmente, para terminar el proceso (Figura 3.4(3)), se marca la hoja de más a la derecha (etiquetada con \mathcal{H}_3) como **U**, su padre (etiquetado con \mathcal{H}_2) es marcado como **D**, y el padre (etiquetado con \mathcal{H}_5) de este último nodo es marcado como **U**.

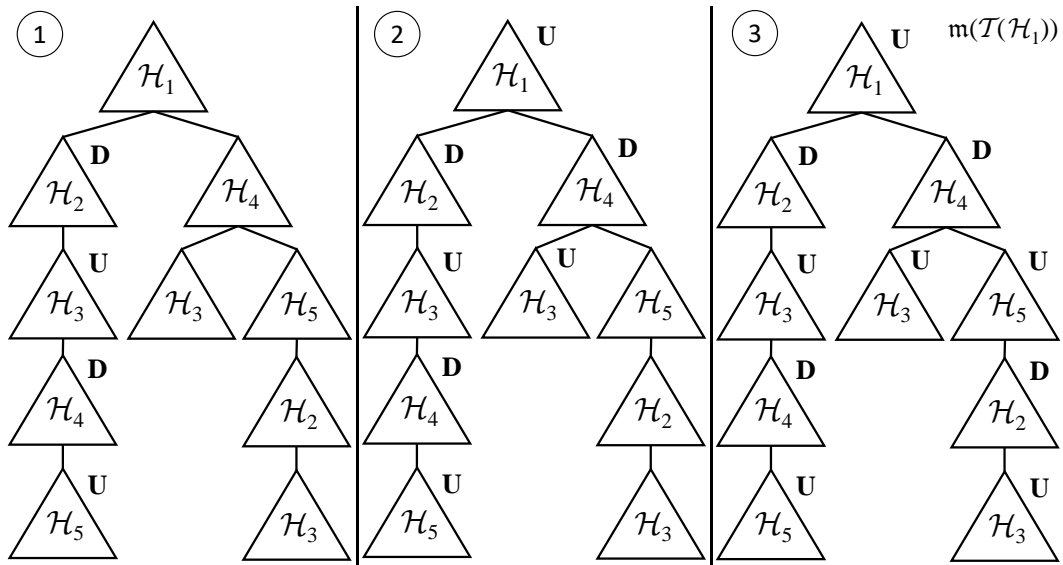


Figura 3.4: Evolución, por etapas, del procedimiento de marcado sobre el árbol de dialéctica para el argumento \mathcal{H}_1 del MEPM de la Figura 3.2.

Si al aplicar la función de marcado sobre un árbol de dialéctica, su raíz es marcada como **U**, esto quiere decir que, si existen derrotadores para el argumento en la raíz sus respectivos nodos están marcados como **D**, con lo cual el argumento de la raíz no pudo ser derrotado, y por ende, ese argumento debe ser aceptado. De lo contrario, si la raíz del árbol de dialéctica de un argumento es marcada como **D**, entonces ese argumento será rechazado.

Definición 3.2.8 (Argumento aceptado). Sea $\mathbf{F} = \langle Ag, Args, \triangleright, \blacktriangleright \rangle$ un MEPM. Se dice que un argumento $\mathcal{A} \in Args$ es aceptado con respecto a \mathbf{F} si y sólo si la raíz de $\mathbf{m}(\mathcal{T}(\mathcal{A}))$

está marcada como **U**. En cambio, si la raíz de $\mathbf{m}(\mathcal{T}(\mathcal{A}))$ está marcada como **D**, se dice que \mathcal{A} es rechazado con respecto a **F**.

De la Figura 3.4 se puede observar que la marca de la raíz de $\mathbf{m}(\mathcal{T}(\mathcal{H}_1))$ es **U**, por lo que \mathcal{A} es aceptado con respecto al MEPM de la Figura 3.2. Por otro lado, la Figura 3.5 muestra el árbol de dialéctica marcado del argumento \mathcal{H}_4 del MEPM de la Figura 3.2. Note que $\mathcal{T}(\mathcal{H}_4)$ es construido a partir de las líneas $[\mathcal{H}_4, \mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3]$, $[\mathcal{H}_4, \mathcal{H}_3]$ y $[\mathcal{H}_4, \mathcal{H}_5, \mathcal{H}_2, \mathcal{H}_3]$. En este caso, se puede ver que la marca de la raíz es **D**, por lo que se puede decir que \mathcal{H}_4 es rechazado con respecto al MEPM de la Figura 3.2.

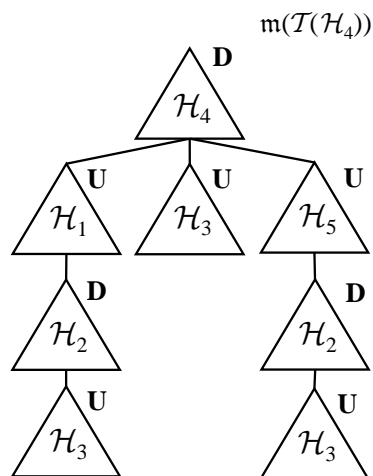


Figura 3.5: Árbol de dialéctica marcado para el argumentos \mathcal{H}_4 del MEPM de la Figura 3.2.

El siguiente algoritmo recursivo, denominado `buildMarkedTree`, muestra cómo construir un árbol de dialéctica marcado. Es importante destacar que este algoritmo no construye el árbol de dialéctica, y luego lo marca, sino que directamente lo marca a medida que lo va construyendo. Además de un MEPM y del argumento para el que se desea construir el árbol, a efectos de garantizar que las líneas de argumentación que existen dentro del árbol sean aceptables y exhaustivas, este algoritmo requiere de una lista auxiliar L con la que se llevará registro de los argumentos que ya fueron considerados dentro de una misma línea, es decir, dentro de una misma rama del árbol. Para obtener un árbol de dialéctica marcado, y así poder determinar si la conclusión de un argumento está garantizada, L deberá ser inicializada como una lista vacía (`[]`). Cabe aclarar, que para la elaboración de este algoritmo se tuvo en cuenta que todo nodo es considerado un árbol en sí mismo, y

que a partir de un nodo se puede acceder a sus descendientes, y por lo tanto, al resto del árbol. Además, para no desviar el foco de atención, se omitirán algunos detalles como la representación de la estructura interna de un nodo y la definición formal de las siguientes funciones auxiliares:

- $addChild(P, H)$, que agrega a H como nodo hijo de P y
- $addLast(L, E)$, que agrega el elemento E al final de la lista L ,
- $defeaters(F, A)$, que retorna los derrotadores del argumento A dentro del MEPM F ,
- $getMark(N)$, retorna la marca del nodo N ,
- $setMark(N, M)$, que marca al nodo N como M ,
- $treeNode(A)$, que retorna un nuevo nodo etiquetado con el argumento A .

Como será comentado más adelante, este algoritmo puede ser optimizado en caso de que sólo se desee determinar si un argumento es aceptado, pero se ha decidido no hacerlo ya que se perdería información potencialmente útil para explicar los resultados del marco de trabajo.

Algoritmo 2 Proceso de evaluación dialéctico - `buildMarkedTree`

Entrada: Un MEPM $\mathbf{F} = \langle Ag, Args, \triangleright, \blacktriangleright \rangle$, un argumento de preferencia $\mathcal{A} \in Args$, y una lista L .

Salida: La raíz del $m(\mathcal{T}(\mathcal{A}))$.

```

R ← treeNode( $\mathcal{A}$ )
setMark(R, U)
addLast(L,  $\mathcal{A}$ )
for each  $\mathcal{B} \in defeaters(\mathbf{F}, \mathcal{A})$  such that  $\mathcal{B}$  not in L do
  L' ← L
  N ← buildMarkedTree( $\mathbf{F}, \mathcal{B}, L'$ )
  addChild(R, N)
  if getMark(N) == U then
    setMark(R, D)
  end if
end for
return R

```

El algoritmo comienza generando un nuevo nodo R al cual se le asigna como etiqueta el argumento \mathcal{A} ($R \leftarrow treeNode(\mathcal{A})$), se lo marca tentativamente como U ($setMark(R, U)$),

y luego se agrega \mathcal{A} a la lista L ($addLast(L, \mathcal{A})$). A continuación, por cada uno de los derrotadores de \mathcal{A} en \mathbf{F} ($\mathcal{B} \in defeaters(\mathbf{F}, \mathcal{A})$) que no hayan sido previamente considerados (\mathcal{B} not in L), se llama a $buildMarkedTree$ ($\mathbf{N} \leftarrow buildMarkedTree(\mathbf{F}, \mathcal{B}, L')$), pero ahora pasando como parámetro al derrotador \mathcal{B} y lista L' que es igual a L ($L' \leftarrow L$). El resultado de esta llamada, que será la raíz de un árbol de argumentación para \mathcal{B} , es almacenado en \mathbf{N} . A continuación, \mathbf{N} es agregado como hijo de \mathbf{R} ($addChild(\mathbf{R}, \mathbf{N})$), y luego, si \mathbf{N} fue marcado como \mathbf{U} , se actualiza la marca de \mathbf{R} a \mathbf{D} ($setMark(\mathbf{R}, \mathbf{D})$). Note, que de cada uno de los hijos de \mathbf{R} , se desprenden las distintas líneas de argumentación aceptables que conforman el árbol, estas son almacenadas en L y L' . Además observe que por construcción, estas líneas son aceptables dado que cada argumento de L , excepto el primero, derrota al argumento que está antes del mismo ($\mathcal{B} \in defeaters(\mathbf{F}, \mathcal{A})$) y que ninguno de estos se repite (\mathcal{B} not in L). El algoritmo finaliza cuando ninguno de los argumentos que están siendo analizados, tenga más derrotadores que no hayan sido previamente incorporados dentro de la línea de argumentación a la que pertenecen. Al finalizar, el algoritmo retorna el nodo raíz de un árbol de argumentación ya marcado. Por el funcionamiento del algoritmo, cuando $L = []$, se da que todas las líneas de argumentación con las que se construye el árbol para \mathcal{A} son exhaustivas, y por lo tanto, el algoritmo termina retornando la raíz de $\mathbf{m}(\mathcal{T}(\mathcal{A}))$. Observe, que el Algoritmo 2 podría ser optimizado si luego de encontrar un hijo marcado como \mathbf{U} , y de actualizar la marca de \mathbf{R} , se dejan de lado los derrotadores restantes. Esto evitaría tener que explorar todo el árbol para conocer la marca de la raíz de un árbol de dialéctica, aunque de ser así, no sería posible recuperar la estructura completa del árbol. Como será mostrado más adelante, conocer la estructura completa de los árboles de dialéctica puede resultar de interés a la hora de brindar explicaciones acerca del funcionamiento del marco de trabajo.

Como se dijo anteriormente, las conclusiones de los argumentos aceptados serán utilizadas para determinar las alternativas que finalmente serán elegidas o recomendadas a los agentes. Las conclusiones de los argumentos aceptados son identificadas como el conjunto de conclusiones garantizadas.

Definición 3.2.9 (Conclusión garantizada). *Sea $\mathbf{F} = \langle Ag, Args, \triangleright, \blacktriangleright \rangle$ un MEPM. Una conclusión α está garantizada en \mathbf{F} si existe un argumento de preferencia $\mathcal{A} \in Args$ tal que $claim(\mathcal{A}) = \alpha$ y \mathcal{A} es aceptado con respecto a \mathbf{F} .*

La Figura 3.4 y la Figura 3.5, muestran los árboles de dialéctica marcados de los argumentos \mathcal{H}_1 y \mathcal{H}_4 , cuyas conclusiones son $pref(jngla, playa)$ y $pref(playa, jngla)$, res-

pectivamente. Observe que la raíz de $\mathbf{m}(\mathcal{T}(\mathcal{H}_1))$ está marcada **U**, mientras que la raíz de $\mathbf{m}(\mathcal{T}(\mathcal{H}_4))$ está marcada **D**. Esto quiere decir que $pref(jngla, playa)$ es una conclusión garantizada, aunque (por el momento) no se puede decir nada acerca de $pref(playa, jngla)$ ya que $\mathcal{H}_2 \in Args_J$ también posee esta conclusión (ver Figura 3.2), por lo que habría que computar el árbol de dialéctica marcado de este argumento para saber si es aceptado, y así, determinar si $pref(playa, jngla)$ está (o no) garantizada. Sin embargo, como será demostrado en la Sección 3.3, no pueden existir dos argumentos aceptados con conclusiones opuestas, por lo que ya en este punto se podría concluir que $pref(playa, jngla)$ no está garantizada. Note que los argumentos \mathcal{H}_3 y \mathcal{H}_5 poseen la misma conclusión que \mathcal{H}_1 , pero con que exista un sólo argumento aceptado que de soporte a una dada conclusión, es suficiente para dar garantía de la misma.

La etapa de evaluación de argumentas termina dando como resultado el conjunto de conclusiones garantizadas, identificado como W , que se obtiene a partir del MEPM instanciado con los argumentos de preferencia provistos por la etapa de generación de argumentos y los ordenes de importancia y prioridad suministrados por los agentes (ver Figura 3.1).

La Figura 3.6 muestra una captura de pantalla de la aplicación (que será presentada en el Capítulo 5) desarrollada para computar y explicar los resultados que serán obtenidos al utilizar el marco de trabajo propuesto en esta tesis. En esta pantalla, se puede ver el grafo de argumentos del MEPM $\mathbf{F}_J = \langle Ag_J, Args_J, \triangleright_J, \blacktriangleright_J \rangle$ (note que están incluidos todos los argumentos del Ejemplo 2.3.1), siguiendo, en parte, el estilo de la Figura 3.2. En la aplicación, los argumentos aceptados son pintados en verde, y los rechazados en rojo. Además, los identificadores, *DePref*-rules y conclusiones de los argumentos se encuentran ubicados en la esquina inferior derecha de cada uno de ellos. Note que los argumentos \mathcal{H}_1 , \mathcal{H}_2 , \mathcal{H}_3 , \mathcal{H}_4 y \mathcal{H}_5 de la Figura 3.2, están identificados como A8, A13, A30, A33 y A36, respectivamente.

Ejemplo 3.2.1. *Considere el MEPM $\mathbf{F}_J = \langle Ag_J, Args_J, \triangleright_J, \blacktriangleright_J \rangle$, donde $Ag_J = \{c, m, p, t\}$, $Args_J$ es el conjunto de argumentos de preferencias generados a partir del escenario $\langle \mathcal{D}_J, \mathcal{C}_J, \mathcal{V}_J, E_J \rangle$ (ver Ejemplo 2.1.1) y el conjunto de *DePref*-rules \mathcal{R}_J (ver Ejemplo 2.2.4), $\triangleright_J = [\triangleright_c, \triangleright_m, \triangleright_p, \triangleright_t]$, y $\blacktriangleright_J = \{c \blacktriangleright m, c \blacktriangleright p, c \blacktriangleright t, m \blacktriangleright t, p \blacktriangleright t\}$ (ver Ejemplo 3.1.2). El conjunto de conclusiones garantizadas de \mathbf{F}_J es $W_J = \{pref(jngla, playa), pref(jngla, crcro), pref(jngla, lago), pref(jngla, mntñs), pref(playa, crcro), pref(playa, lago), pref(playa, mntñs),$*

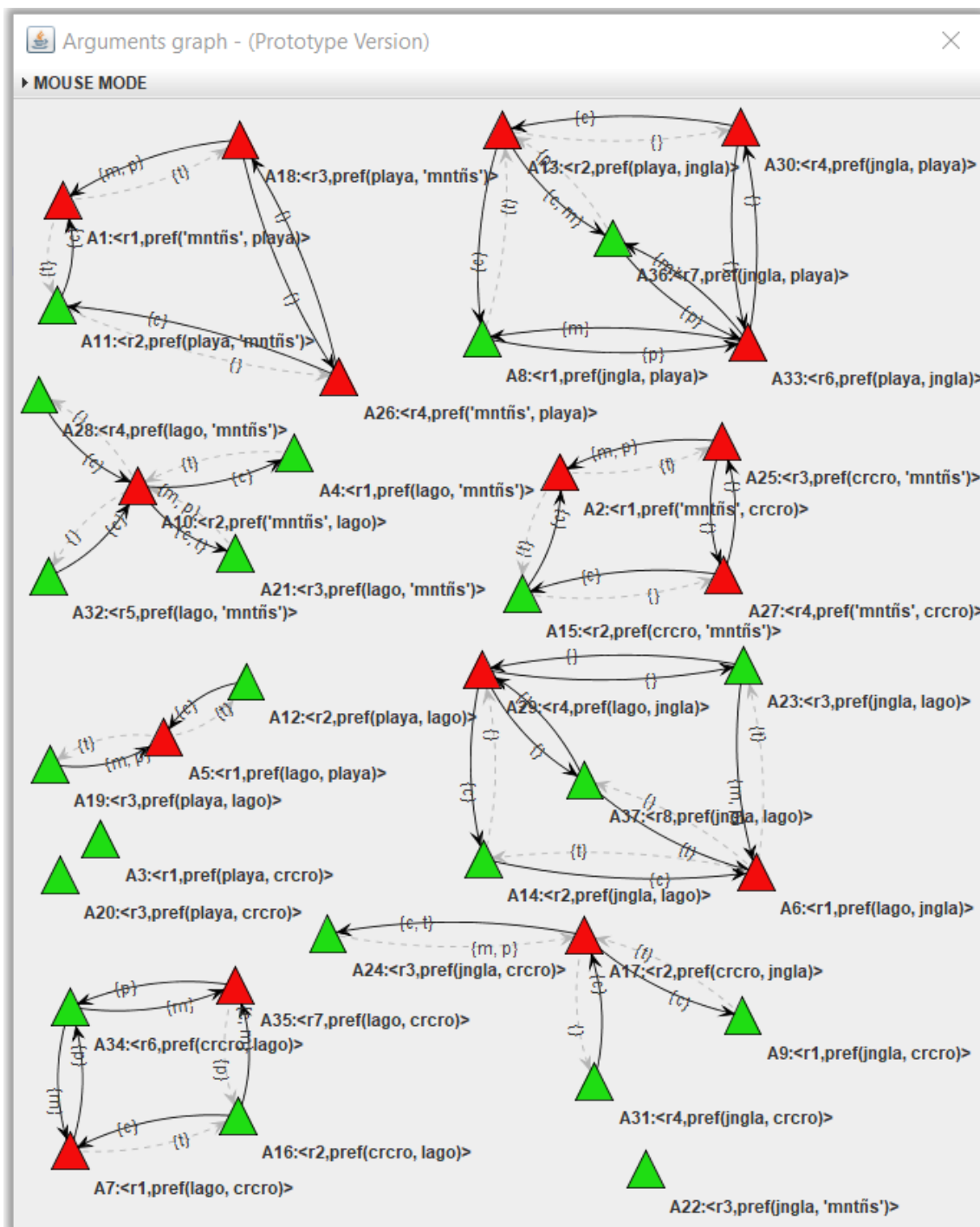


Figura 3.6: Captura de pantalla de la aplicación que será presentada en Capítulo 5 en la que se muestra el grafo de argumentos del MEPM $\mathbf{F}_J = \langle Ag_J, Args_J, \triangleright_J, \blacktriangleright_J \rangle$. Los argumentos aceptados están pintados con verde y los rechazados con rojo.

$\{pref(crcro, lago), pref(crcro, mntñs), pref(lago, mntñs)\}$ y se corresponde con la conclusiones de los argumentos aceptados, mostrados en la Figura 3.6.

En el próximo capítulo, será explicado cómo el marco de trabajo selecciona las alternativas que serán recomendadas al grupo de agentes, a partir del conjunto de conclusiones garantizadas obtenidas al finalizar la etapa de evaluación de argumentos.

3.3. Resultados formales

Para concluir con este capítulo se mostrará, mediante el Teorema 3.3.1 y el Corolario 3.3.1, que el conjunto de conclusiones garantizadas de un MEPM es consistente, esto es, no pueden estar garantizadas dos conclusiones que representen preferencias opuestas. Para poder demostrar esta propiedad, se hará uso de algunas nociones y propiedades auxiliares que serán introducidas a continuación.

Dado un MEPM \mathbf{F} en el que un argumento \mathcal{A} es derrotado por otro argumento \mathcal{B} , se sabe que la secuencia $[\mathcal{A}, \mathcal{B}]$ es una línea de argumentación aceptable en \mathbf{F} . A pesar de que esa línea pueda no ser exhaustiva, se puede asegurar que $[\mathcal{A}, \mathcal{B}]$ será un segmento (ver Definición 3.2.2) de alguna línea de argumentación exhaustiva para \mathcal{A} . Esto significa que uno de los hijos de la raíz del árbol de dialéctica $\mathcal{T}(\mathcal{A})$ será etiquetado con el argumento \mathcal{B} .

Proposición 3.3.1. *Sea \mathbf{F} un MEPM y sean \mathcal{A} y \mathcal{B} dos argumentos tal que \mathcal{B} derrota a \mathcal{A} en \mathbf{F} , la raíz del árbol de dialéctica $\mathcal{T}(\mathcal{A})$ tiene un hijo cuya etiqueta es \mathcal{B} .*

Demostración. Sea \mathbf{F} un MEPM y sean \mathcal{A} y \mathcal{B} dos argumentos tal que \mathcal{B} derrota a \mathcal{A} en \mathbf{F} . Por la Definición 3.2.1, $\lambda = [\mathcal{A}, \mathcal{B}]$ es una línea de argumentación aceptable para \mathcal{A} . Luego, si λ es exhaustiva, de acuerdo con la Definición 3.2.6, λ es una línea de $\mathcal{T}(\mathcal{A})$, y por lo tanto, la raíz de $\mathcal{T}(\mathcal{A})$ tiene un hijo etiquetado con \mathcal{B} . En cambio, si λ no es exhaustiva, debe existir otra línea de argumentación aceptable $\lambda' = [\mathcal{A}, \mathcal{B}, \dots, \mathcal{B}']$ tal que λ' sí es exhaustiva (ver Definición 3.2.5). Luego, como λ' es un línea de $\mathcal{T}(\mathcal{A})$, y \mathcal{B} es el segundo elemento de λ' la raíz de $\mathcal{T}(\mathcal{A})$ debe tener un hijo etiquetado con \mathcal{B} . \square

Dado un árbol de argumentación T y una línea de argumentación λ de T , resulta de interés identificar cuáles son los nodos de T asociados a esta línea. Es decir, cuáles son los nodos dentro del camino (en T) cuya secuencia de etiquetas es igual λ .

Definición 3.3.1 (Nodo asociado). *Sea T un árbol de argumentación y N un nodo dentro de T tal que σ es la secuencia de etiquetas correspondientes a los nodos dentro del camino desde la raíz de T a N . Se dice que N está asociado a una línea de argumentación aceptable λ de T si y sólo si σ es un segmento (inicial) de λ .*

Considere el árbol de dialéctica $\mathcal{T}(\mathcal{H}_1)$ mostrado en la Figura 3.4. Las líneas asociadas al nodo raíz son todas las líneas de ese árbol ya que la secuencia de etiquetas del camino de la raíz hasta la raíz es $[\mathcal{H}_1]$, y este es el segmento inicial de cualquier línea de argumentación para \mathcal{H}_1 . Por su parte, el hijo de la raíz de $\mathcal{T}(\mathcal{H}_1)$ que está etiquetado con \mathcal{H}_4 está asociado a todas las líneas cuyo segmento inicial es $[\mathcal{H}_1, \mathcal{H}_4]$, las cuales son: $[\mathcal{H}_1, \mathcal{H}_4, \mathcal{H}_3]$ y $[\mathcal{H}_1, \mathcal{H}_4, \mathcal{H}_5, \mathcal{H}_2, \mathcal{H}_3]$.

A partir de la definición anterior y la noción de segmento de una línea de argumentación aceptable, es posible definir cuándo un árbol de argumentación es parte (o es un sub-árbol) de otro árbol de argumentación. La idea es que un árbol de argumentación T' será considerado parte de otro árbol de argumentación T cuando haya una porción de la estructura de T que sea igual a T' .

Definición 3.3.2 (Sub-árbol de argumentación). *Sea T un árbol de argumentación para un argumento \mathcal{A} , y sea T' un árbol de argumentación para un argumento \mathcal{B} . Se dice que T' es parte (o es un sub-árbol) de T , denotado $T' \sqsubseteq T$, si y sólo si existe un nodo N en T etiquetado con \mathcal{B} tal que para cada línea de argumentación aceptable λ' de T' , λ' es un segmento de una línea de argumentación λ de T asociada a N .*

Si bien la Definición 3.3.2 está descrita en términos de árboles de argumentación y no de árboles de dialéctica, como todo árbol de dialéctica es un árbol de argumentación en el que todas sus líneas de argumentación aceptables son exhaustivas (ver Definición 3.2.6), esta definición también puede usarse para determinar cuándo un árbol de argumentación es parte de un árbol de dialéctica. Por ejemplo, en la Figura 3.7, el árbol de argumentación T' delimitado con un recuadro de líneas punteadas es parte del árbol de dialéctica $\mathcal{T}(\mathcal{H}_1)$, y también es parte del árbol de dialéctica $\mathcal{T}(\mathcal{H}_4)$. Las líneas de argumentación de T' son $[\mathcal{H}_4, \mathcal{H}_3]$ y $[\mathcal{H}_4, \mathcal{H}_5, \mathcal{H}_2, \mathcal{H}_3]$, y las líneas asociadas al nodo N de $\mathcal{T}(\mathcal{H}_1)$ (ver Figura 3.7) son $[\mathcal{H}_1, \mathcal{H}_4, \mathcal{H}_3]$ y $[\mathcal{H}_1, \mathcal{H}_4, \mathcal{H}_5, \mathcal{H}_2, \mathcal{H}_3]$. Esta claro que todas las líneas de T' son segmentos de las líneas de $\mathcal{T}(\mathcal{H}_1)$ asociadas al nodo N. En cuanto a $\mathcal{T}(\mathcal{H}_4)$, note que las líneas $[\mathcal{H}_4, \mathcal{H}_3]$ y $[\mathcal{H}_4, \mathcal{H}_5, \mathcal{H}_2, \mathcal{H}_3]$ asociadas al nodo M (ver Figura 3.7) son exactamente todas las líneas

de T' , y dado que toda línea de argumentación aceptable es un segmento de sí misma (ver Definición 3.2.2), se tiene que T' es parte de $\mathcal{T}(\mathcal{H}_4)$.

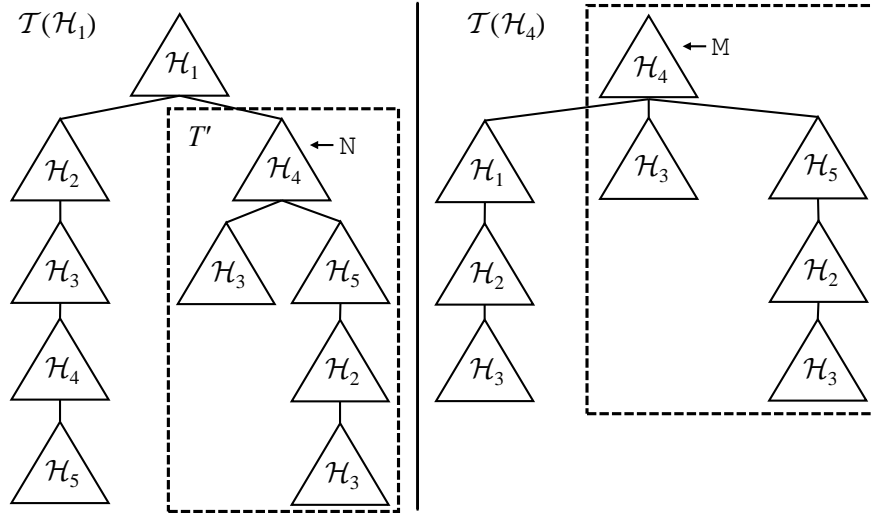


Figura 3.7: Árboles de dialéctica para los argumentos \mathcal{H}_1 y \mathcal{H}_4 del MEPM $\mathbf{F}_{J'}$ de la Figura 3.2. Note que el sub-árbol T' de $\mathcal{T}(\mathcal{H}_1)$ es parte de $\mathcal{T}(\mathcal{H}_4)$.

A continuación, se caracterizará la relación que existe entre el árbol de dialéctica de un argumento y los árboles de dialéctica de sus derrotadores. La Proposición 3.3.2 establece que dado dos argumentos \mathcal{A} y \mathcal{B} , tales que \mathcal{B} derrota a \mathcal{A} , no existe ningún atacante (ni defensor) de \mathcal{B} considerado en $\mathcal{T}(\mathcal{A})$ que no sea considerado en $\mathcal{T}(\mathcal{B})$. En otras palabras, es posible asegurar que el sub-árbol de $\mathcal{T}(\mathcal{A})$ cuya raíz es el hijo de la raíz de $\mathcal{T}(\mathcal{A})$ que está etiquetado con \mathcal{B} (ver Proposición 3.3.1) y que incluye todas las hojas que son alcanzables desde dicho nodo, es parte de $\mathcal{T}(\mathcal{B})$.

Proposición 3.3.2. *Sea \mathbf{F} un MEPM. Sean \mathcal{A} y \mathcal{B} dos argumentos tales que \mathcal{B} derrota a \mathcal{A} en \mathbf{F} . Sea $\mathcal{N}_{\mathcal{B}}$ el hijo de la raíz de $\mathcal{T}(\mathcal{A})$ etiquetado con \mathcal{B} , y sea $T_{\mathcal{B}}$ el sub-árbol de $\mathcal{T}(\mathcal{A})$ cuya raíz es $\mathcal{N}_{\mathcal{B}}$ e incluye todas las hojas que son alcanzables desde $\mathcal{N}_{\mathcal{B}}$. Luego, $T_{\mathcal{B}}$ es parte de $\mathcal{T}(\mathcal{B})$, es decir, $T_{\mathcal{B}} \subseteq \mathcal{T}(\mathcal{B})$.*

Demostración. Sea \mathbf{F} un MEPM. Sean \mathcal{A} y \mathcal{B} dos argumentos tales que \mathcal{B} derrota a \mathcal{A} en \mathbf{F} . Sea $\mathcal{N}_{\mathcal{B}}$ el hijo de la raíz de $\mathcal{T}(\mathcal{A})$ etiquetado con \mathcal{B} (ver Proposición 3.3.1), y sea $T_{\mathcal{B}}$ el sub-árbol de $\mathcal{T}(\mathcal{A})$ cuya raíz es $\mathcal{N}_{\mathcal{B}}$ y que incluye todas las hojas de $\mathcal{T}(\mathcal{A})$ que son alcanzadas desde $\mathcal{N}_{\mathcal{B}}$. En primer lugar, note que $T_{\mathcal{B}}$ es un árbol de argumentación ya que

las secuencia de etiquetas de cada camino desde la raíz a una hoja de $T_{\mathcal{B}}$ es un segmento de una línea de argumentación aceptable de $\mathcal{T}(\mathcal{A})$ (ver Proposición 3.2.1). En segundo lugar, por la Proposición 3.2.1, toda línea de argumentación aceptable λ de $T_{\mathcal{B}}$ es una línea de argumentación aceptable para \mathcal{B} . Luego, si λ es exhaustiva, λ será una de las líneas de $\mathcal{T}(\mathcal{B})$ (ver Definición 3.2.6), y en caso de que λ no sea exhaustiva, será un segmento de una de las líneas $\mathcal{T}(\mathcal{B})$. Además, ya que todas estas líneas comienzan con \mathcal{B} , todas ellas pasan por la raíz de $\mathcal{T}(\mathcal{B})$. Entonces, por Definición 3.3.2, $T_{\mathcal{B}} \sqsubseteq \mathcal{T}(\mathcal{B})$. \square

De forma análoga a cómo los argumentos de soporte e interferencia de una línea de argumentación defienden y atacan (respectivamente) al argumento que se encuentra al principio de la misma, los nodos de un árbol de argumentación T también pueden diferenciarse de acuerdo a si atacan o defienden al argumento asociado a la raíz del árbol.

Definición 3.3.3 (Nodos de soporte e interferencia). *Sea T un árbol de argumentación, N un nodo de T etiquetado con \mathcal{A} , y λ una línea de argumentación aceptable de T asociada a N . Se dice que N es un nodo de soporte (resp. de interferencia) de T si y sólo si $\mathcal{A} \in \mathbf{pros}(\lambda)$ (resp. $\mathcal{A} \in \mathbf{cons}(\lambda)$).*

Por ejemplo, en el árbol de dialéctica $\mathcal{T}(\mathcal{H}_1)$ de la Figura 3.7, los nodos etiquetados con los argumentos \mathcal{H}_1 , \mathcal{H}_3 y \mathcal{H}_5 son nodos de soporte, mientras que los nodos etiquetados con \mathcal{H}_2 y \mathcal{H}_4 son nodos de interferencia de $\mathcal{T}(\mathcal{H}_1)$.

En una línea de argumentación aceptable, que un argumento sea de soporte o interferencia, depende de si se encuentra en una posición par o impar. Esto mismo se traslada a los árboles de argumentación, pero diferenciando los nodos por el nivel en el que se encuentran. Dado una línea de argumentación aceptable $\lambda = [\mathcal{A}_1, \dots, \mathcal{A}_n]$ de un árbol de argumentación T , para todo \mathcal{A}_i ($1 \leq i \leq n$) se cumple que uno de los nodos del nivel i de T debe estar etiquetado con \mathcal{A}_i . Esto significa que si \mathcal{A}_i es un argumento en una posición par (resp. impar) de λ , entonces \mathcal{A}_i estará etiquetando a uno de los nodos de un nivel par (resp. impar) de T .

Proposición 3.3.3. *Sea T un árbol de argumentación, y sea N un nodo de T . Considerando que el nivel de la raíz de T es 1, N es un nodo de soporte (resp. de interferencia) de T si y sólo si N se encuentra en un nivel impar (resp. par) de T .*

Demostración. Sea T un árbol de argumentación, sea N un nodo de T , y considere que el nivel de la raíz de T es 1. Sea $\lambda = [\mathcal{A}_1, \dots, \mathcal{A}_n]$ ($1 \leq n$) una de las líneas de argumentación

aceptable de T que está asociada a \mathbb{N} . De acuerdo con la Definición 3.3.1, existe \mathcal{A}_i en λ ($1 \leq i \leq n$) tal que $\sigma = [\mathcal{A}_1, \dots, \mathcal{A}_i]$ es un segmento de λ y \mathcal{A}_i es la etiqueta de \mathbb{N} . Luego, como σ es la secuencia de etiquetas correspondiente al camino desde la raíz de T hasta \mathbb{N} , y considerando que el nivel de la raíz es 1, i indica el nivel en el que se encuentra \mathbb{N} . Según la Definición 3.2.3, $\mathcal{A}_i \in \text{pros}(\lambda)$ (resp. $\text{cons}(\lambda)$) si y sólo si i es impar (resp. par). Por lo tanto, \mathbb{N} está etiquetado con un nodo de soporte (resp. de interferencia) si y sólo si i es impar (resp. par). Por lo tanto, de acuerdo con la Definición 3.3.3, \mathbb{N} es un nodo de soporte (resp. de interferencia) de T si y solo si i es impar (resp. par). \square

Observe los ejemplos de la Figura 3.8. En ellos se muestra que algunas de las marcas de los nodos de un árbol de argumentación permanecen iguales luego de remover algunos nodos específicos. La siguiente proposición muestra que, en general, dado un árbol de argumentación T , si un nodo de interferencia \mathbb{N} y todos sus descendientes son removidos de T , obteniendo así, un árbol de argumentación T' , entonces todos los ancestros de \mathbb{N} que son nodos de soporte de T y están marcados como \mathbf{U} , también estarán marcados como \mathbf{U} en T' , y todos los ancestros de \mathbb{N} que son nodos de interferencia de T y están marcados como \mathbf{D} , también estarán marcados como \mathbf{D} en T' . Además, esta proposición muestra que si un nodo de soporte \mathbb{N} y todos sus descendientes son removidos de T , obteniendo así, un árbol de argumentación T' , entonces todos los ancestros de \mathbb{N} que son nodos de soporte de T y están marcados como \mathbf{D} , también estarán marcados como \mathbf{D} en T' , y todos los ancestros de \mathbb{N} que son nodos de interferencia de T y están marcados como \mathbf{U} , también estarán marcados como \mathbf{U} en T' . La Proposición 3.3.4 será usada para la demostración del Teorema 3.3.1.

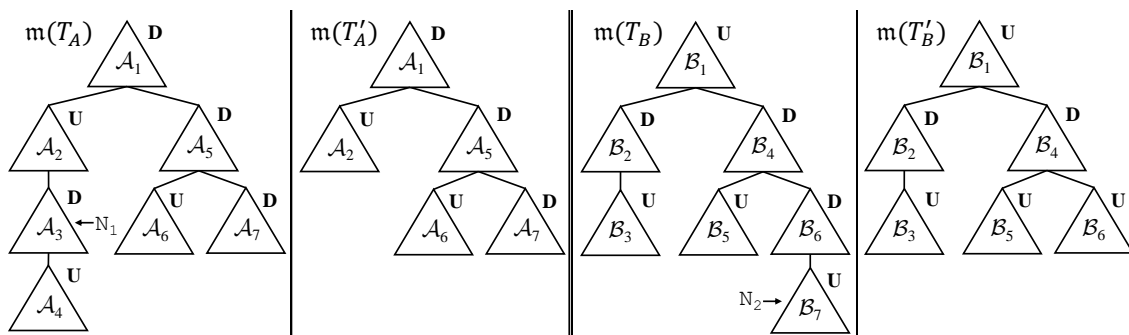


Figura 3.8: Dos pares de árboles de argumentación marcados que muestran dos escenarios en los que las marcas de algunos nodos específicos no cambian luego de remover un nodo junto con todos sus descendientes. T'_A es obtenido a partir de remover el nodo de soporte N_1 y todos sus descendientes de T_A , y T'_B es obtenido a partir de remover el nodo de soporte N_2 y todos sus descendientes de T_B . Como enuncia la Proposición 3.3.4, los ancestros de N_1 que son nodos de interferencia y están marcados como **U** en T_A también serán marcados como **U** en T'_A , y aquellos ancestros de N_1 que son nodos de soporte y están marcados como **D** en T_A también serán marcados como **D** en T'_A . En el otro ejemplo, se puede ver que los ancestros de N_2 que son nodos de interferencia y están marcados como **D** en T_B también serán marcados como **D** en T'_B , y aquellos ancestros de N_2 que son nodos de soporte y están marcados como **U** en T_B también serán marcados como **U** en T'_B .

Proposición 3.3.4. *Sea T un árbol de argumentación, y sea N un nodo de interferencia (resp. un nodo de soporte) de T . Sea T' el árbol de argumentación que se obtiene al remover N y todos sus descendientes de T . Sea M_i el ancestro de N que se encuentra en el nivel i de T ($i > 0$) y sea M'_i el nodo de T' tal que la secuencia de etiquetas correspondientes a los nodos del camino de la raíz de T' a M'_i es la misma que la secuencia de etiquetas correspondientes a los nodos del camino de la raíz de T a M_i . Si M_i es un nodo de soporte (resp. un nodo de interferencia) y está marcado **U** en $m(T)$, M'_i estará marcado **U** en $m(T')$, mientras que si M_i es un nodo de interferencia (resp. un nodo de soporte) y está marcado **D** en $m(T)$, M'_i estará marcado **D** en $m(T')$.*

Demostración. Sea T un árbol de argumentación, y sea N un nodo de interferencia de T . Sea T' el árbol de argumentación que se obtiene al remover N y todos sus descendientes de T . Sea $[M_1, \dots, M_p, N]$ ($p = nivel(N) - 1$) el camino de la raíz (M_1) de T hasta N , siendo M_i el ancestro de N del nivel i ($1 \leq i \leq p$), y sea $[M'_1, \dots, M'_p]$ el camino de T' tal que las etiquetas

de los nodos M_i y M'_i son las mismas. Esta demostración será llevada a cabo utilizando el principio de inducción. Para el caso base, se toma $i = p$, por lo que M_i es el padre de N , identificado como M_p . Por hipótesis, N es un nodo de interferencia, por lo tanto, según la Observación 3.3.3, M_p es un nodo de soporte. Si M_p está marcado **U** en $\mathbf{m}(T)$, por la Definición 3.2.7, todos sus hijos deben estar marcados como **D**. Siendo que ningún otro nodo, a parte de N y sus descendientes, es removido de T para construir T' , a excepción del sub-árbol removido, la estructura debajo de M_p es la misma que la estructura debajo de M'_p . Por lo tanto, todos los hijos de M'_p (si existen) estarán marcados como **U** en $\mathbf{m}(T')$.

Ahora, se probará que para cualquier nodo de interferencia M_{2k} ($1 \leq k < \frac{(p-1)}{2}$) que es marcado como **D** en $\mathbf{m}(T)$, M'_{2k} es marcado como **D** en $\mathbf{m}(T')$. Recuerde que M_{2k} es marcado como **D** en $\mathbf{m}(T)$, M_{2k} tiene al menos un hijo C marcado como **U** en $\mathbf{m}(T)$. Si C no es ancestro de N , entonces M' tiene un hijo C' cuya estructura por debajo es la misma que la de C , por lo que C' es marcado como **U** en $\mathbf{m}(T')$, y por ende, M' es **D** en $\mathbf{m}(T')$. En cambio, si C es ancestro de N entonces $C = M_{2k+1}$. Como M_{2k+1} es un nodo de soporte, si este es marcado como **U** en $\mathbf{m}(T)$, por hipótesis inductiva, M'_{2k+1} es marcado **U** en $\mathbf{m}(T')$. Por lo tanto, M_{2k} es marcado como **D** en $\mathbf{m}(T')$. A continuación, se probará que para todo nodo de soporte M_{2k-1} marcado como **U** en $\mathbf{m}(T)$, M'_{2k-1} es marcado como **U** en $\mathbf{m}(T')$. Si M_{2k-1} marcado como **U** en $\mathbf{m}(T)$, todos los hijos de M_{2k-1} deben estar marcados como **D** en $\mathbf{m}(T)$. Como ningún otro nodo, a parte de N y todos sus descendientes, es removido de T para construir T' , a excepción de los sub-árboles que tienen como raíz a M_{2k} y M'_{2k} , la estructura debajo de M_{2k-1} es la misma que la estructura debajo de M'_{2k-1} . Luego, todos los hijos de M'_{2k-1} , diferentes a M'_{2k} , deben estar marcados como **D** en $\mathbf{m}(T')$. Siendo que M_{2k} es un nodo de interferencia, y está marcado como **D** en $\mathbf{m}(T)$, como fue demostrado arriba, M'_{2k} también está marcado como **D** en $\mathbf{m}(T')$. Por lo tanto, todos los hijos de M'_{2k-1} están marcados como **D** en $\mathbf{m}(T')$, lo que significa que M'_{2k-1} es marcado como **U** en $\mathbf{m}(T')$. Así, queda demostrado que si N es un nodo de interferencia de T para todo ancestro M_i ($1 \leq i \leq p$) de N , si M_i es un nodo de soporte y está marcado como **U** en $\mathbf{m}(T)$, M'_i es marcado como **U** en $\mathbf{m}(T')$, y que si M_i es un nodo de interferencia y está marcado como **D** en $\mathbf{m}(T)$, M'_i es marcado como **D** en $\mathbf{m}(T')$.

La demostración es análoga para el caso en el que N sea un nodo de soporte de T . \square

A continuación será demostrado el Teorema 3.3.1, del mismo se desprende el Corolario 3.3.1 que prueba que el conjunto de conclusiones garantizadas que pueden ser extraídas de un MEPM debe ser consistente, es decir, que no puede darse garantía de una

conclusión α y su complemento $\bar{\alpha}$ al mismo tiempo. Cabe mencionar, que esta propiedad es de gran importancia ya que asegura que el proceso de evaluación dialéctico funciona correctamente.

Teorema 3.3.1. *Sea \mathbf{F} un MEPM. Sean \mathcal{A} y \mathcal{B} dos argumentos de preferencia tales que \mathcal{B} derrota a \mathcal{A} en \mathbf{F} . Si la raíz de $\mathbf{m}(\mathcal{T}(\mathcal{B}))$ está marcada como \mathbf{U} , entonces la raíz de $\mathbf{m}(\mathcal{T}(\mathcal{A}))$ está marcada como \mathbf{D} .*

Demostración. Sea \mathbf{F} un MEPM. Sean \mathcal{A} y \mathcal{B} dos argumentos tales que \mathcal{B} derrota a \mathcal{A} en \mathbf{F} , y considere que la raíz de $\mathbf{m}(\mathcal{T}(\mathcal{B}))$ está marcada como \mathbf{U} . Sean $\mathbf{R}(\mathcal{T}(\mathcal{A}))$ y $\mathbf{R}(\mathcal{T}(\mathcal{B}))$ las raíces de los árboles de dialéctica $\mathcal{T}(\mathcal{A})$ y $\mathcal{T}(\mathcal{B})$, respectivamente. Luego, de acuerdo con la Proposición 3.3.1, se sabe que $\mathbf{R}(\mathcal{T}(\mathcal{A}))$ tiene un hijo $\mathbf{N}_{\mathcal{B}}$ etiquetado con \mathcal{B} (ver Figura 3.9).

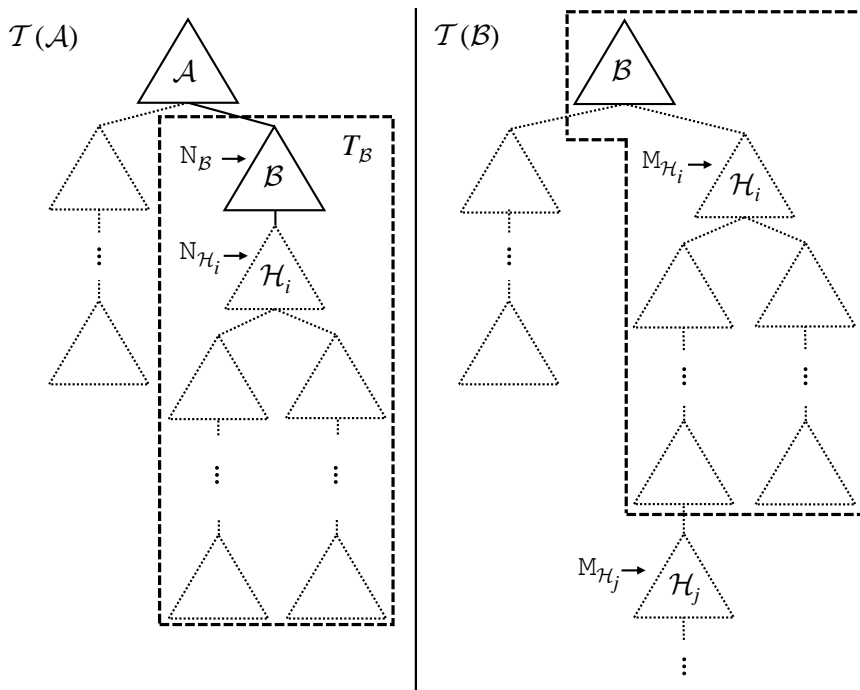


Figura 3.9: Árboles de dialéctica $\mathcal{T}(\mathcal{A})$ y $\mathcal{T}(\mathcal{B})$ mencionados en la demostración del Teorema 3.3.1

En cuanto a $\mathbf{m}(\mathcal{T}(\mathcal{B}))$, según la Definición 3.2.7, si $\mathbf{R}(\mathcal{T}(\mathcal{B}))$ es marcado como \mathbf{U} en $\mathbf{m}(\mathcal{T}(\mathcal{B}))$, es porque $\mathbf{R}(\mathcal{T}(\mathcal{B}))$ es un hoja (lo que significa que \mathcal{B} no tiene derrotadores, ver Proposición 3.3.1), o bien, es porque todos los hijos de $\mathbf{R}(\mathcal{T}(\mathcal{B}))$ (uno por cada derrotador

de \mathcal{B}) están marcados como \mathbf{D} en $\mathbf{m}(\mathcal{T}(\mathcal{B}))$. Si \mathcal{B} no tiene derrotadores, entonces $\mathbf{N}_{\mathcal{B}}$ en $\mathcal{T}(\mathcal{A})$ es una hoja, y por ende $\mathbf{R}(\mathcal{T}(\mathcal{A}))$ es marcado como \mathbf{D} , y la demostración termina. En cambio, si \mathcal{B} tiene derrotadores, por la Proposición 3.3.1, por cada derrotador de \mathcal{B} , existe un hijo de $\mathbf{R}(\mathcal{T}(\mathcal{B}))$ etiquetado con este derrotador. Los argumentos que etiquetan a los hijos de $\mathbf{R}(\mathcal{T}(\mathcal{B}))$, también podrían aparecer como las etiquetas de los hijos de $\mathbf{N}_{\mathcal{B}}$ en $\mathcal{T}(\mathcal{A})$ (ver Figura 3.9). Si $\mathbf{N}_{\mathcal{B}}$ no tiene hijos, entonces $\mathbf{N}_{\mathcal{B}}$ es una hoja, y el análisis es el mismo que el anterior. Entonces, asumiendo que $\mathbf{N}_{\mathcal{B}}$ tiene hijos, si todos ellos son marcados como \mathbf{D} en $\mathbf{m}(\mathcal{T}(\mathcal{A}))$, luego $\mathbf{N}_{\mathcal{B}}$ es marcado \mathbf{U} en $\mathbf{m}(\mathcal{T}(\mathcal{A}))$, y por lo tanto, $\mathbf{R}(\mathcal{T}(\mathcal{B}))$ es marcado como \mathbf{U} . Por el absurdo, suponga que que existe un hijo $\mathbf{N}_{\mathcal{H}_i}$ de $\mathbf{N}_{\mathcal{B}}$ etiquetado con \mathcal{H}_i que es marcado como \mathbf{U} en $\mathbf{m}(\mathcal{T}(\mathcal{A}))$. Sea $T_{\mathcal{B}}$ el sub-árbol de $\mathcal{T}(\mathcal{A})$ cuya raíz es $\mathbf{N}_{\mathcal{B}}$ y que incluye todos los descendientes de $\mathbf{N}_{\mathcal{B}}$, por la Proposición 3.3.2, se sabe que $T_{\mathcal{B}}$ es parte de $\mathcal{T}(\mathcal{B})$ ($T_{\mathcal{B}} \sqsubseteq \mathcal{T}(\mathcal{B})$). Luego, como $\mathbf{N}_{\mathcal{H}_i}$ es hijo $\mathbf{N}_{\mathcal{B}}$, existe un hijo $\mathbf{M}_{\mathcal{H}_i}$ de $\mathbf{R}(\mathcal{T}(\mathcal{B}))$, que también está etiquetado con \mathcal{H}_i . Por hipótesis, $\mathbf{R}(\mathcal{T}(\mathcal{B}))$ está marcado \mathbf{U} en $\mathbf{m}(\mathcal{T}(\mathcal{B}))$, por lo que $\mathbf{M}_{\mathcal{H}_i}$ debe estar marcado como \mathbf{D} en $\mathbf{m}(\mathcal{T}(\mathcal{B}))$. Al haber supuesto que $\mathbf{N}_{\mathcal{H}_i}$ está marcado \mathbf{U} en $\mathbf{m}(\mathcal{T}(\mathcal{A}))$, mientras que $\mathbf{M}_{\mathcal{H}_i}$ está marcado como \mathbf{D} en $\mathbf{m}(\mathcal{T}(\mathcal{B}))$, se puede concluir que los argumentos que fueron considerados a favor y en contra de \mathcal{H}_i en $T_{\mathcal{B}}$ no son los mismos que aquellos que fueron considerados en $\mathcal{T}(\mathcal{B})$. Siendo que $T_{\mathcal{B}} \sqsubseteq \mathcal{T}(\mathcal{B})$, la diferencia en las marcas de $\mathbf{N}_{\mathcal{H}_i}$ y $\mathbf{M}_{\mathcal{H}_i}$ debe ser consecuencia de la existencia de (al menos) un nodo $\mathbf{M}_{\mathcal{H}_j}$ etiquetado con un argumento \mathcal{H}_j tal que ataca (directa o indirectamente) a \mathcal{H}_i en $\mathcal{T}(\mathcal{B})$, y esto no sucede en $\mathcal{T}(\mathcal{A})$ (ver Figura 3.9). Tenga en cuenta que, por la Proposición 3.3.4, se puede asegurar que $\mathbf{M}_{\mathcal{H}_j}$ debe ser un nodo de soporte, ya que de lo contrario, $\mathbf{M}_{\mathcal{H}_j}$ podría ser eliminado de $\mathcal{T}(\mathcal{B})$, y la marca de $\mathbf{M}_{\mathcal{H}_i}$ seguiría siendo \mathbf{D} , lo que está en contradicción con la hipótesis de que $\mathbf{M}_{\mathcal{H}_j}$ es indispensable para que las marcas de $\mathbf{N}_{\mathcal{H}_i}$ y $\mathbf{M}_{\mathcal{H}_i}$ sean diferentes. Luego, la secuencia $\sigma_{\mathcal{B}} = [\mathcal{B}, \mathcal{H}_i, \dots, \mathcal{H}_j]$ (que es un segmento de una línea de argumentación aceptable de $\mathcal{T}(\mathcal{B})$) no puede ser un segmento de ninguna línea de argumentación aceptable de $T_{\mathcal{B}}$, o de lo contrario, \mathcal{H}_j estaría siendo considerado como uno de los atacantes de \mathcal{H}_i en $\mathcal{T}(\mathcal{A})$. Como $\sigma_{\mathcal{B}}$ no es un segmento de ninguna línea de argumentación aceptable de $T_{\mathcal{B}}$ y $T_{\mathcal{B}} \sqsubseteq \mathcal{T}(\mathcal{A})$, $\sigma_{\mathcal{B}}$ no puede ser un segmento de ninguna línea de argumentación aceptable de $\mathcal{T}(\mathcal{A})$. Por lo tanto, la secuencia de etiquetas $\sigma_{\mathcal{A}} = [\mathcal{A}, \mathcal{B}, \mathcal{H}_i, \dots, \mathcal{H}_j]$ (que se obtiene al agregar \mathcal{A} al principio de $\sigma_{\mathcal{B}}$) no puede ser una línea de argumentación aceptable para \mathcal{A} en \mathbf{F} . Por hipótesis, se sabe que \mathcal{B} derrota a \mathcal{A} , por lo que, de acuerdo con la Definición 3.2.1, si $\sigma_{\mathcal{A}}$ no es una línea de argumentación aceptable, mientras que $\sigma_{\mathcal{B}}$ sí, es porque existe un argumento que aparece dos veces en $\sigma_{\mathcal{A}}$, pero solo una vez en $\sigma_{\mathcal{B}}$. Siendo que el único argumento que incorpora $\sigma_{\mathcal{A}}$ con respecto a

$\sigma_{\mathcal{B}}$ es \mathcal{A} , se puede concluir que el argumento que se repite es \mathcal{A} , y que $\mathcal{H}_j = \mathcal{A}$ ya que \mathcal{H}_j es el primer argumento que es considerado en $\mathcal{T}(\mathcal{B})$, pero no en $\mathcal{T}(\mathcal{A})$. Sin embargo, \mathcal{H}_j ataca (directa o indirectamente), y por ende, $\mathcal{H}_j \in \text{pros}(\sigma_{\mathcal{B}})$, osea, $\mathcal{A} \in \text{pros}(\sigma_{\mathcal{B}})$. Esto es una contradicción, ya que por la Observación 3.2.1 se llega a que $\text{claim}(\mathcal{A}) = \text{claim}(\mathcal{B})$, mientras que, por hipótesis, se tiene que \mathcal{B} derrota a \mathcal{A} , lo que implica que sus conclusiones deben ser opuestas (ver Definición 3.1.6). Este absurdo provino de suponer que $N_{\mathcal{H}_i}$ estaba marcado como **U** y que por lo tanto, existía un derrotador de \mathcal{H}_i que es considerado en $\mathcal{T}(\mathcal{B})$, pero no en $\mathcal{T}(\mathcal{A})$. Luego, la marca de $N_{\mathcal{H}_i}$ en $\mathbf{m}(\mathcal{T}(\mathcal{A}))$ (al igual que la de $M_{\mathcal{H}_i}$ en $\mathbf{m}(\mathcal{T}(\mathcal{B}))$) es **D**. Por lo tanto, como todos los hijos de $N_{\mathcal{B}}$ están marcados como **D**, $N_{\mathcal{B}}$ está marcado como **U**, y esto significa que $\mathbf{R}(\mathcal{T}(\mathcal{A}))$ está marcado como **D** en $\mathbf{m}(\mathcal{T}(\mathcal{A}))$. \square

Corolario 3.3.1. *Sea \mathbf{F} un MEPM. Si α está garantizada en \mathbf{F} entonces $\bar{\alpha}$ no está garantizada en \mathbf{F} .*

Demostración. Sea $\mathbf{F} = \langle Ag, Args, \triangleright, \blacktriangleright \rangle$ un MEPM tal que α está garantizada en \mathbf{F} . Luego, de acuerdo con la Definición 3.2.9, existe un argumento $\mathcal{B} \in Args$ tal que $\text{claim}(\mathcal{B}) = \alpha$ y la raíz de $\mathbf{m}(\mathcal{T}(\mathcal{B}))$ está marcada como **U**. Por el absurdo, suponga que $\bar{\alpha}$ está garantizada, y por lo tanto, existe un argumento $\mathcal{A} \in Args$ tal que $\text{claim}(\mathcal{A}) = \bar{\alpha}$ y la raíz de $\mathbf{m}(\mathcal{T}(\mathcal{A}))$ está marcada como **U**. De acuerdo con la Definición 3.1.1, \mathcal{A} y \mathcal{B} están en conflicto. Luego, debe ocurrir que \mathcal{B} derrote a \mathcal{A} en \mathbf{F} , o que \mathcal{A} derrote a \mathcal{B} (pueden ser ambas). Si \mathcal{B} derrota a \mathcal{A} en \mathbf{F} , como la raíz de $\mathbf{m}(\mathcal{T}(\mathcal{B}))$ está marcada como **U**, por el Teorema 3.3.1, la raíz de $\mathbf{m}(\mathcal{T}(\mathcal{A}))$ debe estar marcada como **D**. Sin embargo, esto contradice la suposición de que la raíz de $\mathbf{m}(\mathcal{T}(\mathcal{A}))$ estaba marcada como **U**. En cambio, si \mathcal{A} derrota a \mathcal{B} , como se asume que \mathcal{A} es un argumento aceptado, y por ende, la raíz de $\mathbf{m}(\mathcal{T}(\mathcal{A}))$ debe estar marcada como **U**, por el Teorema 3.3.1, se puede asegurar que la raíz de $\mathbf{m}(\mathcal{T}(\mathcal{B}))$ está marcada como **D**. Sin embargo, esto contradice la hipótesis de que la raíz de $\mathbf{m}(\mathcal{T}(\mathcal{B}))$ estaba marcada como **U**. \square

3.4. Resumen

En este capítulo se explicaron los procesos que se llevan a cabo durante la etapa de evaluación de argumentos, y cuáles son los elementos que intervienen en dichos procesos.

Como fue explicado a lo largo del capítulo, en esta etapa del proceso de decisión se evalúan los argumentos de preferencia (*Args*) que son generados en la etapa de genera-

ción de argumentos. Para cumplir con este objetivo, en primer lugar, se identifican los conflictos y derrotas entre los argumentos. La relación de conflicto entre los argumentos determina cuáles de ellos poseen conclusiones opuestas o contradictorias, es decir, cuando la conclusión de un argumento es $pref(d_x, d_y)$ y la del otro es $pref(d_y, d_x)$. Luego, en base a los órdenes de importancia entre reglas (\triangleright) y el orden de prioridad (\blacktriangleright) entre los agentes (Ag), se puede determinar cuándo un argumento es más fuerte que otro, y con esto, se determinan las derrotas entre los argumentos que están en conflicto. Como es indicado por la Definición 3.1.5, la fortaleza de cada argumento depende de la *DePref*-rule con la que fue creado, de la importancia que le dan los agentes a esta *DePref*-rule, y de la prioridad entre los agentes. Las relaciones entre los argumentos pueden ser visualizadas mediante un grafo de argumentos (ver Definición 3.1.7) como el que se muestra en la Figura 3.2. Además, este grafo permite dilucidar por qué un argumento derrota (o no) a otro.

Una vez identificadas las derrotas entre los argumentos, se ejecuta el proceso de evaluación dialéctico, mediante el cuál se determinan los argumentos aceptados. Para esto, por cada uno de los argumentos se construye su árbol de dialéctica y se marcan sus nodos siguiendo el procedimiento de marcado \mathbf{m} (ver Definición 3.2.7). Dado un argumento \mathcal{A} , el árbol de dialéctica de \mathcal{A} , denotado $\mathcal{T}(\mathcal{A})$, agrupa en una estructura arbórea que sitúa a \mathcal{A} en la raíz, y debajo de este (organizados por niveles) a todos los argumentos que lo atacan y defienden. Luego, el procedimiento de marcado determina, según la estructura del árbol, cuáles de estos argumentos son derrotados, marcándolos como \mathbf{D} , y cuáles no, marcándolos como \mathbf{U} . Si la raíz de un árbol de dialéctica es marcada como \mathbf{U} , esto indica que el argumento en la raíz es defendido de sus derrotadores, lo que significa que debe ser aceptado. En cambio, si la raíz del árbol de dialéctica es marcada como \mathbf{D} , el argumento en su raíz es rechazado. El Algoritmo 2 especifica cómo construir árboles de dialéctica marcados. Finalmente a las conclusiones de los argumentos aceptados se las considera garantizadas, y con ellas se podrá construir un orden de preferencia entre las alternativas, como será explicado en el siguiente capítulo.

Adicionalmente, el capítulo actual cuenta con una sección de resultados formales en la que se demuestra la consistencia del conjunto de conclusiones garantizadas. Esto significa, que el proceso de evaluación dialéctico no puede dar por aceptados dos argumentos de preferencia cuyas conclusiones representen preferencias opuestas, es decir, dos argumentos que estén en conflicto, cuyas conclusiones sean $pref(d_x, d_y)$ y $pref(d_y, d_x)$. Esta propiedad, asegura que el proceso de evaluación dialéctico funciona correctamente.

Capítulo 4

Selección de alternativas y generación de explicaciones

Como fue explicado en el capítulo anterior, durante la etapa de evaluación de argumentos, se determina cuáles son los argumentos aceptados del conjunto de argumentos (*Args*) construidos durante la etapa de generación de argumentos, teniendo en cuenta los intereses individuales (\triangleright) de los agentes (*Ag*), y la prioridad entre los mismos (\blacktriangleright), con el objetivo de obtener el conjunto de conclusiones garantizadas.

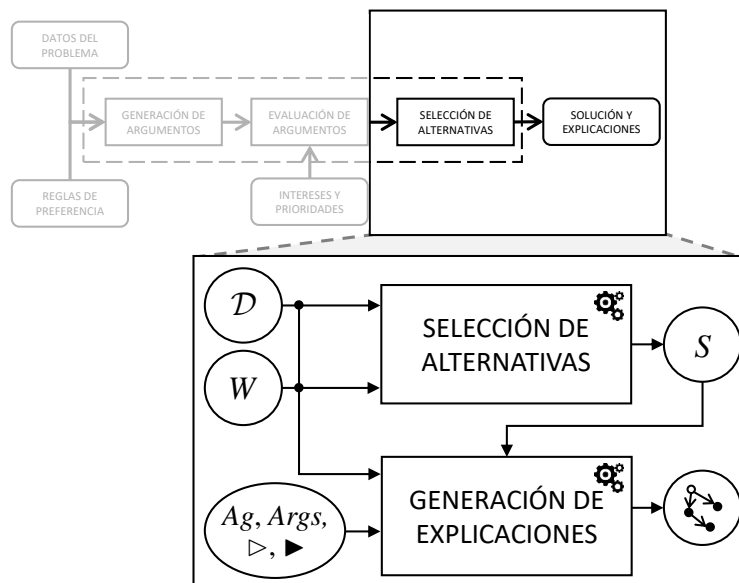


Figura 4.1: Detalle de la etapa de selección de alternativas.

En esta etapa se ejecutan dos procesos: el de selección de alternativas, y el de generación de explicaciones (ver Figura 4.1). El primero, determina el conjunto de alternativas (S) que serán recomendadas al grupo de agentes a partir de las alternativas disponibles (\mathcal{D}) y de las conclusiones garantizadas (W). El segundo proceso, tiene como propósito brindar las explicaciones pertinentes acerca de los resultados obtenidos. Como se verá más adelante, las explicaciones propuestas permitirán, a demanda, obtener los detalles de por qué una alternativa es (o no) elegida.

4.1. Proceso de selección de alternativas

El proceso de selección de alternativas basa su funcionamiento en la relación de preferencia, generada a partir de la clausura transitiva del conjunto de conclusiones garantizadas. Dado un conjunto de conclusiones garantizadas W de un MEPM, la clausura transitiva de W , denotada W^* , es el conjunto de conclusiones tal que $W \subseteq W^*$ y si $pref(d_x, d_y) \in W^*$ y $pref(d_y, d_z) \in W^*$, entonces $pref(d_x, d_z) \in W^*$. A continuación, se muestra un ejemplo para clarificar este concepto.

Ejemplo 4.1.1. Sea $\mathcal{D} = \{d_1, d_2, d_3, d_4, d_5\}$ un conjunto de alternativas y $W = \{pref(d_1, d_3), pref(d_2, d_3), pref(d_3, d_4), pref(d_4, d_5), pref(d_5, d_3)\}$ un conjunto de conclusiones garantizadas. La clausura transitiva de W es $W^* = W \cup \{pref(d_1, d_4), pref(d_1, d_5), pref(d_2, d_4), pref(d_2, d_5), pref(d_3, d_5), pref(d_4, d_3), pref(d_5, d_4), pref(d_3, d_3), pref(d_4, d_4), pref(d_5, d_5)\}$.

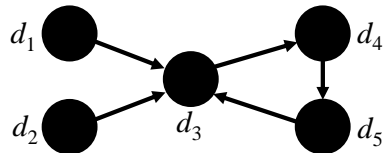


Figura 4.2: Grafo dirigido representando la relación entre los pares de alternativas del conjunto W del Ejemplo 4.1.1.

La Figura 4.2 muestra un grafo dirigido que representa la relación W del Ejemplo 4.1.1. Los nodos de este grafo representan a las alternativas, y si existe un arco de una alternativa X a otra Y es porque $pref(X, Y) \in W$. Note que, $pref(d_1, d_4) \in W^*$ ya que $pref(d_1, d_3)$ y $pref(d_3, d_4) \in W$ y $W \subset W^*$, así como también, $pref(d_1, d_5) \in W^*$ ya que $pref(d_1, d_4) \in$

W^* (aunque $\text{pref}(d_1, d_4) \notin W$) y $\text{pref}(d_4, d_5) \in W^*$. Un razonamiento análogo puede usarse para explicar por qué $\{\text{pref}(d_2, d_4), \text{pref}(d_2, d_5)\} \subset W^*$. Por último, en cuanto a $\text{pref}(d_3, d_5)$, $\text{pref}(d_4, d_3)$, $\text{pref}(d_5, d_3)$, $\text{pref}(d_3, d_3)$, $\text{pref}(d_4, d_4)$ y $\text{pref}(d_5, d_5)$, estas pertenecen a W^* porque $\{\text{pref}(d_3, d_4), \text{pref}(d_4, d_5), \text{pref}(d_5, d_3)\} \subset W$ y $W \subset W^*$.

Claramente, para algunos conjuntos W , puede ocurrir que $W = W^*$. De hecho, esto sucede con el conjunto de conclusiones garantizadas $W_J = \{\text{pref}(jngla, playa), \text{pref}(jngla, crcro), \text{pref}(jngla, lago), \text{pref}(jngla, mntñs), \text{pref}(playa, crcro), \text{pref}(playa, lago), \text{pref}(playa, mntñs), \text{pref}(crcro, lago), \text{pref}(crcro, mntñs), \text{pref}(lago, mntñs)\}$ (ver Ejemplo 3.2.1), donde $W_J = W_J^*$. En base a la clausura transitiva del conjunto de conclusiones garantizadas (W) generado durante la etapa de evaluación de argumentos, es que se podrá definir una relación de preferencias entre las alternativas (\mathcal{D}).

Definición 4.1.1 (Relación de preferencia). *Sea \mathcal{D} un conjunto de alternativas tal que $\{d_x, d_y\} \subseteq \mathcal{D}$, y sea W^* la clausura transitiva de un conjunto de conclusiones garantizadas de un MEPM. Se dice que d_x es al menos tan preferida como d_y de acuerdo con W^* , denotado $d_x \succeq_{W^*} d_y$, si y sólo si $\text{pref}(d_x, d_y) \in W^*$. Luego, cuando $d_x \succeq_{W^*} d_y$ y $d_y \not\prec_{W^*} d_x$, se dice que d_x es estrictamente preferida a d_y ($d_x \succ_{W^*} d_y$); cuando $d_x \succeq_{W^*} d_y$ y $d_y \succeq_{W^*} d_x$, se dice que d_x y d_y son indiferentes ($d_x \sim_{W^*} d_y$); y cuando $d_x \not\prec_{W^*} d_y$ y $d_y \not\prec_{W^*} d_x$ se dice que d_x y d_y son incomparables.*

De acuerdo con el Ejemplo 4.1.1, d_1 y d_2 son incomparables, d_1 y d_2 son ambas estrictamente mejores que d_3 , d_4 y d_5 , y estas últimas tres son equivalentes. A partir de la relación de preferencia estricta \succ_{W^*} , es posible identificar los elementos maximales de la relación, es decir, las mejores alternativas, y por ende, aquellas que serán seleccionadas y presentadas ante el grupo de agentes.

Definición 4.1.2 (Alternativas seleccionadas). *Sea \mathcal{D} un conjunto de alternativas, y sea W^* la clausura transitiva de un conjunto de conclusiones garantizadas de un MEPM. El conjunto de alternativas seleccionadas a partir de \mathcal{D} y W^* es $S = \{d_x \in \mathcal{D} \mid \nexists d_y \in \mathcal{D} \text{ tal que } d_y \succ_{W^*} d_x\}$.*

Note que las mejores alternativas son aquellas para las cuales no existe otra alternativa que sea estrictamente preferida a estas. Siguiendo con el ejemplo anterior, al considerar

las alternativas y la relación de preferencia mostrada en la Figura 4.2, el conjunto de alternativas seleccionadas que se obtiene es $\{d_1, d_2\}$. A continuación, se retoma nuevamente el ejemplo de la familia Jota.

Ejemplo 4.1.2. *Considere el conjunto de alternativas \mathcal{D}_J del escenario de elección del Ejemplo 2.1.1, y el conjunto de conclusiones garantizadas W_J del Ejemplo 3.2.1. El conjunto de alternativas seleccionadas es $S_J = \{jnгла\}$. El grafo dirigido de la Figura 4.3 representa la relación de preferencia $\succeq_{W_J^*}$ que se obtiene a partir de W_J^* . Recuerde que, en este caso particular, $W_J = W_J^*$.*

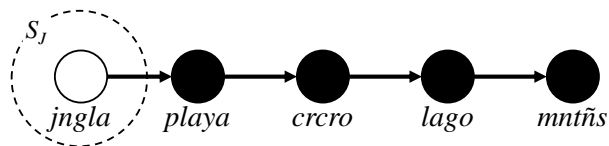


Figura 4.3: Grafo dirigido representando la relación de preferencia $\succeq_{W_J^*}$, obtenida a partir del conjunto de conclusiones garantizadas W_J del Ejemplo 3.2.1. Cada nodo representa una alternativa, y si existe un arco de una alternativa X a otra Y es porque $X \succeq_{W_J^*} Y$. La alternativa seleccionada ($jnгла$) está pintada de color blanco y rodeada por un círculo de líneas punteadas.

El Algoritmo 3 es el que implementa el proceso de selección de alternativas. En principio, el conjunto de alternativas seleccionadas es igual al conjunto de alternativas ($S \leftarrow \mathcal{D}$).

Algoritmo 3 Proceso de selección de alternativas.

Entrada: Un conjunto de alternativas \mathcal{D} y la clausura transitiva W^* de un conjunto de conclusiones garantizadas de un MEPM.

Salida: El conjunto de alternativas seleccionadas S .

$S \leftarrow \mathcal{D}$

for each $d_x \in \mathcal{D}$ **do**

if $\exists d_y \in \mathcal{D}$ **such that** $d_y \succ_{W^*} d_x$ **then**

$S \leftarrow S \setminus \{d_x\}$

end if

end for

return S

Luego, para cada alternativa d_x en \mathcal{D} se verifica si existe una alternativa d_y en \mathcal{D} que sea estrictamente preferida a d_x . En caso de que d_y exista, se remueve d_x del conjunto de alternativas seleccionadas ($S \leftarrow S \setminus \{d_x\}$). Finalmente, el algoritmo termina retornando el conjunto de alternativas que no fueron descartadas (S).

Del ejemplo de la familia Jota, se puede observar que la relación que se podría inducir directamente de las conclusiones de W_J es la misma que $\succeq_{W_J^*}$, esto es porque $W_J = W_J^*$. Sin embargo, si el conjunto de conclusiones garantizadas fuese como el del Ejemplo 4.1.1, esta claro que no sería lo mismo computar la relación de preferencia directamente a partir de W ya que W^* posee muchas más información. Note que W^* completa las conclusiones que pueden deducirse por transitividad, lo que también produce que la relación entre las alternativas dentro de un ciclo sea simétrica, y por lo tanto, sean consideradas equivalentes.

El siguiente ejemplo describe un escenario en el que se propone utilizar el marco de trabajo propuesto en esta tesis para modelar el comportamiento de un robot capaz de apagar incendios. Dicho robot, ante la presencia de diferentes focos ígneos, deberá decidir cuál apagar primero. Los focos pueden variar en intensidad, en el tamaño del área que ocupan, encontrarse en distintas posiciones, o requerir de atención urgente para evitar daños mayores o peligros inminentes. El objetivo principal de este ejemplo, es demostrar que es posible obtener un conjunto de conclusiones garantizadas que sea diferente a su clausura transitiva. Además, este ejemplo, muestra la versatilidad del marco de trabajo para modelar dominios de diferentes características, y que es posible utilizarlo tanto en ambientes multi-agentes, como en ambientes con un sólo agente.

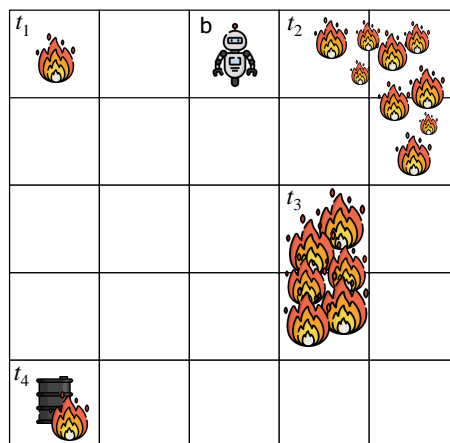


Figura 4.4: Representación del escenario de elección del Ejemplo 4.1.3.

Ejemplo 4.1.3. La Figura 4.4 muestra un posible escenario en el que se requiere que el robot decida cual incendio apagar en primer lugar. En esta figura se puede observar la posición del robot, identificado como \mathbf{b} , y el conjunto de tareas (incendios) $\mathcal{D}_B = \{t_1, t_2, t_3, t_4\}$ que el robot debe realizar (apagar). Si bien el robot debería realizar todas las tareas, solo puede con una a la vez, por lo que tiene que decidir con cuál empezar. Cada una de estas tareas puede ser evaluada según el área incendiada, la intensidad del fuego, la distancia a la que se encuentra del robot, y si es urgente, o no. Estos criterios son representados por el conjunto $\mathcal{C}_B = \{area, int, dst, urg\}$, para los cuales se tiene que $\mathcal{V}_B(area)$ es el rango de enteros positivos ordenados de menor a mayor (cuánto menor es el área es mejor porque el incendio es más fácil de apagar), $\mathcal{V}_B(int)$ es el rango [alta, media, baja] cuyos valores están ordenados de izquierda a derecha de peor a mejor (a menor intensidad es más fácil apagar el fuego), $\mathcal{V}_B(dst)$ es el rango de enteros positivos ordenados de menor a mayor (cuánto más cerca esté el fuego, mejor) y $\mathcal{V}_B(urg)$ es el rango [no, si] tal que si $>_{urg}$ no (es preferible atender urgencias primero). De esta manera, el escenario de elección que modela el problema del Ejemplo 4.1.3, quedaría representado por $\langle \mathcal{D}_B, \mathcal{C}_B, \mathcal{V}_B, E_B \rangle$, donde

$$E_B = \left\{ \begin{array}{cccc} area(t_1, 1), & int(t_1, baja), & dst(t_1, 2), & urg(t_1, no), \\ area(t_2, 3), & int(t_2, media), & dst(t_2, 1), & urg(t_2, no), \\ area(t_3, 2), & int(t_3, alta), & dst(t_3, 3), & urg(t_3, no), \\ area(t_4, 1), & int(t_4, baja), & dst(t_4, 6), & urg(t_4, si) \end{array} \right\}$$

Las preferencias del robot sobre el orden en el que las tareas deben ser atendidas, son modeladas por el conjunto de DePref-rules $\mathcal{R}_B = \{r_9, r_{10}, r_{11}, r_{12}, r_{13}, r_{14}, r_{15}, r_{16}\}$ y el orden de importancia $\triangleright_{\mathbf{b}}$. A continuación, se muestran todas las DePref-rules antes mencionadas y los pares que modelan la relación de importancia entre las mismas, acompañadas de una explicación:

$$r_9 = \left\{ better(X, Y, area) \right\} \Rightarrow pref(X, Y)$$

$$r_{10} = \left\{ better(X, Y, int) \right\} \Rightarrow pref(X, Y)$$

Tanto si el área, como si la intensidad de una alternativa X son menores a la de otra alternativa Y , entonces existe una razón para preferir a X por sobre Y . En general, el área y la intensidad son consideradas igual de importantes.

$$r_{11} = \left\{ better(X, Y, area), min_dist(X, Y, area, 2) \right\} \Rightarrow pref(X, Y)$$

Si el área de una alternativa X es menor que la de otra alternativa Y , y además, la diferencia entre el área de X y Y es de al menos dos, entonces, existe una razón (más fuerte que la anterior) para preferir a X por sobre Y . En este caso el área será considerada más importante que la intensidad ($r_{11} \triangleright_b r_{10}$).

$$r_{12} = \left\{ \text{better}(X, Y, \text{dst}) \right\} \Rightarrow \text{pref}(X, Y)$$

Si la distancia a una alternativa X es menor que a otra alternativa, entonces existe una razón para preferir a X por sobre Y . En general, la distancia es más importante que el área y que la intensidad ($r_{12} \triangleright_b r_9$, $r_{12} \triangleright_b r_{10}$), incluso por más que la diferencia entre el área de X y la de Y sea mayor o igual a dos ($r_{12} \triangleright_b r_{11}$).

$$r_{13} = \left\{ \text{better}(X, Y, \text{urg}) \right\} \Rightarrow \text{pref}(X, Y)$$

Si la X es urgente mientras que Y no, entonces existe una razón para preferir a X por sobre Y . En general, la urgencia es más importante que todos los demás criterios ($r_{13} \triangleright_b r_9$, $r_{13} \triangleright_b r_{10}$, $r_{13} \triangleright_b r_{12}$), incluso por más que la diferencia que entre el área de X y la de Y sea mayor o igual a dos ($r_{13} \triangleright_b r_{11}$).

$$r_{14} = \left\{ \begin{array}{l} \text{worse}(X, Y, \text{dst}), \text{max_dist}(X, Y, \text{dst}, 3), \\ \text{better}(X, Y, \text{area}), \text{better}(X, Y, \text{int}) \end{array} \right\} \Rightarrow \text{pref}(X, Y)$$

A pesar de que la distancia hasta una alternativa X sea mayor que hasta otra alternativa Y , mientras la diferencias entre las distancias no sea mayor a tres, si se cumple que el área y la intensidad de X sean menores a la de Y , entonces existe una razón para preferir a X por sobre Y . En este caso, la importancia pierde importancia frente al área y la intensidad ($r_{14} \triangleright_b r_{12}$).

$$r_{15} = \left\{ \text{better}(X, Y, \text{dst}), \text{min_dist}(X, Y, \text{dst}, 5) \right\} \Rightarrow \text{pref}(X, Y)$$

Si la distancia hasta una alternativa X es menor que hasta otra alternativa Y , y además, la diferencia entre entre las distancias es de al menos cinco puntos, entonces existe una razón para preferir a X por sobre Y . En este caso, la distancia es considerada más importante que todos los demás criterios ($r_{15} \triangleright_b r_{14}$, $r_{15} \triangleright_b r_{13}$, $r_{15} \triangleright_b r_{11}$, $r_{15} \triangleright_b r_{10}$, $r_{15} \triangleright_b r_9$).

$$r_{16} = \left\{ \begin{array}{l} \text{better}(X, Y, \text{area}), \text{better}(X, Y, \text{int}), \\ \text{better}(X, Y, \text{dst}), \text{min_dist}(X, Y, \text{dst}, 2) \end{array} \right\} \Rightarrow \text{pref}(X, Y)$$

Si el área, la intensidad y la distancia de una alternativa X son menores que la de otra alternativa Y , y además, la diferencia entre entre las distancias es de al menos dos puntos, entonces existe una razón para preferir a X por sobre Y . En este caso, la urgencia pierde relevancia frente a todos los demás criterios ($r_{16} \triangleright_b r_{13}$).

La Figura 4.5 muestra una captura de pantalla de la aplicación que fue desarrollada en el contexto de esta tesis, y que será presentada con más detalles en el Capítulo 5. En esta pantalla se puede ver el grafo de argumentos del MEPM $\mathbf{F}_J = \langle Ag_B, Args_B, \triangleright_B, \blacktriangleright_B \rangle$, donde $Ag = \{b\}$, $Args_B$ son los argumentos de preferencia generados a partir de $\langle \mathcal{D}_B, \mathcal{C}_B, \mathcal{V}_B, E_B \rangle$ y \mathcal{R}_B , $\triangleright_B = [\triangleright_b]$ y $\blacktriangleright_B = \emptyset$. La aplicación distingue los argumentos aceptados con color verde, y los rechazados con color rojo. De aquí, se puede observar que $W_B = \{pref(t_1, t_2), pref(t_1, t_3), pref(t_2, t_4), pref(t_2, t_3), pref(t_4, t_1), pref(t_4, t_3)\}$.

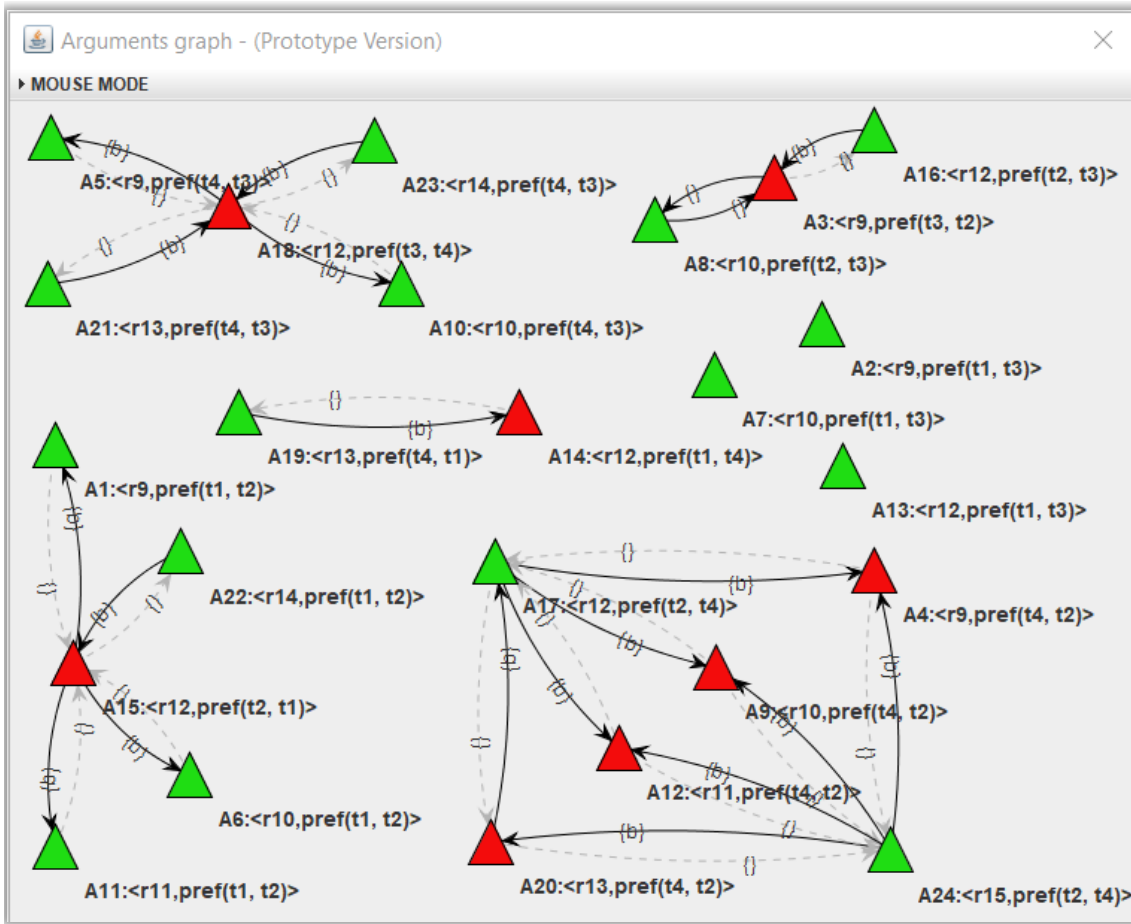


Figura 4.5: Argumentos aceptados y rechazados del MEPM \mathbf{F}_B .

La Figura 4.6 muestra la relación de preferencia $\succeq_{W_B^*}$ entre las alternativas en \mathcal{D}_B . Se puede ver que $t_1 \sim_{W_B^*} t_2 \sim_{W_B^*} t_4 \succeq_{W_B^*} t_3$. Esto significa que resulta indiferente que el robot realice primero la tarea t_1 , t_2 o t_4 , ya que $S_B = \{t_1, t_2, t_4\}$. Note que las relaciones establecidas por las conclusiones de W_B , no son las mismas que las establecidas por $\succeq_{W_B^*}$.

En particular, se tiene que $t_1 \succeq_{W_B^*} t_4$, $t_2 \succeq_{W_B^*} t_1$ y $t_4 \succeq_{W_B^*} t_2$, mientras que $\text{pref}(t_1, t_4)$, $\text{pref}(t_2, t_1)$ y $\text{pref}(t_4, t_2)$ no pertenecen a W_B .

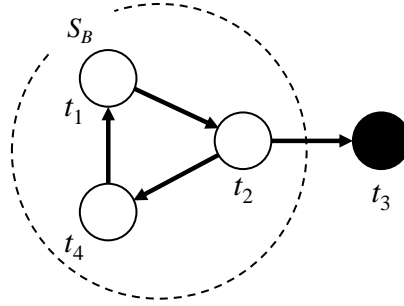


Figura 4.6: El conjunto de alternativas seleccionadas S_B y la relación de preferencia $\succeq_{W_B^*}$ entre las alternativas en \mathcal{D}_B obtenida a partir de las conclusiones garantizadas W_B . No todos los pares de la relación de preferencia son mostrados, aunque los pares restantes pueden inferirse por transitividad.

El ejemplo anterior muestra que es posible obtener un conjunto de conclusiones garantizadas que induzcan una relación de preferencia cíclica entre las alternativas. La política que se ha tomado ante esta situación es considerar que todas las alternativas dentro del ciclo son igual de buenas. Es por esta razón, que se utiliza la clausura transitiva de W para obtener el conjunto de alternativas seleccionadas.

4.2. Resultados formales

En esta sección se presenta un conjunto de proposiciones que permiten caracterizar el comportamiento general del proceso de selección, por lo tanto, los resultados del marco de trabajo. La siguiente proposición, muestra que todas las alternativas dentro de un ciclo serán consideradas indiferentes, y que si son indiferentes es porque pertenecen a un mismo ciclo.

Proposición 4.2.1. *Sea \mathcal{D} un conjunto de alternativas, y sea W el conjunto de conclusiones garantizadas de un MEPM. Para cualquier par de alternativas $d_x, d_y \in \mathcal{D}$ se cumple que $d_x \sim_{W^*} d_y$ si y sólo si existe una secuencia de alternativas $C = [d_1, \dots, d_m, \dots, d_n]$ ($1 < m < n$) tal que C es un ciclo que comienza y termina en d_x ($d_1 = d_n = d_x$), que pasa por d_y ($d_m = d_y$), y tal que $\text{pref}(d_i, d_{i+1}) \in W$, con $1 \leq i < n$.*

Demostración. Sea \mathcal{D} un conjunto de alternativas, y sea W el conjunto de conclusiones garantizadas de un MEPM.

\Rightarrow) Sean $d_x, d_y \in \mathcal{D}$ dos alternativas tales que $d_x \sim_{W^*} d_y$. De acuerdo con la Definición 4.1.1, $d_x \succeq_{W^*} d_y$ y $d_y \succeq_{W^*} d_x$. Luego, como $d_x \succeq_{W^*} d_y$, de acuerdo con la Definición 4.1.1, $\text{pref}(d_x, d_y) \in W^*$. Como W^* es la clausura transitiva de W , debe existir una secuencia de alternativas $C_1 = [d_1, \dots, d_m]$ tal que $d_1 = d_x$, $d_m = d_y$, y para todo j ($1 \leq j < m$) $\text{pref}(d_j, d_{j+1}) \in W$. Por la misma razón, como $d_y \succeq_{W^*} d_x$, debe existir una secuencia de alternativas $C_2 = [d_m, \dots, d_n]$ ($m < n$) tal que $d_m = d_y$, $d_n = d_x$, y para todo k ($m \leq k < n$) $\text{pref}(d_k, d_{k+1}) \in W$. Por lo tanto, uniendo C_1 y C_2 , es posible construir una secuencia de alternativas $C = [d_1, \dots, d_m, \dots, d_n]$ en la que $d_1 = d_n = d_x$, $d_m = d_y$, y para todo i ($1 \leq i < n$) $\text{pref}(d_i, d_{i+1}) \in W$.

\Leftarrow) Sea $C = [d_1, \dots, d_y, \dots, d_n]$ una secuencia de alternativas tal que $d_1 = d_n = d_x$, y para todo i ($1 \leq i < n$) $\text{pref}(d_i, d_{i+1}) \in W$. Siendo W^* la clausura transitiva de W , $\text{pref}(d_x, d_y) \in W^*$ y $\text{pref}(d_y, d_x) \in W^*$. Luego, por la Definición 4.1.1, $d_x \succeq_{W^*} d_y$ y $d_y \succeq_{W^*} d_x$. Esto significa que $d_x \sim_{W^*} d_y$. \square

La proposición que se enuncia a continuación, asegura que si una alternativa d_x que es indiferente a otra d_y , es seleccionada, entonces d_y también será seleccionada. Por transitividad, si no existe ninguna alternativa que sea estrictamente preferida a d_x , entonces tampoco puede existir ninguna alternativa que sea estrictamente preferida a d_y .

Proposición 4.2.2. *Sea S el conjunto de alternativas seleccionadas a partir de un conjunto de alternativas \mathcal{D} y la clausura transitiva W^* de un conjunto de conclusiones garantizadas de un MEPM. Sean $d_x, d_y \in \mathcal{D}$ dos alternativas tales que $d_x \sim_{W^*} d_y$. Si $d_x \in S$, entonces $d_y \in S$.*

Demostración. Sea S el conjunto de alternativas seleccionadas a partir de un conjunto de alternativas \mathcal{D} y la clausura transitiva W^* de un conjunto de conclusiones garantizadas de un MEPM. Sean $d_x, d_y \in \mathcal{D}$ dos alternativas tales que $d_x \sim_{W^*} d_y$ y $d_x \in S$. Suponga por el absurdo que $d_y \notin S$. De acuerdo con la Definición 4.1.2, esto sólo es posible si existe $d_z \in \mathcal{D}$ tal que $d_z \succ_{W^*} d_y$. Por lo tanto, según la Definición 4.1.1, $\text{pref}(d_z, d_y) \in W^*$. Además, como $d_x \sim_{W^*} d_y$, según la Definición 4.1.1, $\text{pref}(d_y, d_x) \in W^*$. Luego, como W^* es una relación transitiva, al tener que $\text{pref}(d_z, d_y) \in W^*$ y $\text{pref}(d_y, d_x) \in W^*$, se puede asegurar que $\text{pref}(d_z, d_x) \in W^*$. Note que si $\text{pref}(d_x, d_z) \in W^*$, como $\text{pref}(d_y, d_x) \in W^*$, también

se tendría que $\text{pref}(d_y, d_z) \in W^*$. Si ese fuera el caso, d_y y d_z serían indiferentes (ver Definición 4.1.1), lo que está en contradicción con que $d_z \succ_{W^*} d_y$. Por ende, $\text{pref}(d_x, d_z) \notin W^*$. Entonces, de acuerdo con la Definición 4.1.1, $d_z \succ_{W^*} d_x$. Sin embargo, esto contradice la hipótesis de que $d_x \in S$. Esta contradicción provino de suponer que $d_y \notin S$. \square

La siguiente proposición, muestra que el conjunto de alternativas seleccionadas no puede ser vacío, si es que existe al menos una alternativa para elegir.

Proposición 4.2.3. *Sea S el conjunto de alternativas seleccionadas a partir de un conjunto de alternativas \mathcal{D} y la clausura transitiva W^* de un conjunto de conclusiones garantizadas de un MEPM. Si $\mathcal{D} \neq \emptyset$, entonces $S \neq \emptyset$.*

Demostración. Sea \mathcal{D} un conjunto de alternativas tal que $\mathcal{D} \neq \emptyset$, W^* la clausura transitiva de un conjunto de conclusiones garantizadas de un MEPM, y S el conjunto de alternativas a partir de \mathcal{D} y W^* . De acuerdo con la Definición 4.1.2, una alternativa $d_x \in \mathcal{D}$ es seleccionada si y sólo si no existe otra alternativa $d_y \in \mathcal{D}$ tal que $d_y \succ_{W^*} d_x$. Luego, si $\succ_{W^*} = \emptyset$ (ya sea porque para todo par de alternativas $d_x, d_y \in \mathcal{D}$, $d_x \sim_{W^*} d_y$, o bien, son incomparables), $S = \mathcal{D}$. En cambio, si $\succ_{W^*} \neq \emptyset$, debe existir al menos un elemento $d_x \in \mathcal{D}$ que sea maximal con respecto a \succ_{W^*} . Si d_x es maximal con respecto a \succ_{W^*} , significa que no existe $d_y \in \mathcal{D}$ tal que $d_y \succ_{W^*} d_x$. Por lo tanto, $d_x \in S$, y entonces, $S \neq \emptyset$. \square

Finalmente, la siguiente Proposición 4.2.4 asegura que si existe una alternativa d_x tal que para toda alternativa d_y , d_x es al menos tan buena como d_y , entonces d_x debe ser seleccionada.

Proposición 4.2.4. *Sea $\langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$ un escenario de elección. Sea S el conjunto de alternativas seleccionadas a partir de \mathcal{D} y la clausura transitiva W^* del conjunto de conclusiones garantizadas de un MEPM $\langle \text{Ag}, \text{Args}, \triangleright, \blacktriangleright \rangle$ tal que Args es construido a partir de $\langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$ y un conjunto de DePref-rules \mathcal{R} . Si existe una alternativa $d_y \in \mathcal{D}$ tal que para toda otra alternativa $d_x \in \mathcal{D}$ se cumple que si $c(d_x, v) \in E$, existe $c(d_y, u) \in E$ con $u \geq_c v$, entonces $d_y \in S$.*

Demostración. Sea $\langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$ un escenario de elección en el que existe una alternativa $d_y \in \mathcal{D}$ tal que para toda otra alternativa $d_x \in \mathcal{D}$ se cumple que si $c(d_x, v) \in E$, existe $c(d_y, u) \in E$ con $u \geq_c v$. Sea S el conjunto de alternativas seleccionadas a partir de \mathcal{D} y la clausura transitiva W^* del conjunto de conclusiones garantizadas de un MEPM

$\langle Ag, Args, \triangleright, \blacktriangleright \rangle$ tal que $Args$ es construido a partir de $\langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$ y un conjunto de *DePref*-rules \mathcal{R} . De acuerdo con la Proposición 2.2.2, no existe una *DePref*-rule $r \in \mathcal{R}$ tal que E pueda satisfacer a $r(d_x, d_y)$, con $d_x \in \mathcal{D}$. Por lo tanto, ningún argumento $\mathcal{A} \in Args$ puede ser tal que $\text{claim}(\mathcal{A}) = \text{pref}(d_x, d_y)$ (ver Definición 2.3.1). Luego, por la Definición 3.2.9, $\text{pref}(d_y, d_x) \notin W$. Como esto vale para todo $d_x \in \mathcal{D}$, $\text{pref}(d_x, d_y) \notin W^*$, lo que (de acuerdo con la Definición 4.1.1) significa que no existe $d_x \in \mathcal{D}$ tal que $d_x \succ_{W^*} d_y$. Por lo tanto, de acuerdo con la Definición 4.1.2, $d_x \in S$. \square

4.3. Generación de explicaciones

La confianza de los usuarios en los sistemas de toma de decisiones es algo crucial para que sean considerados útiles. Sería extraño encontrar a alguien que acepte la recomendación de un sistema en el que no confía, y mucho más que le permita tomar decisiones por ella/él. De acuerdo con Tintarev y Masthoff ([TM07] y [TM15]), una forma de fortalecer la confianza en este tipo de sistemas es a través de explicaciones que permitan entender el proceso de razonamiento mediante el cual se obtienen los resultados. Si las explicaciones son buenas, resulta más fácil confiar en el sistema ya que a pesar de que la recomendación no sea la esperada, se podrá hacer uso de las explicaciones para determinar si el resultado obtenido es correcto. Además, debido a que muchos sistemas, capaces de tomar decisiones de forma autónoma, están siendo aplicados para resolver problemas de distintos dominios (ingeniería, economía, salud, transporte, etc.) que afectan en gran medida a la sociedad, el requerimiento de explicaciones para este tipo de sistemas está empezando a ser considerado por los organismos de regulación gubernamentales. Por ejemplo, el Reglamento General de Protección de Datos (o GDPR, por sus siglas en inglés), que es aplicable en toda la Unión Europea desde mayo del 2018, establece en su Considerando 71 que los ciudadanos tendrán derecho a solicitar una explicación de cualquier decisión que haya resultado de un proceso automatizado (ver [Eur16]). Por este motivo, el marco de trabajo presentado en esta tesis cuenta con un proceso dedicado exclusivamente a la generación de explicaciones (ver Figura 4.1). Este proceso es parte de la etapa de selección ya que, al igual que el conjunto de alternativas seleccionadas, las explicaciones también son consideradas una de las salidas (o resultados) que ofrece esta propuesta. De hecho, las explicaciones son tan importantes como los resultados que se obtengan.

En esta sección se explicará en detalle la información provista por el proceso de generación de explicaciones para que los usuarios puedan comprender los resultados obtenidos. Como se puede apreciar en la Figura 4.1, este proceso toma las mismas entradas que el proceso de selección, y junto con los elementos que intervienen en la etapa de evaluación de argumentos elabora un grafo interactivo, mediante el cual los usuarios tendrán acceso a las explicaciones. Como será detallado más adelante, este grafo estará asociado a otras dos estructuras gráficas, que juntas con el primero, se complementan para mostrar distintos niveles de detalles acerca del proceso de decisión. Con esta herramienta, los usuarios podrán obtener información relacionada con cada uno de los distintos procesos: desde cuáles son los argumentos generados y cómo se relacionan entre ellos, hasta cuáles son las alternativas seleccionadas, cómo es el orden entre ellas y qué *DePref*-rules sustentan este orden.

La estructura subyacente de las explicaciones brindadas por el proceso de generación de explicaciones lleva el nombre de grafo explicativo. Este último, integra las explicaciones acerca de los resultados obtenidos a partir de todos los sub-procesos que conforman el proceso de decisión. En una primera instancia, los grafos explicativos permiten visualizar el orden de preferencia entre las alternativas, y las *DePref*-rules que sustentan dicho orden, permitiendo a los usuarios entender por qué una alternativa es seleccionada, o no. Además, estos grafos permiten acceder a los grafos de argumentos y árboles de dialéctica que explican los resultados del proceso argumentativo del cual se obtuvieron las conclusiones garantizadas con las que se construyó el orden de preferencia.

En [GCRS13], los autores estudiaron cómo brindar explicaciones adecuadas acerca de la garantía de una conclusión con respecto a un marco de trabajo argumentativo abstracto. Para ello definen la noción de *δ -explanation* (o explicación dialéctica), la cual explica los resultados mediante el uso de árboles de dialéctica. A continuación, se define el concepto de explicación dialéctica adaptado para explicar los resultados de un MEPM.

Definición 4.3.1 (Explicación dialéctica). *Una explicación dialéctica para una conclusión α con respecto a un MEPM \mathbf{F} es una tupla $\langle \mathbb{T}_{\mathcal{U}}(\alpha), \mathbb{T}_{\mathcal{U}}(\bar{\alpha}), \mathbb{T}_{\mathcal{D}}(\alpha) \cup \mathbb{T}_{\mathcal{D}}(\bar{\alpha}) \rangle$, donde $\mathbb{T}_{\mathcal{U}}(\alpha)$ (resp. $\mathbb{T}_{\mathcal{D}}(\alpha)$) es el conjunto de árboles de dialéctica marcados para los argumentos aceptados (resp. rechazados) de \mathbf{F} cuya conclusión es α .*

Una explicación dialéctica para una conclusión α contiene todos los árboles de dialéctica que están involucrados con α y con $\bar{\alpha}$. Esto se debe a que para entender por qué una

conclusión está (o no) garantizada uno podría querer analizar qué es lo que ocurre, no sólo con los argumentos que sustentan esta conclusión, sino también lo que ocurre con aquellos que sustentan a su complemento. Note que si α es una conclusión garantizada con respecto a un cierto MEPM \mathbf{F} , entonces $\mathbb{T}_{\mathbf{U}}(\alpha) \neq \emptyset$ ya que al menos debe existir un argumento \mathcal{A} en \mathbf{F} tal que $\text{claim}(\mathcal{A}) = \alpha$ y \mathcal{A} esté aceptado. Además, a partir del Corolario 3.3.1, se puede asegurar que si α es una conclusión garantizada, $\mathbb{T}_{\mathbf{U}}(\bar{\alpha}) = \emptyset$ ya que no puede existir ningún argumento aceptado cuya conclusión sea $\bar{\alpha}$; y por lo tanto, para todo argumento \mathcal{B} de \mathbf{F} tal que $\text{claim}(\mathcal{B}) = \bar{\alpha}$, $\mathbf{m}(\mathcal{T}(\mathcal{B})) \in \mathbb{T}_{\mathbf{D}}(\bar{\alpha})$.

La Figura 4.7 muestra (en la ventana superior) la explicación dialéctica para la conclusión $\text{pref}(t_4, t_2)$ con respecto al MEPM \mathbf{F}_B del Ejemplo 4.1.3. Los nodos de los árboles de dialéctica están representados con triángulos, y abajo a su derecha se puede encontrar el identificador del argumento con el que están etiquetados. Los nodos marcados \mathbf{U} están pintados de celeste, y los nodos marcados \mathbf{D} están pintados de amarillo. En la ventana inferior, se muestra el grafo de argumentos con aquellos que sustentan a las conclusiones $\text{pref}(t_4, t_2)$ y $\text{pref}(t_2, t_4)$. Todos estos argumentos son los que aparecen etiquetando los nodos de los árboles de dialéctica de la ventana superior.

Si sólo se utilizarían las explicaciones dialéctica para explicar los resultados del proceso de decisión, los usuarios deberían dilucidar el orden de preferencia que se construye a partir de las conclusiones garantizadas, sin más ayuda que conocer cuáles son los argumentos aceptados. Es por eso que surgen los grafos explicativos, para facilitar el entendimiento de la etapa de ordenamiento de las alternativas, agrupando y resumiendo los resultados que surgen de la etapa de argumentación.

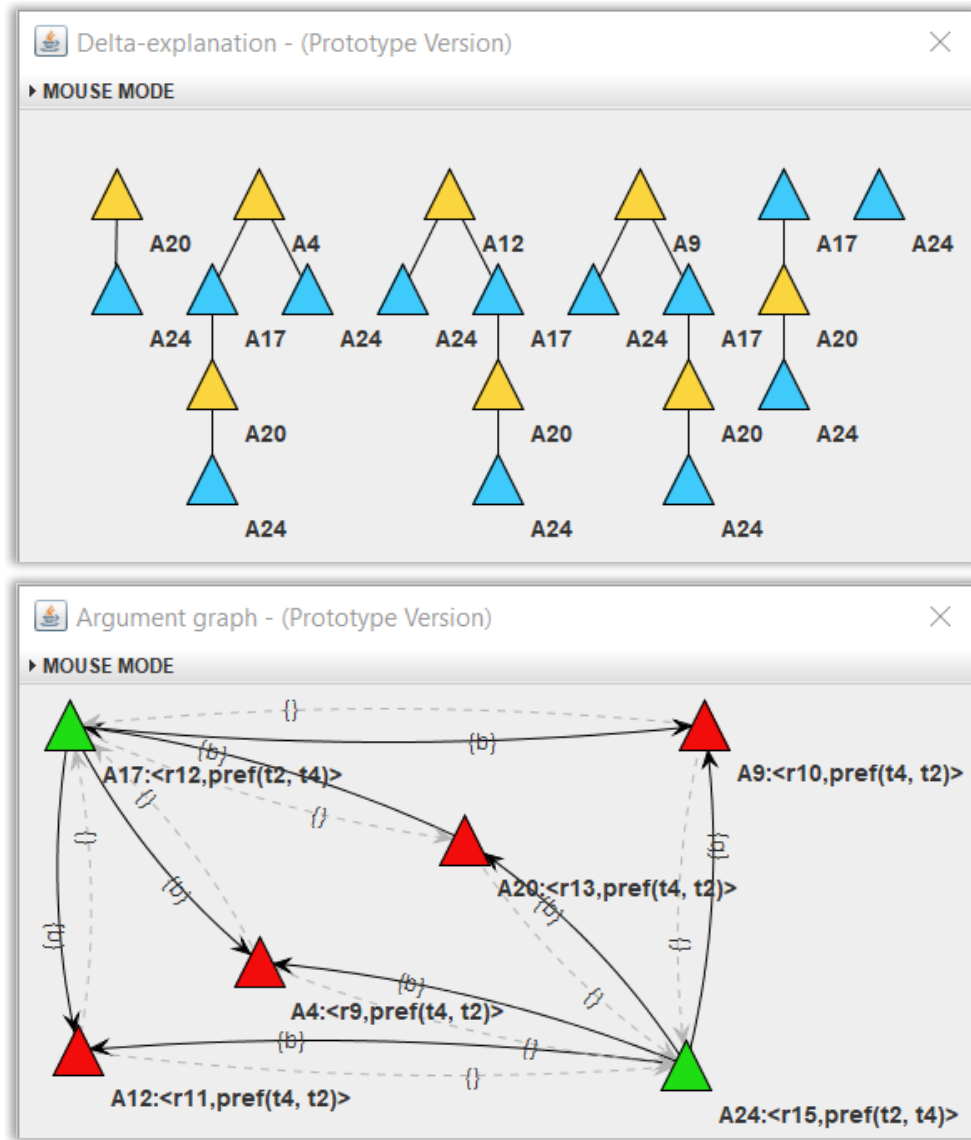


Figura 4.7: Capturas de pantalla mostrando cómo la aplicación que será presentada en el Capítulo 5 brinda explicaciones acerca del por qué la conclusión $pref(t_4, t_2)$ está garantizada en el MEPM F_B del ejemplo 4.1.3.

Definición 4.3.2 (Grafo explicativo). Sea \mathcal{D} un conjunto de alternativas, y W el conjunto de conclusiones garantizadas de un MEPM $F = \langle Ag, Args, \triangleright, \blacktriangleright \rangle$. Un grafo explicativo es un grafo dirigido $\langle \mathcal{D}, \rightarrow \rangle$, donde \mathcal{D} son sus nodos, y $\rightarrow = \{(d_x, d_y) \mid pref(d_x, d_y) \in W\}$ sus arcos. Cada arco (d_x, d_y) tiene asociado (o está vinculado a):

- i) El conjunto de *DePref*-rules con las que se construyeron los argumentos aceptados que sustentan la conclusión $pref(d_x, d_y)$, es decir, $\{r \mid \exists \mathcal{A} \in \text{Args que es aceptado en } \mathbf{F}, \text{ con } r = \mathbf{rule}(\mathcal{A}) \text{ y } pref(d_x, d_y) = \mathbf{claim}(\mathcal{A})\}$,
- ii) La explicación dialéctica (con respecto a \mathbf{F}) de la conclusión $pref(d_x, d_y)$, y
- iii) El grafo de argumentos para el MEPM $\langle Ag, \text{Args}', \triangleright, \blacktriangleright \rangle$, donde Args' es el conjunto de todos los argumentos con los que están etiquetados los nodos de los árboles de dialéctica que conforman la explicación dialéctica de la conclusión $pref(d_x, d_y)$.

La Figura 4.8 muestra el grafo explicativo que justifica el conjunto de alternativas seleccionadas S_B del Ejemplo 4.1.3. Cada nodo de un grafo explicativo representa una de las alternativas del escenario de elección, mientras que los arcos de un grafo explicativo muestran la relación entre las alternativas que es inducida a partir de las conclusiones garantizadas, obtenidas en la etapa de argumentación. Note que para la implementación de los grafos explicativos, con el objetivo de que las explicaciones sean aún más completas, se decidió colorear de blanco los nodos correspondientes a las alternativas seleccionadas, y de gris oscuro los nodos correspondientes a las demás alternativas (ver Figura 4.8). Por medio del grafo de la Figura 4.8, es posible identificar que las mejores alternativas son t_1 , t_2 y t_4 , que todas estas están en un ciclo (son equivalentes), y que son mejores que t_3 . Además, se pueden ver las *DePref*-rules que fueron utilizadas para construir los argumentos aceptados que sustentan la relación de preferencia. Por ejemplo, el arco (t_2, t_4) está etiquetado con el conjunto $\{r_{12}, r_{15}\}$ ya que los argumentos $\langle r_{12}(t_2, t_4), pref(t_2, t_4) \rangle$ (identificado como A17 en la Figura 4.7) y $\langle r_{15}(t_2, t_4), pref(t_2, t_4) \rangle$ (identificado como A24 en la Figura 4.7) son aceptados. Note que dicho arco está vinculado a la explicación dialéctica y al grafo de argumentos que aparecen en la Figura 4.7.

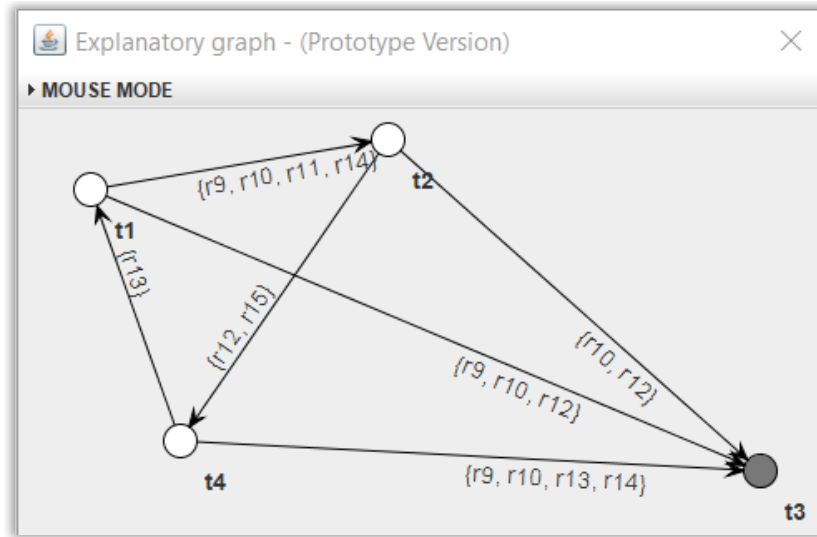


Figura 4.8: Captura de pantalla de cómo la aplicación que será presentada en el Capítulo 5 exhibe el grafo explicativo correspondiente al problema planteado en el Ejemplo 4.1.3.

Retomando el ejemplo de la familia Jota, la Figura 4.9 muestra, en la parte superior izquierda, la captura de pantalla del grafo explicativo para la solución que retorna el proceso con respecto al problema que enfrentan los miembros de la familia. Observe que tanto el orden entre las alternativas, como el conjunto de alternativas seleccionadas, coinciden con los presentados en el Ejemplo 4.1.2. A la derecha del grafo explicativo, se muestra el grafo de argumentos, y debajo, la explicación dialéctica asociada al arco (*jngla, playa*) del grafo explicativo. Las herramientas gráficas cuentan con una gran ventaja sobre otras herramientas que sólo utilizan texto para explicar los resultados de un sistema. Principalmente, debido a que permiten resumir mucha información en una sola imagen, que si tuviese que ser escrita probablemente se necesitarían varios párrafos, lo que terminaría abrumando al usuario. Además, como se comentará en el Capítulo 5, las explicaciones son interactivas por lo que el usuario podrá decidir cuánta información visualizar.

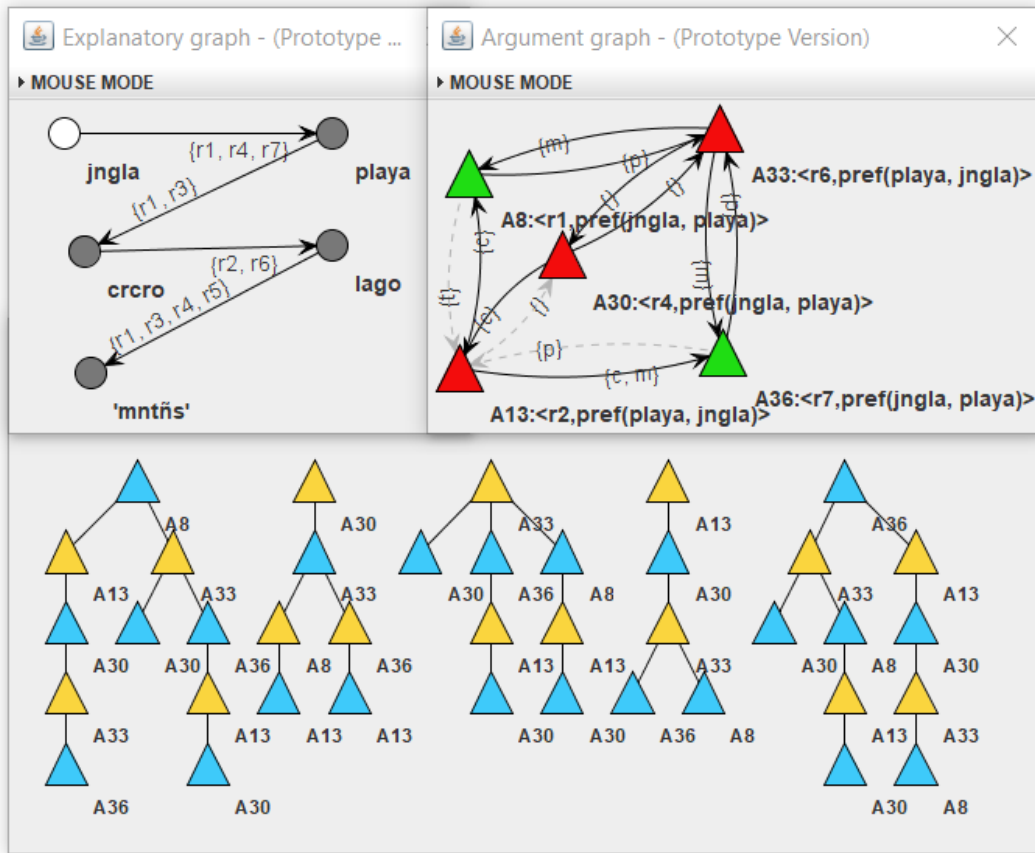


Figura 4.9: Capturas de pantalla de la aplicación que será presentada en el Capítulo 5. Arriba se muestra el grafo explicativo de la solución que ofrece el proceso de decisión al problema de la familia Jota, abajo se muestra el grafo de argumentos y la explicación dialéctica correspondiente al arco (*jngla, playa*).

La importancia de las explicaciones en sistemas de toma de decisiones, sistemas de recomendación, o sistemas de soporte para la toma de decisiones, también ha sido estudiada desde un punto de vista más social o filosófico, por lo que existen autores que han tratado de identificar las características que deberían tener las explicaciones. Por ejemplo, en [TM15], los autores identifican diferentes propósitos para los cuales se pueden utilizar las explicaciones, y cómo estas deberían ser diseñadas para cumplir con su propósito. En línea con esto, se podría decir que los grafos explicativos mejoran el nivel de transparencia y persuasión del marco de trabajo presentado en esta tesis. Según [TM15], una explicación brinda transparencia si explica el funcionamiento del sistema, es decir, explica cómo es que el sistema llega a una dada recomendación o solución. Los grafos explicati-

vos permiten a los usuarios visualizar la información que fue utilizada para construir el conjunto de alternativas seleccionadas, muestran el orden entre las alternativas que justifica el comportamiento del proceso de selección, y además, conecta este orden con los grafos de argumentos y las explicaciones dialécticas que justifican la estructura del orden, dando como resultado una explicación exhaustiva del comportamiento del marco de trabajo, lo que hace que el proceso de decisión sea totalmente transparente. Más aún, ante resultados inesperados y/o erróneos, los grafos explicativos pueden ser utilizados como una herramienta de depuración, haciendo que sea posible identificar errores que hayan surgido durante la etapa de modelado del problema, como por ejemplo, un valor o una *DePref*-rules incorrecta. Esto, trae consigo el beneficio adicional de aumentar el nivel de confianza de los usuarios para con el sistema ya que los mismos podrían identificar más fácilmente si los resultados son correctos cuando estos no son los esperados. En cuanto a la persuasión, los autores indican que una explicación puede persuadir a un usuario si trata de convencerlo de que la recomendación dada es mejor que el resto de las opciones. Esta es también una propiedad de los grafos explicativos ya que no sólo permiten entender por qué una alternativa es seleccionada, sino que además permite a los usuarios indagar acerca de las razones por las cuales una alternativa fue descartada.

Por otro lado, según Miller [Mil19], los seres humanos prefieren explicaciones contrastivas. Una explicación para la ocurrencia de un dado evento es contrastiva, si explica por qué ha ocurrido dicho evento en lugar de que hayan ocurrido otros. Observe que los grafos explicativos proveen explicaciones contrastivas de distintos niveles. En primer lugar, el orden entre las alternativas que muestra un grafo explicativo permite responder a preguntas como: ¿por qué la alternativa X es seleccionada en lugar de la alternativa Y ? o ¿por qué la alternativa X no es seleccionada, mientras que la alternativa Y sí? Estas preguntas se responden *a priori* visualizando el orden entre las alternativas ya que con el mismo es posible dilucidar cuáles son los elementos maximales de la relación de preferencia que se infiere a partir del conjunto de conclusiones garantizadas. Y en segundo lugar, al dar acceso (mediante los arcos del grafo explicativo) al grafo de argumentos y a las explicaciones dialécticas asociadas a las conclusiones que sustentan la relación de preferencia, se pueden responder preguntas como: ¿por qué X es preferida a Y , y no ocurre lo contrario? o ¿por qué un argumento es aceptado, y no rechazado?. Note que el grafo de argumentos no sólo muestra los argumentos aceptados y rechazados que sustentan las distintas conclusiones, sino que además, permite responder preguntas como: ¿por qué dos argumentos están en conflicto? o ¿por qué un argumento \mathcal{A} derrota a otro \mathcal{B} ? las cuales también ayudan a

comprender por qué un argumento es aceptado, o no.

4.4. Trabajando sin un agente de máxima prioridad

En el ejemplo de la Familia Jota, desarrollado a lo largo de los capítulos anteriores, se plantea un escenario en el que existe un único agente con máxima prioridad (ver Ejemplo 3.1.2). En dicho ejemplo, el orden de prioridad entre los agentes era $\blacktriangleright_J = \{c \blacktriangleright m, c \blacktriangleright p, c \blacktriangleright t, m \blacktriangleright t, p \blacktriangleright t\}$, por lo que Carolina (c) tenía prioridad sobre sus padres y hermano (m, p y t, respectivamente). En esta sección se mostrará el comportamiento del marco de trabajo propuesto en esta tesis sobre un nuevo problema de decisión en el que no existe un agente que tenga prioridad sobre todos los demás. Con este nuevo ejemplo, se mostrará que esta propuesta tiene en cuenta la opinión de todos los agentes a la hora de computar los resultados, y que es un error pensar que solamente considera los intereses de los agentes de mayor prioridad como podría mal interpretarse del ejemplo de la Familia Jota donde prevalece la opinión de Carolina.

Ejemplo 4.4.1. *Un grupo de médicos debe tomar la decisión acerca de qué tratamiento aplicar a un paciente para tratar de sanar su enfermedad. Este grupo está formado por Nilda, Osvaldo, Roberto y Susana, identificados como: n, o, r y s, respectivamente. Nilda y Osvaldo cuentan con muchos años de experiencia, y entre los dos dirigen la clínica. Roberto, es el asistente y mano derecha de Nilda, y Susana, es una residente a cargo de Osvaldo y Roberto. El orden de prioridad que representa la jerarquía entre estos agentes está dado por $\blacktriangleright_M = \{n \blacktriangleright r, n \blacktriangleright s, r \blacktriangleright s, o \blacktriangleright s\}$. Note que entre Nilda y Osvaldo, como entre Osvaldo y Roberto, no existe prioridad.*

En este escenario de elección son considerados cuatro tratamientos (o alternativas) disponibles: $\mathcal{D}_M = \{trat_1, trat_2, trat_3, trat_4\}$. Los tratamientos son evaluados en base a su eficacia, duración y la gravedad de los efectos colaterales (o GEC), siendo estos aspectos representados por el conjunto de criterios $\mathcal{C}_M = \{efi, dur, gec\}$. La eficacia se mide con un número que indica el porcentaje de veces que el tratamiento resultó efectivo, por lo que $\mathcal{V}_M(efi)$ es el rango de enteros positivos ordenados de menor a mayor (cuánto mayor es la eficacia, mejor es el tratamiento). La duración se mide utilizando el rango de valores establecidos por $\mathcal{V}_M(dur) = [de_por_vida, larga, moderada, corta]$ (cuanto menor es la duración, mejor es el tratamiento). Y por último, la gravedad de los efectos colaterales se

mide utilizando el rango $\mathcal{V}_M(gec) = [graves, moderados, leves, no_produce]$ (cuanto menor es la gravedad de los efectos colaterales, mejor es el tratamiento). Las características de cada tratamiento están especificadas por el siguiente conjunto de evidencia:

$$E_M = \left\{ \begin{array}{lll} efi(trat_1, 85), & dur(trat_1, corta), & gec(trat_1, leves), \\ efi(trat_2, 80), & dur(trat_2, moderada), & gec(trat_2, no_produce), \\ efi(trat_3, 95), & dur(trat_3, de_por_vida), & gec(trat_3, moderados), \\ efi(trat_4, 90), & dur(trat_4, larga), & gec(trat_4, leves) \end{array} \right\}$$

Para evaluar los tratamientos los cuatros agentes están de acuerdo en utilizar las siguientes DePref-rules:

$$\begin{aligned} r_{17} &= \left\{ better(X, Y, efi) \right\} \Rightarrow pref(X, Y) \\ r_{18} &= \left\{ better(X, Y, dur) \right\} \Rightarrow pref(X, Y) \\ r_{19} &= \left\{ better(X, Y, gec) \right\} \Rightarrow pref(X, Y) \\ r_{20} &= \left\{ \begin{array}{l} worse(X, Y, efi), \min(X, efi, 80) \\ better(X, Y, gec), \min(X, gec, no_produce) \end{array} \right\} \Rightarrow pref(X, Y) \\ r_{21} &= \left\{ \begin{array}{l} better(X, Y, efi), \min_dist(X, Y, efi, 10), \\ worse(X, Y, dur), \max_dist(X, Y, dur, 1), \\ \min(X, dur, larga) \end{array} \right\} \Rightarrow pref(X, Y) \end{aligned}$$

Las DePref-rules r_{17} , r_{18} y r_{19} establecen que si un tratamiento X es mejor que otro tratamiento Y con respecto a cualquiera de los criterios de comparación, entonces hay una razón para considerar que X sea preferido a Y . La DePref-rule r_{20} indica que si un tratamiento X es menos eficaz que otro tratamiento Y , pero la eficacia de X al menos del 80 %, y además, X no produce efectos colaterales mientras que Y sí, entonces hay una razón para considerar que X sea preferido a Y . Por último, r_{21} , indica que un tratamiento X podría ser considerado preferido a otro Y a pesar de que Y tenga una menor duración, siempre y cuando, la diferencia de duración no sea mayor uno, X no dure para toda la vida (que su duración sea larga o menor), y además, X sea al menos 10 % más eficaz que Y .

Nilda y Osvaldo coinciden en que la GEC es más importante que la eficacia ($r_{19} \triangleright_n r_{17}$ y $r_{19} \triangleright_o r_{17}$), mientras que, Roberto y Susana consideran lo contrario ($r_{17} \triangleright_r r_{19}$ y $r_{17} \triangleright_s r_{19}$). Además, Osvaldo considera que la gravedad de los efectos es también más

importantes que la duración ($r_{19} \triangleright_o r_{18}$), al contrario que Roberto, que piensa que la duración es más importante ($r_{18} \triangleright_r r_{19}$). Nilda y Susana no tienen una postura con respecto a la importancia entre estos últimos dos criterios. No obstante, Susana considera que eficacia es más importante que la duración ($r_{17} \triangleright_s r_{18}$). Por último, todos los agentes están de acuerdo en que las condiciones que plantea la DePref-rules r_{20} son suficientes para considerar que la GEC es más importante que la eficacia ($r_{20} \triangleright_n r_{17}$, $r_{20} \triangleright_o r_{17}$, $r_{20} \triangleright_r r_{17}$ y $r_{20} \triangleright_s r_{17}$), así como también que las condiciones que plantea la DePref-rules r_{21} son suficientes para considerar que eficacia sea más importante que la duración ($r_{21} \triangleright_n r_{18}$, $r_{21} \triangleright_o r_{18}$, $r_{21} \triangleright_r r_{18}$ y $r_{21} \triangleright_s r_{18}$).

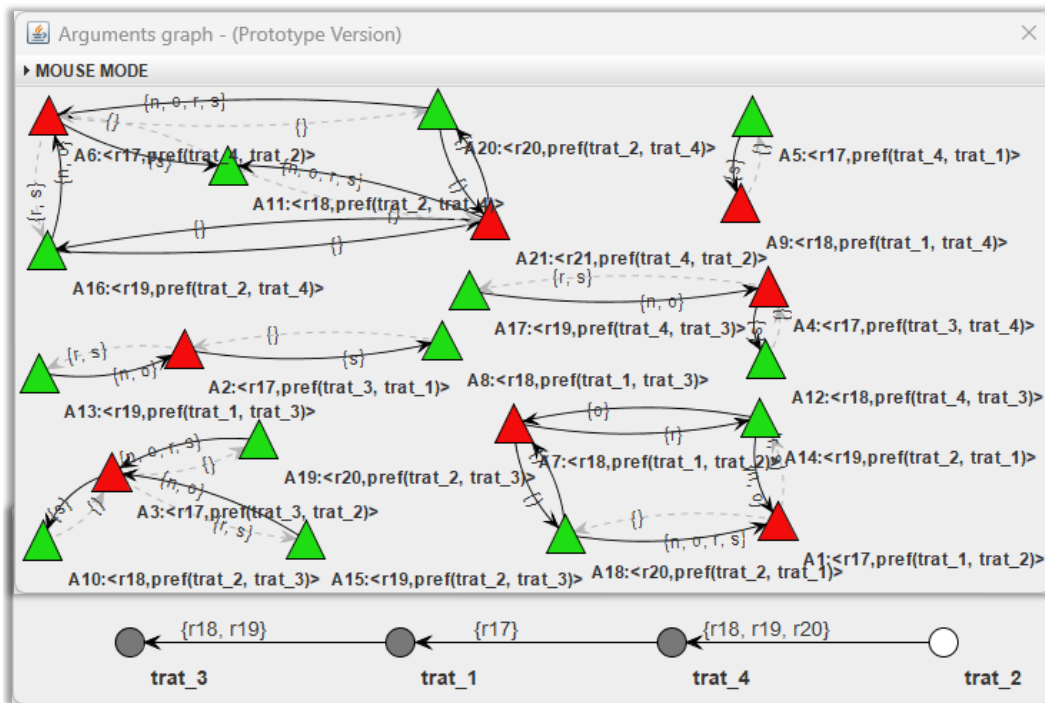


Figura 4.10: (Arriba) Grafo de argumentos para el MEPM $\mathbf{F}_M = \langle Ag_M, Arg_{SM}, \triangleright_M, \blacktriangleright_M \rangle$. (Abajo) Grafo explicativo mostrando la relación de preferencia entre los tratamientos construido en base a las conclusiones garantizadas de \mathbf{F}_M .

En la parte superior de la Figura 4.10 se muestra el grafo de argumentos para el MEPM $\mathbf{F}_M = \langle Ag_M, Arg_{SM}, \triangleright_M, \blacktriangleright_M \rangle$, donde $Ag_M = \{n, o, r, s\}$, $\blacktriangleright_M = \{n \blacktriangleright r, n \blacktriangleright s, r \blacktriangleright s, o \blacktriangleright s\}$, $\triangleright_M = [\triangleright_n, \triangleright_o, \triangleright_r, \triangleright_s]$ y Arg_{SM} es el conjunto de todos los argumentos de preferencia que se pueden construir a partir de $\langle \mathcal{D}_M, \mathcal{C}_M, \mathcal{V}_M, E_M \rangle$ y $\mathcal{R}_M = \{r_{17}, r_{18}, r_{19}, r_{20}, r_{21}\}$. En la parte inferior de la Figura 4.10 aparece el grafo explicativo que muestra el orden de

preferencia que se obtiene a partir de las conclusiones garantizadas de \mathbf{F}_M . Observe que $trat_2$ es la única alternativa seleccionada.

El grafo de argumentos de la Figura 4.10, resulta útil para analizar la influencia que cada agente tiene sobre la recomendación final, es decir, que el tratamiento $trat_2$ sea el que los médicos apliquen al paciente. Por ejemplo, note que **A16** derrota a **A6** ya que tanto Nilda como Osvaldo consideran que r_{19} es más importante que r_{17} y la opinión de Nilda tiene prioridad sobre la opinión de Roberto y sobre la de Susana que piensan lo contrario. Además, note que **A7** y **A14** se derrotan mutuamente ya que para Osvaldo r_{19} es más importante que r_{18} , mientras que Roberto opina lo contrario, y no existe prioridad entre ellos. La opinión de estos agentes tiene el mismo peso, y por lo tanto, la fortaleza de dichos argumentos es la misma. Por último, observe lo que ocurre entre **A5** y **A9**. En este caso, se tiene que **A5** derrota a **A9** dado que Susana considera que r_{17} es más importante que r_{18} , y como nadie se opone a eso, a pesar de que Susana sea el agente con menor prioridad, su opinión igual cuenta. De hecho, el siguiente ejemplo, muestra que si no se tuviera en cuenta la opinión de Susana, el orden de preferencia no sería el mismo.

Ejemplo 4.4.2. *Considere nuevamente el escenario de elección $\langle \mathcal{D}_M, \mathcal{C}_M, \mathcal{V}_M, E_M \rangle$, el conjunto de DePref-rules \mathcal{R}_M , el conjunto de agentes Ag_M , y los ordenes de importancia \triangleright_M del Ejemplo 4.4.1. La Figura 4.11 (arriba) muestra el grafo de argumentos para el MEPM $\mathbf{F}'_M = \langle Ag'_M, Args_M, \triangleright'_M, \blacktriangleright'_M \rangle$, donde $Ag'_M = \{n, o, r\}$, $\blacktriangleright'_M = \{n \blacktriangleright r\}$, $\triangleright'_M = [\triangleright_n, \triangleright_o, \triangleright_r]$ y $Args_M$ (el mismo que para el Ejemplo 4.4.1) es el conjunto de todos los argumentos de preferencia que se pueden construir a partir de $\langle \mathcal{D}_M, \mathcal{C}_M, \mathcal{V}_M, E_M \rangle$ y \mathcal{R}_M . En la parte inferior de la Figura 4.11 aparece el grafo explicativo que muestra el orden de preferencia que se obtiene a partir de las conclusiones garantizadas de \mathbf{F}'_M .*

La única diferencia entre la situación planteada en el Ejemplo 4.4.1 y la planteada en el Ejemplo 4.4.2, es que en este último, no se considera a Susana como parte del grupo de agentes. Esto es equivalente a ignorar por completo su opinión. Note que a pesar de que en el Ejemplo 4.4.1, Susana es el agente con menor prioridad, su opinión es tenida en cuenta por el formalismo, y es por ese motivo que se pueden observar diferencias entre la relación de preferencia que se obtiene luego de evaluar los argumentos en ambos ejemplos.

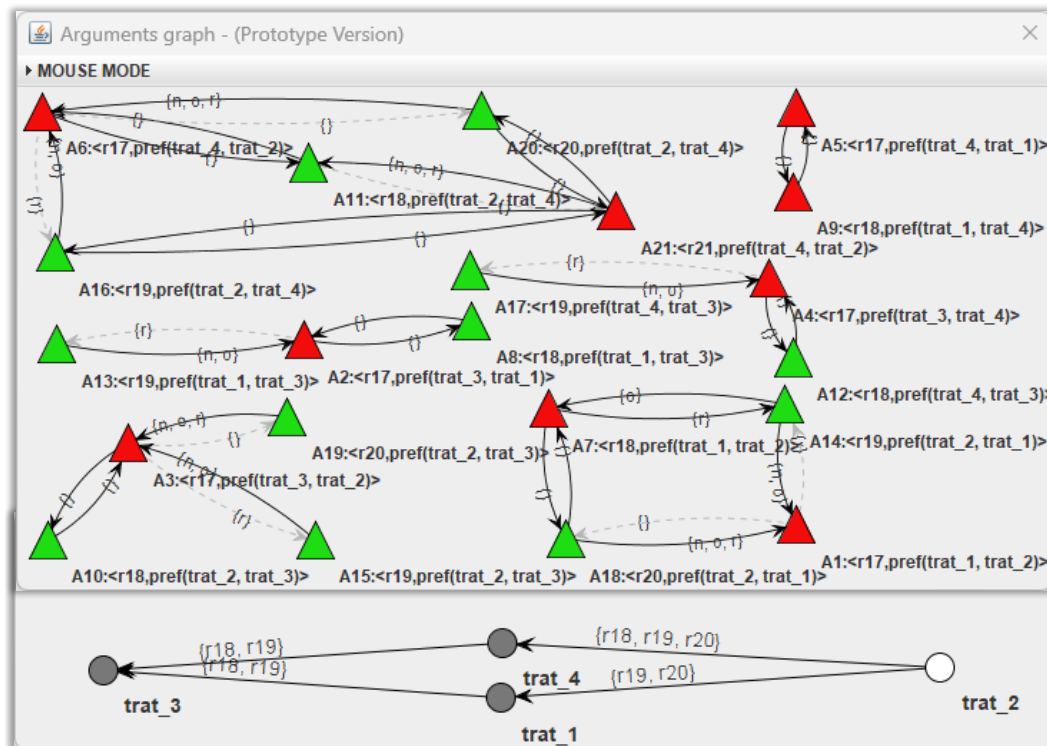


Figura 4.11: (Arriba) Grafo de argumentos para el MEPM $\mathbf{F}'_M = \langle Ag'_M, Args_M, \triangleright'_M, \blacktriangleright'_M \rangle$. (Abajo) Grafo explicativo mostrando la relación de preferencia entre los tratamientos construido en base a las conclusiones garantizadas de \mathbf{F}'_M .

Esto demuestra que el marco de trabajo propuesto en esta tesis tiene en cuenta la opinión de todos los agentes a la hora de computar los resultados. Por lo tanto, sería un error pensar que esta propuesta sólo hace caso a los intereses de los agentes de mayor prioridad.

4.5. Trabajando con evidencia incompleta

Tanto en el ejemplo de la Familia Jota (ver Ejemplo 2.1.1), como en el ejemplo del robot anti-incendios (ver Ejemplo 4.1.3), y como en el ejemplo de los médicos (ver Ejemplo 4.4.1) se consideraron conjuntos de evidencia completos. Es decir, todas las alternativas tenían asociado un valor con respecto a cada uno de los criterios. Note que a pesar de que el conjunto de evidencia debe ser consistente (ver Definición 2.1.1), no es un requisito que sea completo. La principal consecuencia de que el conjunto de evidencia no sea completo es que

podría resultar más difícil (o incluso imposible) comparar las alternativas, dado que habrá menos información para poder hacer esto. En esta sección se analizará el comportamiento del marco de trabajo en escenarios donde existan alternativas que no tienen un valor asociado en al menos uno de los criterios.

Ejemplo 4.5.1. *Considere nuevamente el dominio de aplicación considerado en el Ejemplo 4.4.1, en el que un grupo de médicos ($Ag_M = \{\mathbf{n}, \mathbf{o}, \mathbf{r}, \mathbf{s}\}$) deben decidir que tratamiento aplicar a uno de sus pacientes. Además, considere que los médicos utilizarán los mismos criterios para evaluar los tratamientos ($C_M = \{efi, dur, gec\}$) y los mismos rangos de valores que en el ejemplo mencionado. Sin embargo, para este nuevo ejemplo, el conjunto de alternativas esta dado por $\mathcal{D}_I = \{trat_5, trat_6, trat_7\}$ y el conjunto de evidencia es el siguiente:*

$$E_I = \left\{ \begin{array}{lll} efi(trat_5, 70), & dur(trat_5, larga), & gec(trat_5, no_produce), \\ & dur(trat_6, corta), & gec(trat_6, moderados), \\ efi(trat_7, 80), & dur(trat_7, moderada), & \end{array} \right\}$$

Note que en este caso la evidencia no es completa ya que no se conoce la eficacia de *trat_6*, ni tampoco la GEC de *trat_7*. Hipotéticamente, esto podría deberse a que estos tratamientos podrían ser experimentales, y no se tendría acceso a dichos datos, o que si bien se podría acceder a estos, no serían lo suficientemente certeros, por lo que resultaría preferible dejarlos sin completar.

Como será mostrado a continuación, a pesar de que la evidencia esté incompleta, el marco de trabajo propuesto en esta tesis funciona correctamente. De hecho, como la etapa de generación de argumentos puede ejecutarse sin problemas, las demás etapas que forman parte del proceso de decisión tampoco se verán afectadas. La principal consecuencia de que el conjunto de evidencia no sea completo es que podría resultar más difícil (o incluso imposible) comparar las alternativas, dado que habrá menos información para poder hacerlo. Esto resulta evidente si se observa que, por ejemplo, el tratamiento *trat_6* no va a poder ser comparado mediante ninguna *DePref*-rule que necesite evaluar la eficacia, así como tampoco *trat_7* va a poder ser comparado mediante *DePref*-rules que evalúen la GEC.

Ejemplo 4.5.2. *(Continuación) Al ejecutar el proceso de generación de argumentos tomando como entrada el nuevo escenario de elección con evidencia incompleta*

$\langle \mathcal{D}_I, \mathcal{C}_M, \mathcal{V}_M, E_I \rangle$ y el conjunto de *DePref*-rules $\mathcal{R}_M = \{r_{17}, r_{18}, r_{19}, r_{20}, r_{21}\}$ del Ejemplo 4.4.1, se obtiene el conjunto de argumentos

$$Args_I = \left\{ \begin{array}{l} \langle r_{17}(trat_7, trat_5), pref(trat_7, trat_5) \rangle \\ \langle r_{18}(trat_6, trat_5), pref(trat_6, trat_5) \rangle \\ \langle r_{18}(trat_6, trat_7), pref(trat_6, trat_7) \rangle \\ \langle r_{18}(trat_7, trat_5), pref(trat_7, trat_5) \rangle \\ \langle r_{19}(trat_5, trat_6), pref(trat_5, trat_6) \rangle \end{array} \right\}$$

Como se mencionó anteriormente, el proceso de generación de argumentos funciona igual, por más que el conjunto de evidencia no sea completo. Para verificar si una *DePref*-rule es satisfecha, este proceso recae en la semántica de cada una de los predicados de la regla (ver Definición 2.2.3). Por la forma en la que la semántica de los predicados está definida, se necesita que el conjunto de evidencia contenga las piezas de evidencia correspondientes para poder evaluar las alternativas y determinar si el predicado es satisfecho, o no. Si la evidencia no es completa, es probable que alguno de los predicados no sea satisfecho dado que no hay información para asegurarlo. Esto sucede, por ejemplo, si se instancia *DePref*-rule r_{17} con $trat_5$ y $trat_6$ (o $trat_6$ y $trat_5$), como no se conoce la eficacia de $trat_6$, $better(trat_5, trat_6, efi)$ no puede ser satisfecha por E_M , y por lo tanto, no existen argumentos basados en r_{17} que sustenten alguna preferencia entre $trat_5$ y $trat_6$. Note que algo similar sucede al instanciar r_{17} con $trat_6$ y $trat_7$ (o $trat_7$ y $trat_6$), ya que en definitiva ninguna *DePref*-rule que evalúe el criterio efi podrá ser usada para comparar a $trat_6$ con cualquier otra alternativa. En general, si existe una alternativa d tal que no existe una pieza de evidencia $c(d, v)$ para un dado criterio c , entonces ninguna *DePref*-rule que evalúe a c y sea instanciada con d (y cualquier otra alternativa), va a poder ser satisfecha. Esto explica por qué en el Ejemplo 4.5.2 no hay argumentos construidos a partir de las *DePref*-rules r_{20} o r_{21} , ya que estas reglas evalúan los tres criterios, y para dos de las tres alternativas no existe información con respecto a al menos uno de los criterios. Una vez terminado el proceso de generación de argumentos, el resto de los procesos internos que forman parte del proceso de decisión son llevados a cabo con normalidad.

Ejemplo 4.5.3. (Continuación) La Figura 4.12 muestra los argumentos aceptados y rechazados (determinados por medio del proceso de evaluación dialéctico) del MEPM $\mathbf{F}_I = \langle Args_I, Ag_M, \triangleright_M, \blacktriangleright_M \rangle$, donde $Ag_M = \{\mathbf{n}, \mathbf{o}, \mathbf{r}, \mathbf{s}\}$, $\triangleright_M = [\triangleright_n, \triangleright_o, \triangleright_r, \triangleright_s]$ y $\blacktriangleright_M =$

$\{n \triangleright r, n \triangleright s, r \triangleright s, o \triangleright s\}$ es el grupo de agentes, los ordenes de importancia y el orden de prioridad del Ejemplo 4.4.1. De esta figura, se puede observar que el conjunto de conclusiones garantizadas de \mathbf{F}_I es $W_I = \{pref(trat_6, trat_7), pref(trat_7, trat_5)\}$. Luego, al ejecutar el proceso de selección de alternativas con \mathcal{D}_I y W_I^* se tiene que $S_I = \{trat_6\}$ ya que $trat_6 \succ_{W_I^*} trat_7 \succ_{W_M^*} trat_5$.

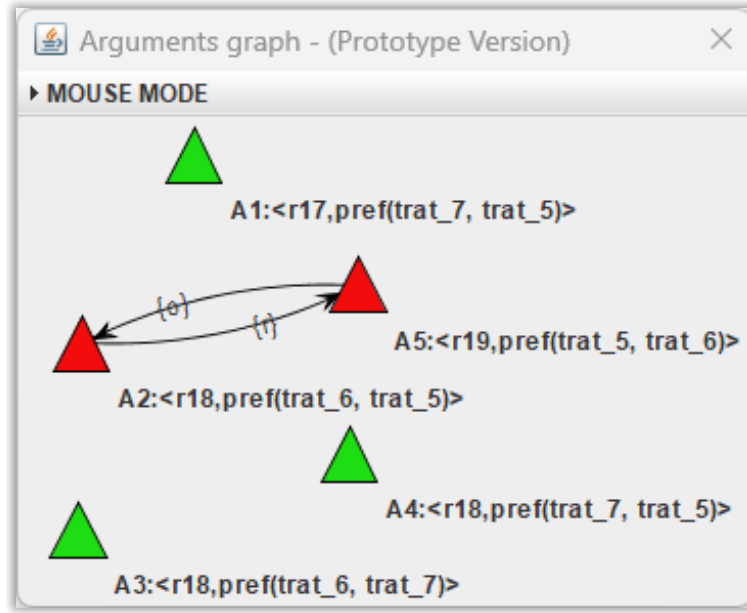


Figura 4.12: Argumentos aceptados y rechazados del MEPM \mathbf{F}_M .

A pesar de que el proceso de decisión puede ejecutarse correctamente en escenarios donde la evidencia no es completa, los usuarios deberían tener en cuenta las consecuencias de esto antes de usarlo. Como se explicó antes, cuanto menos evidencia haya acerca de una alternativa, menor va a ser la cantidad de reglas con las que se pueda comparar esa alternativa con otras, y por ende, menor va a ser la cantidad de argumentos a favor y/o en contra de preferir esa alternativa. Observe lo que sucedería si existiese un tratamiento $trat_8$ para el que sólo se supiera que su eficacia es del 80% ($trat_8$ sería como $trat_7$, pero sin información acerca de su duración). Si esto pasa, entonces la única forma de comparar este tratamiento con los demás, sería a través de r_{17} , y el único argumento que se generaría sería $\langle r_{17}(trat_8, trat_5), pref(trat_8, trat_5) \rangle$. Como el resto de la evidencia es igual, todos los argumentos en $Args'_I$ se conservarían, y sólo habría que agregar este nuevo argumento, por lo que el nuevo conjunto de argumentos sería $Args'_I = Args_I \cup \{\langle r_{17}(trat_8, trat_5), pref(trat_8, trat_5) \rangle\}$. Suponiendo que los ordenes

de importancia entre las reglas y las prioridades entre los agentes no cambian, las conclusiones garantizadas que se obtendrían a partir de este conjunto de argumentos sería $W'_I = W_I \cup \{pref(trat_8, trat_5)\}$, por lo que $trat_8 \succ_{W'_I} trat_5$, y al igual que antes, $trat_6 \succ_{W'_I} trat_7 \succ_{W'_I} trat_5$. Note que $trat_8$ sería incomparable con $trat_6$ y $trat_7$. Luego, el conjunto de alternativas seleccionadas sería $S'_I = \{trat_6, trat_8\}$.

En algunos casos, cuanto menos información se tenga acerca de una alternativa, más escasos van a ser los argumentos en contra de preferir a esta alternativa. Por la forma en la que está definido el conjunto de alternativas seleccionadas (ver Definición 4.1.2), se le da más prioridad a las razones en contra de preferir una alternativa que a las razones a favor de preferirla, siendo que este algoritmo descarta toda aquella alternativa para la que exista una alternativa que sea mejor. Por lo tanto, las alternativas para las cuáles sólo se tiene información acerca de sus puntos fuertes y se desconocen los valores de las características en las que podrían ser débiles, son más propensas a ser seleccionadas, que aquellas para las que se conocen todas sus características.

A continuación, se mostrará qué es lo que sucede al eliminar algunas piezas de evidencia del escenario de elección del problema de la familia Jota (ver Ejemplo 2.1.1).

Ejemplo 4.5.4. *Considere nuevamente el escenario de elección $\langle \mathcal{D}_J, \mathcal{C}_J, \mathcal{V}_J, E_J \rangle$ que se presenta en el Ejemplo 2.1.1. Recuerde que dicho escenario representa la información que tenían los miembros de la familia Jota (Máma, Papá, Carolina y Tomás), con respecto a los posible lugares dónde pueden ir de vacaciones para celebrar los 15 años de Carolina. El conjunto de alternativas es $\mathcal{D}_J = \{mntñs, playa, lago, jngla, crcro\}$, el conjunto de criterios es $\mathcal{C}_J = \{días, entmto, srvicio\}$, $\mathcal{V}_J(días)$ es el rango de enteros de 1 a 30 ordenados de mayor a menor (cuantos más días de estadía, mejor), $\mathcal{V}_J(entmto)$ y $\mathcal{V}_J(srvicio)$ son el rango de valores [pésimo, malo, regular, bueno, excelente] ordenados de izquierda a derecha de peor a mejor valor, y el conjunto de evidencia es:*

$$E_J = \left\{ \begin{array}{lll} días(mntñs, 14), & entmto(mntñs, regular), & srvicio(mntñs, pésimo), \\ días(playa, 11), & entmto(playa, bueno), & srvicio(playa, excelente), \\ días(lago, 20), & entmto(lago, malo), & srvicio(lago, bueno), \\ días(jngla, 14), & entmto(jngla, regular), & srvicio(jngla, excelente), \\ días(crcro, 10), & entmto(crcro, bueno), & srvicio(crcro, bueno) \end{array} \right\}$$

Ahora, suponga que en realidad los miembros de la familia Jota no tienen acceso a la información acerca de la calidad del servicio (srvicio) que ofrece el complejo turístico en

la jungla (*jngla*), ni tampoco están seguros de cuantos días (*días*) quedarse en la playa (*playa*). El nuevo conjunto de evidencia sería:

$$E'_J = \left\{ \begin{array}{lll} \text{días}(\text{mntñs}, 14), & \text{entmto}(\text{mntñs}, \text{regular}), & \text{svccio}(\text{mntñs}, \text{pésimo}), \\ & \text{entmto}(\text{playa}, \text{bueno}), & \text{svccio}(\text{playa}, \text{excelente}), \\ \text{días}(\text{lago}, 20), & \text{entmto}(\text{lago}, \text{malo}), & \text{svccio}(\text{lago}, \text{bueno}), \\ \text{días}(\text{jngla}, 14), & \text{entmto}(\text{jngla}, \text{regular}), & \\ \text{días}(\text{crcro}, 10), & \text{entmto}(\text{crcro}, \text{bueno}), & \text{svccio}(\text{crcro}, \text{bueno}) \end{array} \right\}$$

La Figura 4.13 muestra una captura de pantalla de la aplicación que será presentada en el Capítulo 5 en la que se puede observar el grafo de argumentos correspondiente al MEPM $\langle Ag_J, Args'_J, \triangleright_J, \blacktriangleright_J \rangle$, donde $Ag_J = \{c, m, p, t\}$ es el conjunto de agentes que representan a los miembros de la familia Jota, $\triangleright_J = [\triangleright_c, \triangleright_m, \triangleright_p, \triangleright_t]$ es la lista de ordenes de importancia entre las DePref-rules correspondiente a cada uno de los agentes, $\blacktriangleright_J = \{c \blacktriangleright m, c \blacktriangleright p, c \blacktriangleright t, m \blacktriangleright t, p \blacktriangleright t\}$ es el orden de prioridad entre los agentes (ver Ejemplo 3.1.2), y donde $Args'_J$ es el conjunto de todos los argumentos de preferencia que se pueden construir a partir del escenario $\langle \mathcal{D}_J, E'_J, \mathcal{C}_J, \mathcal{V}_J \rangle$ y el conjunto de DePref-rules \mathcal{R}_J del Ejemplo 2.2.4. Note que la cantidad de argumentos generados es bastante menor que la cantidad de argumentos generados a partir de $\langle \mathcal{D}_J, \mathcal{C}_J, \mathcal{V}_J, E_J \rangle$ y \mathcal{R}_J (ver Figura 3.6). En particular, observe que para este nuevo escenario, el único argumento cuya conclusión hace referencia a la preferencia entre playa y jngla es $\langle r_2(\text{playa}, \text{jngla}), \text{pref}(\text{playa}, \text{jngla}) \rangle$ (identificado como A9 en la Figura 4.13). Recuerde que en el escenario anterior, los argumentos generados cuya conclusión hacen referencia a la preferencia entre playa y jngla son todos aquellos mostrados en el grafo de argumentos de la Figura 4.9. Al no contar con la información de la cantidad de días, que era el punto más débil del viaje a la playa, y al no contar con la información acerca de la calidad de servicio del viaje a la jungla, que era el punto más fuerte de esta alternativa, sólo es posible comparar estas alternativas por medio del nivel de entretenimiento (*entmto*) que ofrecen. Es por ese motivo que sólo se genera el argumento $\langle r_2(\text{playa}, \text{jngla}), \text{pref}(\text{playa}, \text{jngla}) \rangle$, ya que

$$r_2 = \text{better}(X, Y, \text{entmto}) \Rightarrow \text{pref}(X, Y)$$

es la única DePref-rule dentro de \mathcal{R}_J que evalúa únicamente el nivel de entretenimiento.

creer que *jngla* es preferida a *lago*, por lo que la segunda alternativa termina siendo considerada mejor que antes. La Figura 4.14 muestra el grafo explicativo que se genera a partir del conjunto de conclusiones garantizadas que se obtienen del MEPM de la Figura 4.13. Observe que ahora *playa* es la alternativa seleccionada ya que es preferida a todas las demás. Además, en comparación a la relación de preferencia que se obtenía al considerar el conjunto de evidencia completa (ver Figura 4.3), observe que ahora *jngla* ya no es preferida a *lago*, ni a *crcro*, sino que ahora, estas tres alternativas son equivalentes.

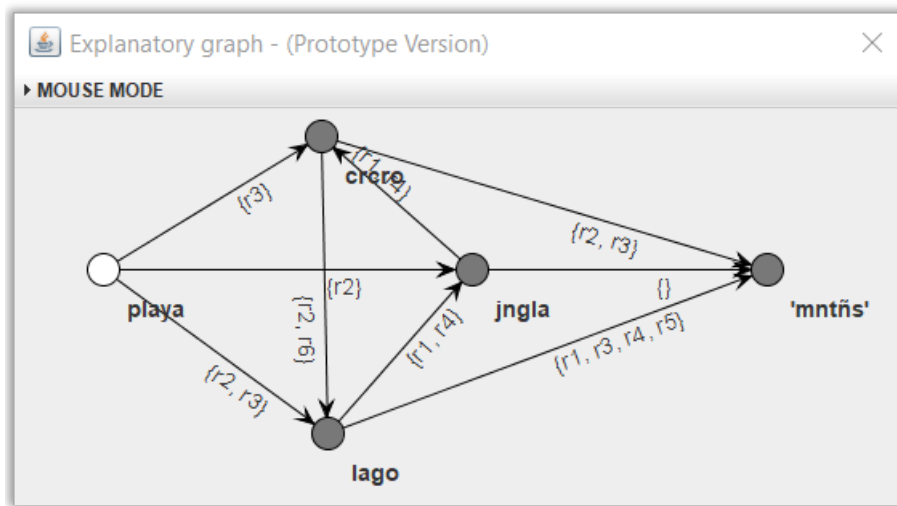


Figura 4.14: Captura de pantalla de la aplicación que será presentada en el Capítulo 5 en la que se puede observar el grafo explicativo que se genera a partir de las conclusiones garantizadas que se obtienen del MEPM de la Figura 4.13.

4.6. Caracterización del conjunto de argumentos ante cambios en la evidencia o en las *DePref*-rules

En esta sección, se analizará cómo un cambio en la evidencia o en el conjunto de *DePref*-rules puede afectar al conjunto de argumentos previamente generados. Se probará formalmente que si una nueva pieza de evidencia es agregada al conjunto de evidencia (Proposición 4.6.1), o si una nueva regla es agregada al conjunto de *DePref*-rules (Proposición 4.6.2), no es necesario eliminar ni modificar los argumentos computados previos al cambio. Más aún, se identificarán las *DePref*-rules y las alternativas con las que de-

ben instanciarse para generar los argumentos que se pueden construir luego de agregar la nueva pieza de evidencia (Proposición 4.6.3), o la nueva *DePref*-rule (Proposición 4.6.4). Además, se indicará una forma de identificar aquellos argumentos que deberán ser eliminados cuando una pieza de evidencia (Proposición 4.6.5) o una *DePref*-rule (Proposición 4.6.6) es removida. A partir de estas propiedades, es posible reducir el computo que se llevaría a cabo durante la etapa de generación de argumentos ante un cambio en la evidencia o en el conjunto de *DePref*-rules. Esto se debe a que las propiedades antes mencionadas aseguran que los cambios sólo afectarán a un grupo reducido de argumentos, por lo que resulta innecesario re-computar nuevamente todos los argumentos, y además, limitan la cantidad de *DePref*-rules y/o alternativas que se deben analizar para construir los nuevos argumentos. Esto puede resultar de suma importancia en situaciones como las del Ejemplo 4.1.3, en donde las condiciones del entorno son muy dinámicas, o incluso para ahorrar tiempo de computo a la hora de reparar errores que se hayan cometido durante la etapa de carga de la evidencia o el conjunto de *DePref*-rules, como por ejemplo, que se haya omitido una *DePref*-rule, o que haya habido un error de tipeo al cargar los valores de alguna pieza de evidencia.

Por ejemplo, suponga que se desea modificar el comportamiento del robot anti-incendios introduciendo una nueva *DePref*-rule. Como será demostrado a continuación, este cambio no afectará a los argumentos computados previamente, por lo que no es necesario volver a ejecutar el proceso de generación de argumentos desde cero. En lugar de esto, el proceso podrá ejecutarse considerando sólo la nueva regla, y agregando, al conjunto de argumentos pre-existente, los argumentos que puedan ser generados a partir de esta nueva regla. De forma similar, suponga que antes de que el robot pueda decidir cuál foco ígneo apagar primero, la intensidad del fuego en uno de ellos aumenta. Está claro que este cambio afecta en cómo este foco se compara con los demás (aunque sólo en cuanto a la intensidad). Sin embargo, como será mostrado más adelante, este cambio no afecta a cómo los demás focos se comparan entre ellos. Esto significa que una porción de los argumentos calculados, antes de que el cambio ocurra, seguirán existiendo luego de haber percibido dicho cambio. Por lo tanto, en este caso el proceso de generación de argumentos también puede restringirse para que tenga en cuenta sólo las alternativas cuya descripción cambió, y sólo, las *DePref*-rules que evalúen el criterio correspondiente al valor que cambió.

Incorporando información

La siguiente proposición indica que al agregar una nueva pieza de evidencia todos los argumentos generados previamente seguirán existiendo. Esta propiedad se deduce de forma directa de la definición de la semántica de las *DePref*-rules (Definición 2.2.3) y de la definición de argumento de preferencia (Definición 2.3.1).

Proposición 4.6.1. *Sean $\Pi = \langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$ y $\Pi^+ = \langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E^+ \rangle$ dos escenarios de elección tales que $E^+ = E \cup \{c(d, v)\}$, y sea \mathcal{R} un conjunto de *DePref*-rules. Si $Args$ es el conjunto de todos los argumentos que se pueden construir a partir de Π y \mathcal{R} , y $Args^+$ es el conjunto de todos los argumentos que se pueden construir a partir de Π^+ y \mathcal{R} , entonces $Args \subseteq Args^+$.*

Demostración. Sean $\Pi = \langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$ y $\Pi^+ = \langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E^+ \rangle$ dos escenarios de elección tales que $E^+ = E \cup \{c(d, v)\}$, y sea \mathcal{R} un conjunto de *DePref*-rules. Como $E \subset E^+$, por la Definición 2.2.3, si E satisface a una *DePref*-rule instanciada $r(d_x, d_y)$, E^+ también debe satisfacerla. Luego, por la Definición 2.3.1, si un argumento es construido a partir de Π y $r \in \mathcal{R}$, entonces este argumento también debe poder ser construido a partir de Π^+ y r . Por lo tanto, si $Args$ es el conjunto de todos los argumentos que se pueden construir a partir de Π y \mathcal{R} , y $Args^+$ es el conjunto de todos los argumentos que se pueden construir a partir de Π^+ y \mathcal{R} , se tiene que $Args \subseteq Args^+$. \square

La proposición que se presenta a continuación indica que al agregar una nueva *DePref*-rule todos los argumentos generados previamente seguirán existiendo.

Proposición 4.6.2. *Sea $\Pi = \langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$ un escenario de elección, y sean \mathcal{R} y \mathcal{R}^+ dos conjuntos de *DePref*-rules tales que $\mathcal{R}^+ = \mathcal{R} \cup \{r'\}$. Si $Args$ es el conjunto de todos los argumentos que se pueden construir a partir de Π y \mathcal{R} , y $Args^+$ es el conjunto de todos los argumentos que se pueden construir a partir de Π y \mathcal{R}^+ , entonces $Args \subseteq Args^+$.*

Demostración. Sea $\Pi = \langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$ un escenario de elección, y sean \mathcal{R} y \mathcal{R}^+ dos conjuntos de *DePref*-rules tales que $\mathcal{R}^+ = \mathcal{R} \cup \{r'\}$. Sea $Args$ el conjunto de todos los argumentos que se pueden construir a partir de Π y \mathcal{R} , y sea $Args^+$ el conjunto de todos los argumentos que se pueden construir a partir de Π y \mathcal{R}^+ . Sea \mathcal{A} un argumento construido a partir de Π y $r \in \mathcal{R}$, luego $\mathcal{A} \in Args$. Como $r \in \mathcal{R}$ y $\mathcal{R} \subset \mathcal{R}^+$, se tiene que $\mathcal{A} \in Args^+$. Por lo tanto, $Args \subseteq Args^+$ \square

Ejemplo 4.6.1. *Considere nuevamente el escenario de elección $\langle \mathcal{D}_B, \mathcal{C}_B, \mathcal{V}_B, E_B \rangle$ y el conjunto de DePref-rules \mathcal{R}_B del Ejemplo 4.1.3. Suponga que se desea modificar el comportamiento del robot anti-incendios agregando la siguiente DePref-rule:*

$$r_{16bis} = \left\{ \begin{array}{l} equal(X, Y, area), equal(X, Y, int), \\ better(X, Y, dst) \text{ min_dist}(X, Y, dst, 3) \end{array} \right\} \Rightarrow pref(X, Y)$$

Luego, al ejecutar el proceso de generación de argumentos tomando como entradas a $\langle \mathcal{D}_B, \mathcal{C}_B, \mathcal{V}_B, E_B \rangle$ y a $\mathcal{R}_B^+ = \mathcal{R}_B \cup \{r_{16bis}\}$, se obtiene como resultado el conjunto de argumentos $Args_B^+ = Args_B \cup \{\langle r_{16bis}(t_1, t_4), pref(t_1, t_4) \rangle\}$. Recuerde que

$$E_B = \left\{ \begin{array}{llll} area(t_1, 1), & int(t_1, baja), & dst(t_1, 2), & urg(t_1, no), \\ area(t_2, 3), & int(t_2, media), & dst(t_2, 1), & urg(t_2, no), \\ area(t_3, 2), & int(t_3, alta), & dst(t_3, 3), & urg(t_3, no), \\ area(t_4, 1), & int(t_4, baja), & dst(t_4, 6), & urg(t_4, si) \end{array} \right\}$$

Por lo tanto, las únicas alternativas que satisfacen los predicados $equal(X, Y, area)$ y $equal(X, Y, int)$ son t_1 y t_4 . La Figura 4.15 muestra todos los argumentos para $pref(t_1, t_4)$ y su complemento que son generados antes y después del cambio.

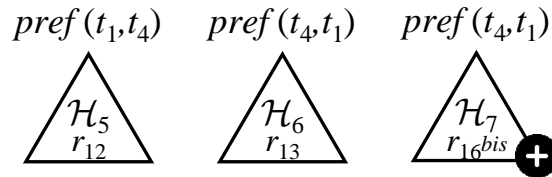


Figura 4.15: Argumentos para $pref(t_1, t_4)$ y su complemento que son construidos a partir de $\langle \mathcal{D}_B, \mathcal{C}_B, \mathcal{V}_B, E_B \rangle$ y $\mathcal{R}_B^+ = \mathcal{R}_B \cup \{r_{16bis}\}$. El único argumento generado después del cambio está marcado con el símbolo “+”, el resto de los argumentos ya habían sido generados antes de haber agregado la nueva regla a \mathcal{R}_B .

La próxima proposición especifica cuáles DePref-rules deben ser re-evaluadas, y cómo deben ser instanciadas, en caso de que una nueva pieza de evidencia sea agregada.

Proposición 4.6.3. Sean $\Pi = \langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$ y $\Pi^+ = \langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E^+ \rangle$ dos escenarios de elección tales que $E^+ = E \cup \{c(d, v)\}$, y sea \mathcal{R} un conjunto de *DePref*-rules. Sea $Args$ el conjunto de todos los argumentos que se pueden generar a partir de Π y \mathcal{R} , y sea $Args^+$ el conjunto de todos los argumentos que se pueden generar a partir de Π^+ y \mathcal{R} . Si $\langle r(d_x, d_y), pref(d_x, d_y) \rangle \in (Args^+ \setminus Args)$, entonces r tiene una premisa que evalúa a c y $d \in \{d_x, d_y\}$.

Demostración. Sean $\Pi = \langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$ y $\Pi^+ = \langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E^+ \rangle$ dos escenarios de elección tales que $E^+ = E \cup \{c(d, v)\}$, y sea \mathcal{R} un conjunto de *DePref*-rules. Sea $Args$ el conjunto de todos los argumentos que se pueden generar a partir de Π y \mathcal{R} , y sea $Args^+$ el conjunto de todos los argumentos que se pueden generar a partir de Π^+ y \mathcal{R} . Si $\langle r(d_x, d_y), pref(d_x, d_y) \rangle \in (Args^+ \setminus Args)$, de acuerdo con la Definición 2.3.1, E^+ satisface a $r(d_x, d_y)$, pero E no. Siguiendo la Definición 2.2.3, para que esto suceda, $r(d_x, d_y)$ debe poseer una premisa tal que E^+ la satisfaga, pero E no. Esto sólo sería posible si $c(d, v)$ es necesario para satisfacer dicha premisa, lo que implica que la premisa debe evaluar a c y que $d \in \{d_x, d_y\}$. \square

La siguiente proposición asegura que al agregar una nueva *DePref*-rule todo argumento que pueda aparecer luego de este cambio, será generado a partir de esta nueva *DePref*-rule, tal y como ocurre en el Ejemplo 4.6.1.

Proposición 4.6.4. Sea $\Pi = \langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$ un escenario de elección, y sean \mathcal{R} y \mathcal{R}^+ dos conjuntos de *DePref*-rules tales que $\mathcal{R}^+ = \mathcal{R} \cup \{r'\}$. Sea $Args$ el conjunto de todos los argumentos que se pueden construir a partir de Π y \mathcal{R} , y sea $Args^+$ el conjunto de todos los argumentos que se pueden construir a partir de Π y \mathcal{R}^+ . Para todo argumento $\mathcal{A} \in (Args^+ \setminus Args)$ se cumple que $rule(\mathcal{A}) = r$.

Demostración. Sea $\Pi = \langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$ un escenario de elección, y sean \mathcal{R} y \mathcal{R}^+ dos conjuntos de *DePref*-rules tales que $\mathcal{R}^+ = \mathcal{R} \cup \{r'\}$. Sea $Args$ el conjunto de todos los argumentos que se pueden construir a partir de Π y \mathcal{R} , y sea $Args^+$ el conjunto de todos los argumentos que se pueden construir a partir de Π y \mathcal{R}^+ . De acuerdo con la Definición 2.3.1, si existe un argumento $\mathcal{A} \in (Args^+ \setminus Args)$, \mathcal{A} debe ser construido a partir de una *DePref*-rule que pertenezca a \mathcal{R}^+ , pero no pertenezca a \mathcal{R} . Dado que $\mathcal{R}^+ \setminus \mathcal{R} = \{r'\}$, luego $rule(\mathcal{A}) = r$. \square

La importancia de las proposiciones anteriores radica en que estas aseguran que no es necesario evaluar todas las *DePref*-rules nuevamente a la hora de agregar una pieza de

evidencia o agregar una nueva *DePref*-rule. Más aún, identifican claramente cuáles son las *DePref*-rule que debe ser evaluadas, lo que permite optimizar el proceso de generación de argumentos.

Removiendo información

Finalmente, las dos proposiciones que siguen a continuación especifican cómo identificar los argumentos que deben ser removidos cuando una pieza de evidencia o *DePref*-rule es eliminada.

Proposición 4.6.5. *Sean $\Pi = \langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$ y $\Pi^- = \langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E^- \rangle$ dos escenarios de elección tales que $E^- = E \setminus \{c(d, v)\}$, y sea \mathcal{R} un conjunto de *DePref*-rules. Sea $Args$ el conjunto de todos los argumentos que se pueden construir a partir de Π y \mathcal{R} , y sea $Args^-$ el conjunto de todos los argumentos que se pueden generar a partir de Π^- y \mathcal{R} . Si $\langle r(d_x, d_y), pref(d_x, d_y) \rangle \in (Args \setminus Args^-)$, entonces r tiene una premisa que evalúa a c y $d \in \{d_x, d_y\}$.*

Demostración. Por hipótesis, $E = E^- \cup \{c(d, v)\}$. Entonces, por la Proposición 4.6.1, $Args^- \subseteq Args$. Luego, por la Proposición 4.6.3, para cada argumento $\langle r(d_x, d_y), pref(d_x, d_y) \rangle \in (Args \setminus Args^-)$ se cumple que r tiene una premisa que evalúa a c y $d \in \{d_x, d_y\}$. \square

Proposición 4.6.6. *Sea $\Pi = \langle \mathcal{D}, \mathcal{C}, \mathcal{V}, E \rangle$ un escenario de elección, y sean \mathcal{R} y \mathcal{R}^- dos conjuntos de *DePref*-rules tales que $\mathcal{R}^- = \mathcal{R} \setminus \{r\}$. Sea $Args$ el conjunto de todos los argumentos que se pueden construir a partir de Π y \mathcal{R} , y sea $Args^-$ el conjunto de todos los argumentos que se pueden construir a partir de Π y \mathcal{R}^- . Para todo argumento $\mathcal{A} \in (Args \setminus Args^-)$ se cumple que $rule(\mathcal{A}) = r$.*

Demostración. Por hipótesis, $\mathcal{R} = \mathcal{R}^- \cup \{r\}$. Entonces, por la Proposición 4.6.2, $Args^- \subseteq Args$. Luego, por la Proposición 4.6.4, si $\mathcal{A} \in (Args \setminus Args^-)$ se cumple que $rule(\mathcal{A}) = r$. \square

Junto con el resto de las proposiciones mencionadas anteriormente, estas últimas dos proposiciones permiten caracterizar cómo se ve modificado el conjunto de argumentos frente a cambios en la evidencia o en el conjunto de *DePref*-rules ya que es posible identificar cuáles argumentos eliminar y cuáles *DePref*-rules evaluar para obtener el nuevo conjunto de argumentos, acorde a los cambios realizados.

Editando información

Gracias a las propiedades estudiadas anteriormente, ante una edición de la información que se tenía, no es necesario generar todos los argumentos nuevamente ya que se puede asegurar que parte de los argumentos generados anteriormente seguirán estando presentes, aún después del cambio.

Para aprovechar esta ventaja debe pensarse el cambio en la evidencia cómo un proceso de dos pasos: en el primero se remueve la pieza de evidencia o *DePref*-rule actual, y en el segundo se agrega la pieza de evidencia o *DePref*-rule que reemplaza a la anterior. Con esto en mente, se debe proceder de la siguiente manera: primero, se deben identificar los argumentos que deben ser eliminados debido a la remoción (ver Proposición 4.6.5 y Proposición 4.6.6); y segundo, se debe ejecutar el proceso de generación de argumentos, pero restringido para considerar sólo las alternativas y *DePref*-rules que sean necesarias (ver Proposición 4.6.3 y Proposición 4.6.4).

Ejemplo 4.6.2. *Considere nuevamente el escenario de elección $\langle \mathcal{D}_B, \mathcal{C}_B, \mathcal{V}_B, E_B \rangle$ y el conjunto de *DePref*-rules $\mathcal{R}^+ B = \mathcal{R}_B \cup \{r_{16^{bis}}\}$ del Ejemplo 4.6.1. Suponga que antes de ejecutar el proceso de evaluación dialéctico, el fuego de t_2 aumenta su intensidad de media a alta (recuerde que $int(t_2, media) \in E_B$).*

*Por la Proposición 4.6.5, se sabe que sólo pueden verse afectados aquellos argumentos que fueron construidos a partir de una *DePref*-rule que evalúe la intensidad y que haya sido instanciada con t_2 . Recuerde que las *DePref*-rules pertenecientes a \mathcal{R}_B^+ que evalúan la intensidad son:*

$$\begin{aligned}
 r_{10} &= \left\{ \text{better}(X, Y, int) \right\} \Rightarrow \text{pref}(X, Y) \\
 r_{14} &= \left\{ \begin{array}{l} \text{worse}(X, Y, dst), \text{max_dist}(X, Y, dst, 3), \\ \text{better}(X, Y, area), \text{better}(X, Y, int) \end{array} \right\} \Rightarrow \text{pref}(X, Y) \\
 r_{16} &= \left\{ \begin{array}{l} \text{better}(X, Y, area), \text{better}(X, Y, int), \\ \text{better}(X, Y, dst), \text{min_dist}(X, Y, dst, 2) \end{array} \right\} \Rightarrow \text{pref}(X, Y) \\
 r_{16^{bis}} &= \left\{ \begin{array}{l} \text{equal}(X, Y, area), \text{equal}(X, Y, int), \\ \text{better}(X, Y, dst), \text{min_dist}(X, Y, dst, 3) \end{array} \right\} \Rightarrow \text{pref}(X, Y)
 \end{aligned}$$

*La Figura 4.16 muestra los argumentos relacionados con t_2 que se obtendrían antes y después de remover $int(t_2, media)$ de E_B . Note que sólo son eliminados aquellos argumentos cuyas *DePref*-rules evalúan la intensidad, estos son \mathcal{H}_9 , \mathcal{H}_{11} , \mathcal{H}_{13} y \mathcal{H}_{19} .*

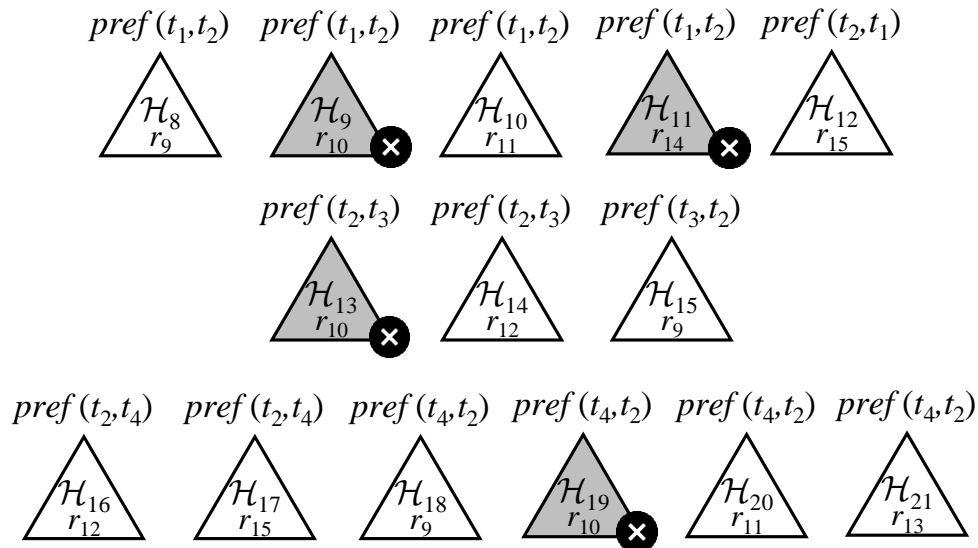


Figura 4.16: Argumentos relacionados con t_2 que se obtendrían antes y después de remover $int(t_2, media)$ de E_B . Los argumentos que deben ser eliminados están marcados con el símbolo “x” y coloreados de gris oscuro.

Luego, por la Proposición 4.6.3, se sabe que los nuevos argumentos que vayan a ser generados, luego de agregar $int(t_2, alta)$ a $E_B^- = E_B \setminus \{int(t_2, media)\}$, deberán serlo a partir de aquellas DePref-rules que evalúen la intensidad, y utilizando t_2 para instanciarlas. De estas reglas, las únicas que instanciadas con t_2 son satisfechas por el nuevo conjunto de evidencia $E_B^+ = E_B^- \cup \{int(t_2, alta)\}$ son $r_{10}(t_1, t_2)$, $r_{10}(t_4, t_2)$, y $r_{14}(t_1, t_2)$. La Figura 4.17 muestra los argumentos relacionados con t_2 que se obtendrían antes y después de agregar $int(t_2, alta)$ a $E_B^- = E_B \setminus \{int(t_2, media)\}$. Note que sólo se agregan los argumentos \mathcal{H}_{22} , \mathcal{H}_{23} y \mathcal{H}_{24} cuyas reglas coinciden con las mencionadas anteriormente. El resto de los argumentos se mantienen igual que antes.

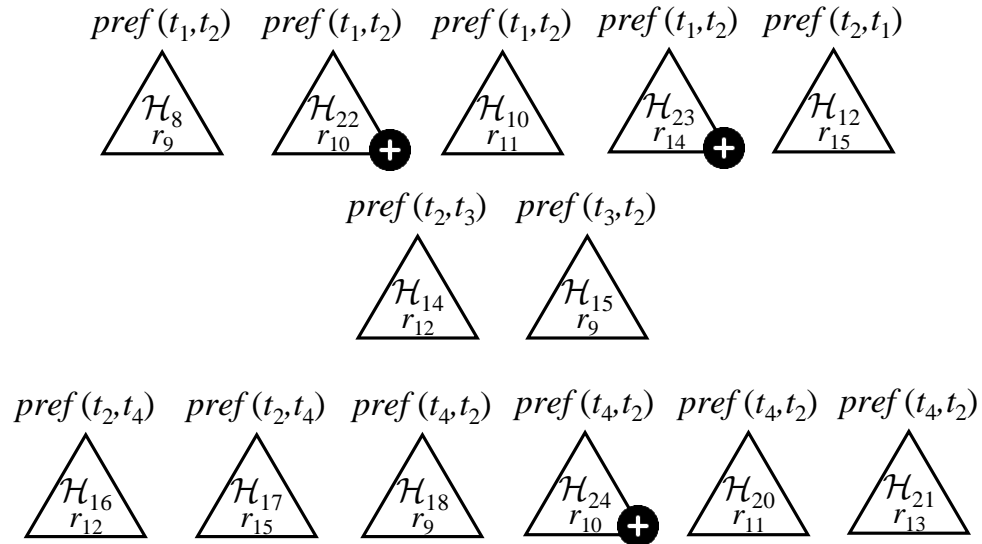


Figura 4.17: Argumentos relacionados con t_2 que se obtendrían antes y después de agregar $int(t_2, alta)$ a $E_B^- = E_B \setminus \{int(t_2, media)\}$. Los nuevos argumentos están marcados con el símbolo “+”.

4.7. Resumen

En este capítulo se explicaron los procesos que forman parte de la etapa de selección de alternativas. En esta etapa, se construye un orden de preferencia entre las alternativas (\mathcal{D}) en base al conjunto de conclusiones garantizadas (W), que fue obtenido a partir del proceso de evaluación dialéctico. Como se explicó en la Sección 4.1, este orden es utilizado por el proceso de selección de alternativas, el cual distingue los elementos maximales de dicho orden, y con ellos construye el conjunto de alternativas seleccionadas (S), es decir, las alternativas que el marco de trabajo recomienda a los agentes. Además, al final de la Sección 4.1 se caracteriza el conjunto de alternativas seleccionadas.

El otro de los procesos que forma parte de la etapa de selección de alternativas, es el proceso de generación de explicaciones. Este proceso, produce una composición de explicaciones gráficas, que junto con el conjunto de alternativas seleccionadas, constituyen la salida del proceso de decisión. Como fue explicado en la Sección 4.3, la explicación de los resultados es tan importante como los resultados en sí, ya que estas ayudan a fortalecer la confianza de los usuarios en el sistema, y que de esta manera estén dispuestos a

utilizarlo. Las explicaciones se presentan por medio de un grafo, llamado grafo explicativo (ver Definición 4.3.2), con el cual se puede visualizar el orden de preferencia entre las alternativas. Cada uno de los nodos de este grafo representa una alternativa, y los arcos del grafo representan la relación de preferencia que se induce a partir del conjunto de conclusiones garantizadas. Esto es, (d_x, d_y) es un arco del grafo explicativo si y sólo si $pref(d_x, d_y) \in W$. Además, cada arco (d_x, d_y) está asociado a tres elementos: i) el conjunto de *DePref*-rules con las que se construyeron los argumentos aceptados cuya conclusión es $pref(d_x, d_y)$, ii) la explicación dialéctica asociada a la conclusión $pref(d_x, d_y)$ (ver Definición 4.3.1), y iii) un grafo de argumentos (ver Definición 3.1.7) reducido en el que sólo se muestran los argumentos que aparecen en la explicación dialéctica. Estos elementos permiten que los usuarios puedan obtener una explicación exhaustiva acerca de por qué una alternativa es seleccionada, por qué una alternativa es preferida a otra, o cómo fue construido el orden de preferencia.

En la Sección 4.4 se analiza con mayor profundidad la influencia que tiene cada agente sobre el comportamiento del formalismo. En particular, se analiza un ejemplo de aplicación en el que no existe un agente que tenga más prioridad que todos los demás. Mediante dicho ejemplo se demostró la capacidad del marco de trabajo para tener en cuenta los intereses de todos los agentes, incluyendo aquellos de menor prioridad.

En este capítulo también se estudió el comportamiento del marco de trabajo al considerar escenarios de elección en los que el conjunto de evidencia no es completo, es decir, cuando al menos una de las alternativas no tiene asociado un valor con respecto a alguno de los criterios. Como se explica a lo largo de la Sección 4.5, esto no es un impedimento para el correcto funcionamiento del marco de trabajo, pero es una cuestión que debe ser tomada en cuenta ya que cuanto menos información se tenga de una alternativa, más difícil será compararla con las demás.

Finalmente, en la Sección 4.6 se caracterizó el efecto que tienen los cambios en la evidencia y/o en el conjunto de *DePref*-rules sobre el conjunto de argumentos de preferencia. Como se explica a lo largo de dicha sección, estos cambios afectan a un número reducido de argumentos que pueden ser fácilmente identificados. A partir de las propiedades que allí se presentan, es posible reducir el computo que se requiere a la hora de ejecutar el proceso de generación de argumentos ya que al identificar los argumentos afectados por el cambio, resulta innecesario re-computar nuevamente todos los argumentos.

Capítulo 5

Implementación del formalismo

En este capítulo, se describirá la aplicación desarrollada con el propósito de implementar el formalismo para toma de decisiones grupales multi-criterio introducido en los capítulos anteriores. Esta aplicación es de código abierto y puede ser descargada desde un repositorio de GitHub¹. El motor de decisión fue implementado usando SWI-Prolog (versión 7.6.4), mientras que la interfaz gráfica fue implementada usando Java (versión 7). Para la conexión entre Java y SWI-Prolog se utilizó la biblioteca de terceros JPL7² la cual provee una interfaz bidireccional entre JAVA y SWI-Prolog.

La aplicación ha ido evolucionando desde su concepción. En principio, fue desarrollada para implementar el formalismo propuesto en [BBTG19] el cual sólo trataba con problemas de decisión individual. Allí, también se muestra un prototipo en papel de los grafos explicativos. La primera versión de la aplicación se describe en [BBTG20]. En dicho trabajo se muestra una primer versión de las pantallas para interactuar con el sistema, tanto para la definición del problema, como para ahondar en las explicaciones. Luego, en [BBTG23], se muestra una versión de la aplicación capaz de abordar problemas de decisión con múltiples agentes, y se analiza con mayor nivel de detalle las posibles interacciones con las explicaciones. Cabe aclarar que para la versión presentada en esta tesis, se incorporan las etiquetas sobre los arcos del grafo de argumentos con el objetivo de que sea más claro entender la relación de derrota entre ellos.

La Figura 5.1 muestra la pantalla principal de la aplicación. En la parte izquierda, se encuentran todos los botones con los que el usuario puede interactuar para cargar la

¹https://github.com/buronbrarda/madf_with_dprefrules/tree/multi-agent

²<https://jpl7.org/index>

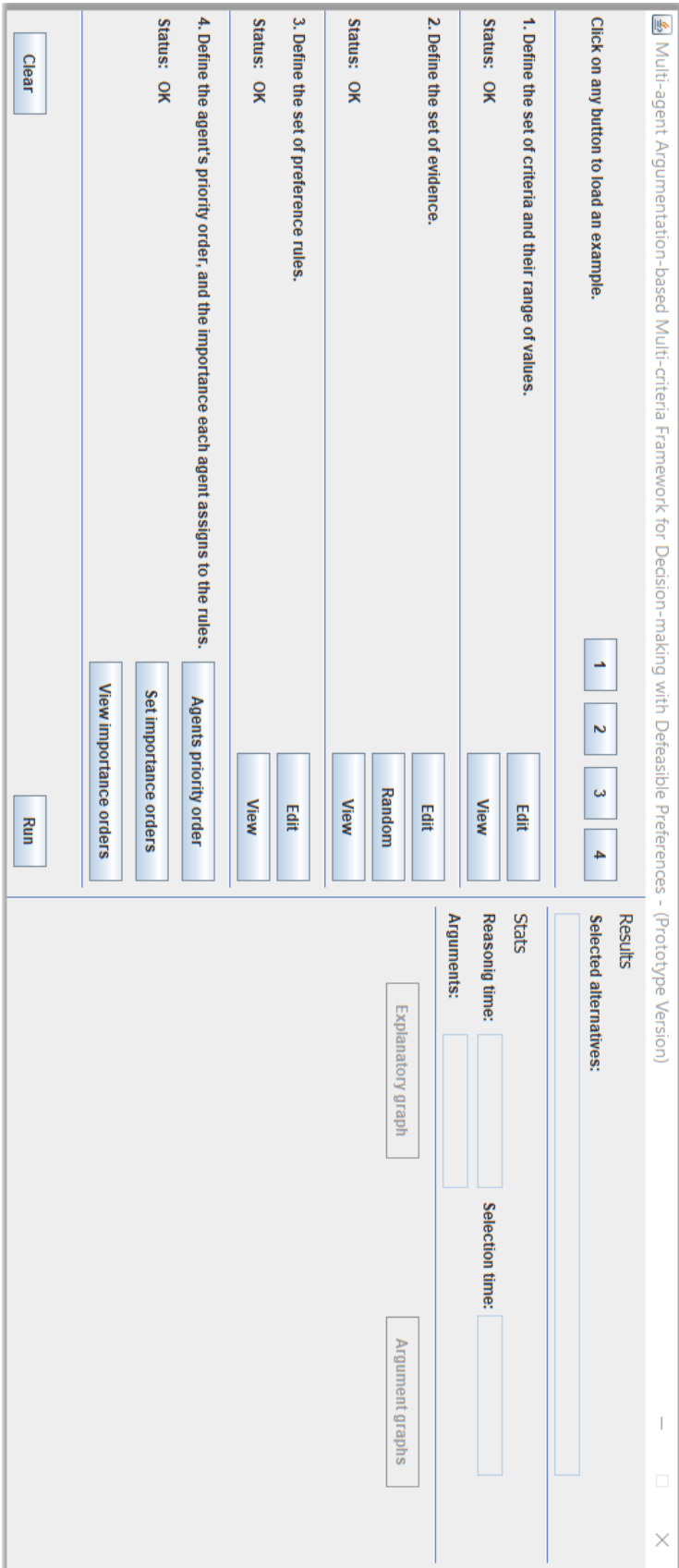


Figura 5.1: Pantalla principal de la aplicación.

información correspondiente al problema de decisión que se desee modelar. En la parte de la derecha, se encuentra la pantalla en la que se mostrarán los resultados arrojados por el motor de decisión, así como también otros datos como el tiempo que el sistema tardó en calcular la solución o la cantidad de argumentos que fueron generados. Además, la pantalla de resultados ofrece acceso al grafo explicativo y a un grafo de argumentos que incluye todos los argumentos generados.

5.1. Definiendo el problema de decisión

La pantalla de carga de información ha sido diseñada para guiar al usuario durante todo el proceso de modelado. El usuario deberá especificar los datos del problema de decisión siguiendo una serie de pasos. Lo primero que tendrá que hacer es especificar los criterios y sus respectivos rangos de valores. Luego, deberá definir el conjunto de evidencia. Seguido a eso, deberá establecer el conjunto de *DePref*-rules, y finalmente definir los ordenes de importancia y las prioridades entre los agentes. Cada uno de los pasos es dependiente de los anteriores, por lo que al producirse un error en alguno de estos, el usuario no podrá continuar con los pasos posteriores hasta haber corregido el error. La Figura 5.2 muestra la pantalla principal de aplicación luego de haber cargado los datos correspondientes al escenario de elección (ver Ejemplo 2.1.1), *DePref*-rules (ver Ejemplo 2.2.4), intereses y prioridades (ver Ejemplo 3.1.2) del problema de decisión que enfrentan los miembros de la familia Jota.

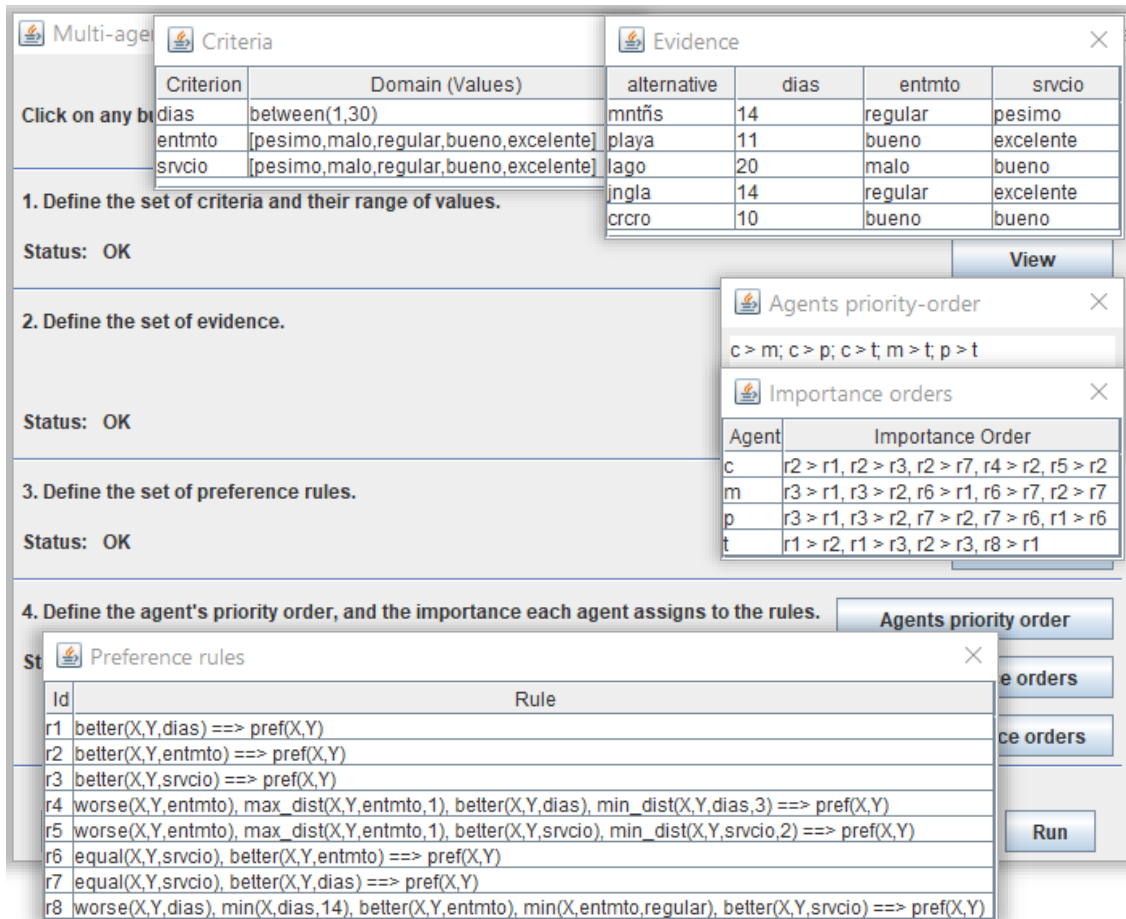


Figura 5.2: Pantalla principal de la aplicación mostrando los datos que modelan el problema de decisión que enfrentan los miembros de la familia Jota.

Como se mencionó antes, en el primer paso, el usuario debe especificar todos los criterios de comparación del escenario de elección, junto con sus rangos de valores legales (ver ventana *Criteria* de la Figura 5.2). Una vez cargados estos datos, la aplicación verificará que no existan errores de sintaxis. En cuanto a los nombres de los elementos, la única restricción es que no contengan espacios y que sean todos distintos entre sí. En cuanto a los rangos de valores, estos pueden ser de dos clases: numéricos, como es el caso de *días*; o discretos, como es el caso de *entmto* y *svrcio*. Los rangos numéricos se definen haciendo uso de la forma `better(worst, best)`, donde $\text{worst} \in \mathbb{R}$ especifica el peor valor posible y $\text{best} \in \mathbb{R}$ el mejor valor posible. Por ejemplo, con el rango `between(1, 30)` (utilizado para caracterizar los valores posibles correspondientes a la cantidad de días de viaje) se especifica que la cantidad de días puede variar entre 1 y 30, y que a mayor cantidad de días, mejor es

la alternativa. Sólo aquellos números entre **worst** y **best** (incluidos) serán considerados valores legales. Si **worst** < **best** entonces, los números de mayor magnitud serán considerados mejores que aquellos de menor magnitud (como sucede en el caso anterior); en cambio, si **worst** > **best**, los números de menor magnitud serán considerados mejores que aquellos de mayor magnitud. Por ejemplo, imagine que conoce la cantidad de quejas que recibe por día la empresa que organiza cada uno de los viaje, entonces uno podría utilizar el intervalo **between(10, 0)** para expresar que: cuanto menor sea la cantidad de quejas, mejor es la alternativa, como máximo una empresa puede recibir hasta 10 quejas por día, y como mínimo cero. Por otra parte, los rangos discretos se definen utilizando la notación de listas de SWI-Prolog, identificando cada posible valor como un elemento de la lista, tal y como se puede apreciar en la Figura 5.2 con el rango [pésimo, malo, regular, bueno, excelente], asociados a los criterios **entmto** y **svcio**. El orden de los elementos en la lista establece cuán bueno es un valor. Cuanto más cerca del principio se encuentre, es decir, más a la izquierda, peor será el valor. Así, se tiene que (por ejemplo) **pésimo** es peor que **malo**, que estos dos son peores que **regular**, y así siguiendo con el resto de los valores. Si al ejecutar el paso de verificación, la aplicación encuentra algún problema, en **Status** se mostrará el mensaje “**ERROR**” y aparecerá un cuadro de diálogo con un mensaje que ayude al usuario a buscar el error que se cometió. Por ejemplo, en el caso de la Figura 5.3, el error está en que el rango de valores especificado no sigue el formato de una lista de Prolog ya que no posee los corchetes que indican el principio y final de la lista. Cuando no hay errores en la carga, **Status** mostrará el mensaje “**OK**”, y la aplicación habilitará la carga de datos del siguiente paso.

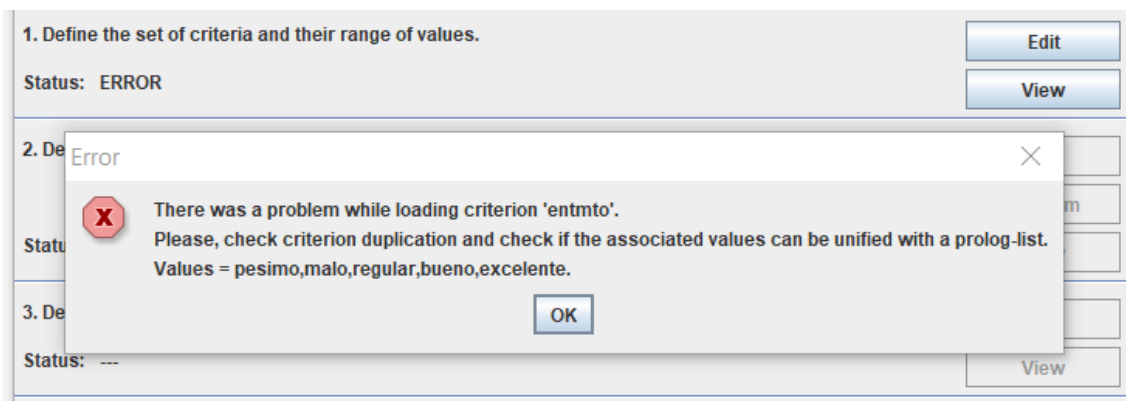
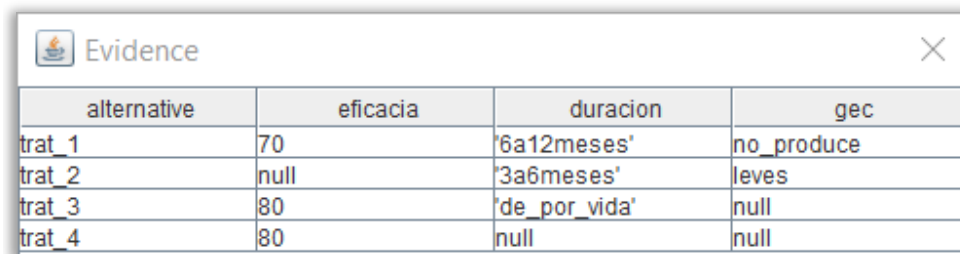


Figura 5.3: Captura de pantalla mostrando el mensaje de error que aparece cuando el proceso de verificación encuentra que el rango de valores no es válido.

En el segundo paso, el usuario debe definir el conjunto de evidencia. Para esto, el usuario tendrá que definir para cada alternativa los valores correspondientes a cada uno de los criterios (ver ventana **Evidence** de la Figura 5.2). Note que en este paso, el conjunto de alternativas es definido de forma implícita al definir los valores correspondientes a cada una de ellas. En caso de que para una alternativa no exista el valor correspondiente a alguno de los criterios, los usuarios deberán indicarlo asignando el valor **null** (ver Figura 5.4). Para que el proceso de verificación no falle, todas las alternativas deben tener un valor asignado para cada uno de los criterios, ya sea un valor dentro del rango de valores posibles, o bien **null** en caso de que no tengan asignado ningún valor (y por ende, el conjunto de evidencia sea incompleto). Además, para este paso, la aplicación no sólo verificará que los valores asignados a las alternativas sean válidos (es decir, que los valores para los criterios coincidan con aquellos definidos en el paso anterior), sino que también, verifica que la evidencia sea consistente. Esto es, que no existan dos valoraciones distintas para una misma alternativa y un mismo criterio (ver Definición 2.1.1). Al igual que antes, si no hay errores, **Status** mostrará el mensaje “OK”, y habilitará la carga de datos del paso siguiente, es decir, el paso 3 (ver Figura 5.1).



alternative	eficacia	duracion	gec
trat_1	70	'6a12meses'	no_produce
trat_2	null	'3a6meses'	leves
trat_3	80	'de_por_vida'	null
trat_4	80	null	null

Figura 5.4: Captura de pantalla mostrando un ejemplo de cómo se ve la tabla de evidencia cuando alguna alternativa no tiene asignado un valor para un dado criterio. Observe que estos valores se corresponden con los del conjunto de evidencia del Ejemplo 4.5.3. Recuerde que para este ejemplo, se desconocía la eficacia de *trat_2*, la gravedad de los efectos colaterales (GEC) de *trat_3*, y la duración y GEC de *trat_4*.

El tercer paso requiere que se cargue el conjunto de *DePref*-rules que codifican las razones detrás de las preferencias de los agentes. Una vez cargadas, la aplicación verificará su sintaxis y que sean coherentes (ver Definición 2.2.1). Al igual que en los pasos anteriores, si todo está correcto, **Status** mostrará el mensaje “OK” y se habilitará la carga de datos

del paso siguiente, mientras que si hubo algún error se mostrará el mensaje “ERROR” y aparecerá un cuadro de diálogo, como el de la Figura 5.5, indicando cuál de las *DePref*-rules cargadas es incorrecta.

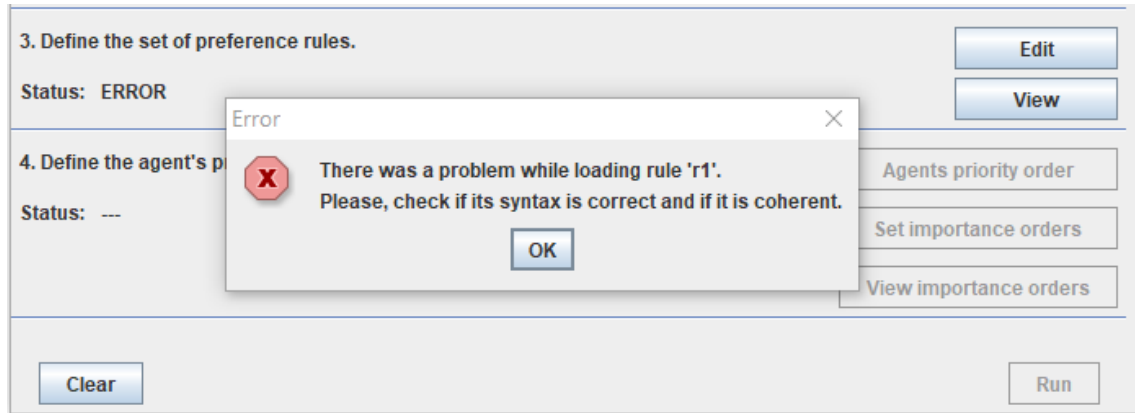


Figura 5.5: Captura de pantalla mostrando el mensaje de error que aparecería si el proceso de verificación determina que una de las *DePref*-rules no es válida. Note que este cartel no hace referencia a la *DePref*-rule r_1 mostrada anteriormente.

Finalmente, en el cuarto y último paso, se deben indicar los órdenes de importancia correspondientes a cada uno de los agentes, así como también, la prioridad entre los agentes. Al igual que lo que ocurrido con las alternativas al cargar el conjunto de evidencia, los agentes son definidos de forma implícita al introducir su orden de importancia (ver ventana **Importance Orders** de la Figura 5.2). Dentro de un orden de importancia, la sentencia $r_i > r_j$ indica que la *DePref*-rule r_i es más importante que la *DePref*-rule r_j . Estas sentencias deben estar separadas por “;”. Por ejemplo, en la Figura 5.2, se puede ver que el agente c (Carolina), considera que $r_2 > r_1$, $r_2 > r_3$, $r_2 > r_7$, $r_4 > r_2$ y $r_5 > r_2$. Una vez cargados los agentes y sus órdenes, será posible especificar la prioridad entre los agentes (ver ventana **Agents priority-order** de la Figura 5.2). Para el orden de prioridad, la sentencia $a > b$ indica que el agente a tiene prioridad sobre el agente b , y las distintas sentencias deben estar separadas por “;”. De acuerdo con el ejemplo de la Figura 5.2, se tiene que $c > m$, $c > p$, $c > t$, $m > t$ y $p > t$. Al finalizar la carga de los datos, la aplicación verificará que en las sentencias de los órdenes de importancia sólo aparezcan los identificadores de las *DePref*-rules que fueron definidas en el paso 3, y que en las sentencias del orden de prioridad entre agentes aparezcan los agentes que

fueron (previamente) definidos junto con los órdenes de importancia. Si no hay ningún error, **Status** mostrará el mensaje “OK” y se habilitará el botón de **Run** con el que se podrá dar arranque al motor de decisión, y así, poder obtener el conjunto de alternativas seleccionadas.

5.2. Explicaciones interactivas

Una de las ventajas más importantes de contar con esta aplicación, es la de poder interactuar por medio de una interfaz gráfica con las explicaciones que brinda el sistema. Esta interfaz fue diseñada para que el usuario pueda hacer tantas consultas como desee, pero sin sobrecargarlo o abrumarlo con información, que podría no ser de su interés. Es por eso, que la aplicación muestra al usuario sólo una mínima cantidad de información, que luego, por medio de interacciones, el usuario podrá ampliar y obtener así toda la información que le resulte necesaria, sin tener que procesar aquello que no le interese. Cabe aclarar, que el concepto de explicaciones interactivas de esta propuesta difiere del presentado en [RCT18, RCBT20, RCB⁺21]. En dichos trabajos, el objetivo principal de las interacciones entre los usuarios y las explicaciones es el de obtener feedback para mejorar los resultados, y no el de indagar acerca del resultado. En los capítulos anteriores se presentaron imágenes mostrando la implementación de los grafos explicativos, las explicaciones dialécticas y los grafos de argumentos (ver Figura 4.7 y Figura 4.8). A continuación, se explicará cómo interactuar con estas pantallas.

Además de las interacciones básicas, como mover y seleccionar los nodos, una de las formas de interactuar más interesantes que poseen los grafos explicativos es la que permite a los usuarios ahondar en la explicación. Mediante esta interacción, la aplicación provee la justificación de por qué existe una preferencia entre una alternativa y otra, es decir, la justificación de la existencia de cada uno de los arcos del grafo explicativo. Según la Definición 4.3.2, un arco desde una alternativa d_x a otra d_y en un grafo explicativo existe si y sólo si $pref(d_x, d_y)$ está garantizada. Por lo tanto, para justificar la existencia de este arco se debería presentar la explicación dialéctica correspondiente a esta conclusión, y el grafo de argumentos con los argumentos que intervienen en los árboles de dicha explicación dialéctica.

Suponga que luego de cargar los datos del problema de la familia Jota en la aplicación, tal y como se muestra en la Figura 5.2, se hace clic sobre el botón **Run**. Además de mostrar

los resultados, la aplicación mostrará un grafo como el que se muestra en la Figura 5.6. Observe que este grafo es similar al mostrado en la Figura 4.9 (solo que los nodos están distribuidos de manera diferente), del cual se puede inferir la misma relación de preferencia que se presenta en el Ejemplo 4.1.2.

Al hacer clic derecho en alguno de los arcos de un grafo explicativo, se despliega un menú en el que se puede seleccionar la acción “**Explain**” (ver Figura 5.6). Al hacer esto, se mostrará la explicación dialéctica y el grafo de argumentos asociados a ese arco, los cuales explican y justifican su existencia. Por ejemplo, para entender por qué la alternativa *lago* es mejor que *crcro*, uno debe hacer clic derecho en el arco (*lago*, *crcro*), y luego clic en **Explain** (ver Figura 5.6). Esto producirá que se muestren la explicación dialéctica para la conclusión $pref(lago, crcro)$, y el grafo de argumentos en el que aparecen sólo los argumentos que apoyan o están en contra de la conclusión $pref(lago, crcro)$.

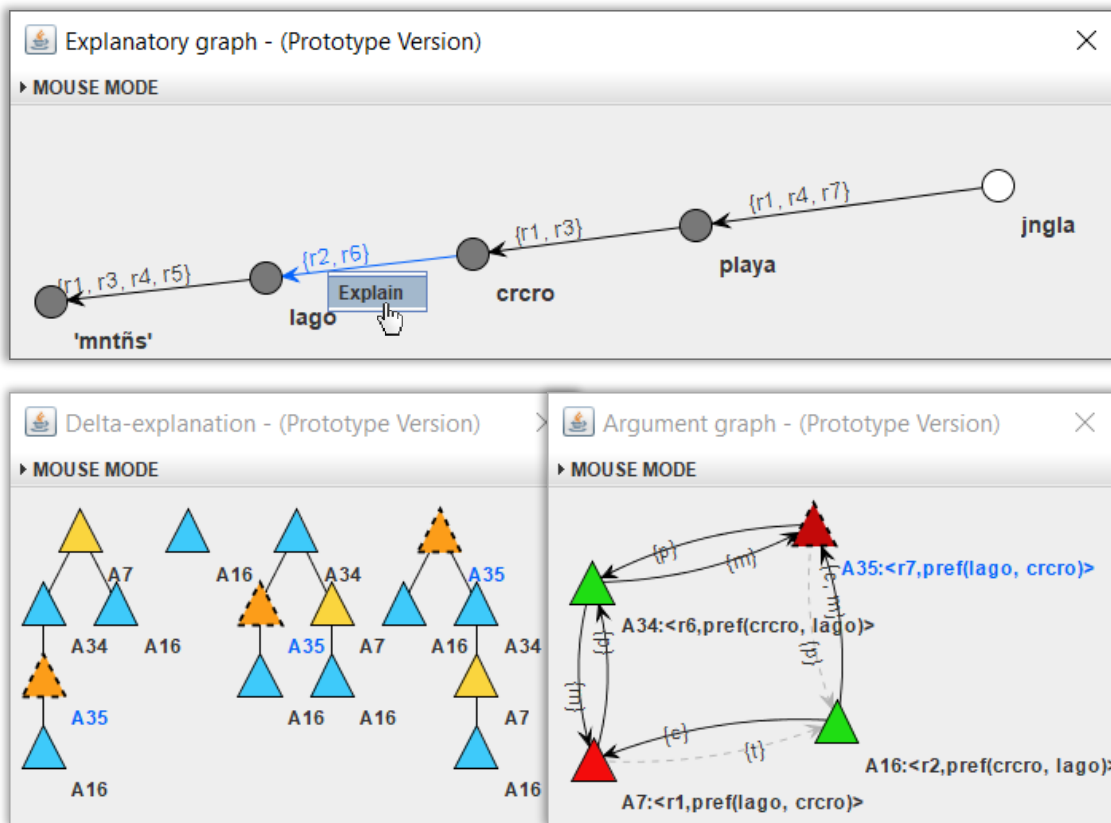


Figura 5.6: En la parte superior se muestra una captura de pantalla en el cual se puede ver el menú que aparece al hacer clic derecho en uno de los arcos de un grafo explicativo. En la parte inferior, se muestra (a la izquierda) la explicación dialéctica y (a la derecha) el grafo de argumentos que explican y justifican la existencia del arco (*crcro*, *lago*) (alternativas correspondientes al problema de la Familia Jota, ver Ejemplo 2.1.1). Note que el argumento A35 del grafo de argumentos, así como también todos los nodos de los árboles de dialéctica etiquetados con dicho argumento, están remarcados con un color más oscuro y un borde más ancho de línea punteada. Esto es producto de que el usuario haya seleccionado ya sea el argumento en sí, o alguno de los nodos etiquetados con este argumento.

La Figura 5.6, también nos muestra cómo la aplicación remarca la relación que existe entre los argumentos del grafo de argumentos y los nodos de los árboles de dialéctica de la explicación dialéctica. Cuando un usuario hace clic en uno de los argumentos, se remarcen todos los nodos de todos los árboles de dialéctica que estén marcados con este argumento. Análogamente, si un usuario hace clic en un nodo, no solo se remarcen todos los demás

nodos que estén etiquetados con el mismo argumento, sino que además, se remarca dicho argumento en el grafo de argumentos. Este detalle visual permite identificar con mayor facilidad la influencia que cada argumento tiene durante el proceso de evaluación de los demás argumentos con los que interactúa.

Una de las metas en el diseño de las interfaces de la aplicación fue la de no sobrecargar de información al usuario innecesariamente. Es por tal motivo, que no se muestran todos los arcos de un grafo explicativo desde un primer momento, sino que se muestran solo aquellos que son necesarios para dilucidar el orden, teniendo en cuenta que la relación de preferencia entre las alternativas es transitiva. Sin embargo, en caso de que un usuario quisiera obtener explicaciones acerca de un arco no mostrado, puede hacerlo solicitando que se muestre dicho arco, y luego pedir las explicaciones como se mostró antes. Para mostrar un nuevo arco, el usuario debe seleccionar los dos nodos que conectaría ese arco (manteniendo apretada la tecla SHIFT mientras selecciona los nodos), hacer clic derecho en uno de ellos, y luego seleccionar la opción “Add extra-edge”.

En la ventana de arriba de la Figura 5.7, se puede ver como, luego de haber seleccionado los nodos *jngla* y *crcro* y haber hecho clic derecho en uno de ellos, aparece la opción de Add extra-edge. Luego, de seleccionar dicha acción, aparece el arco (*jngla*, *crcro*), tal y como se puede apreciar en la ventana de abajo de la Figura 5.7.

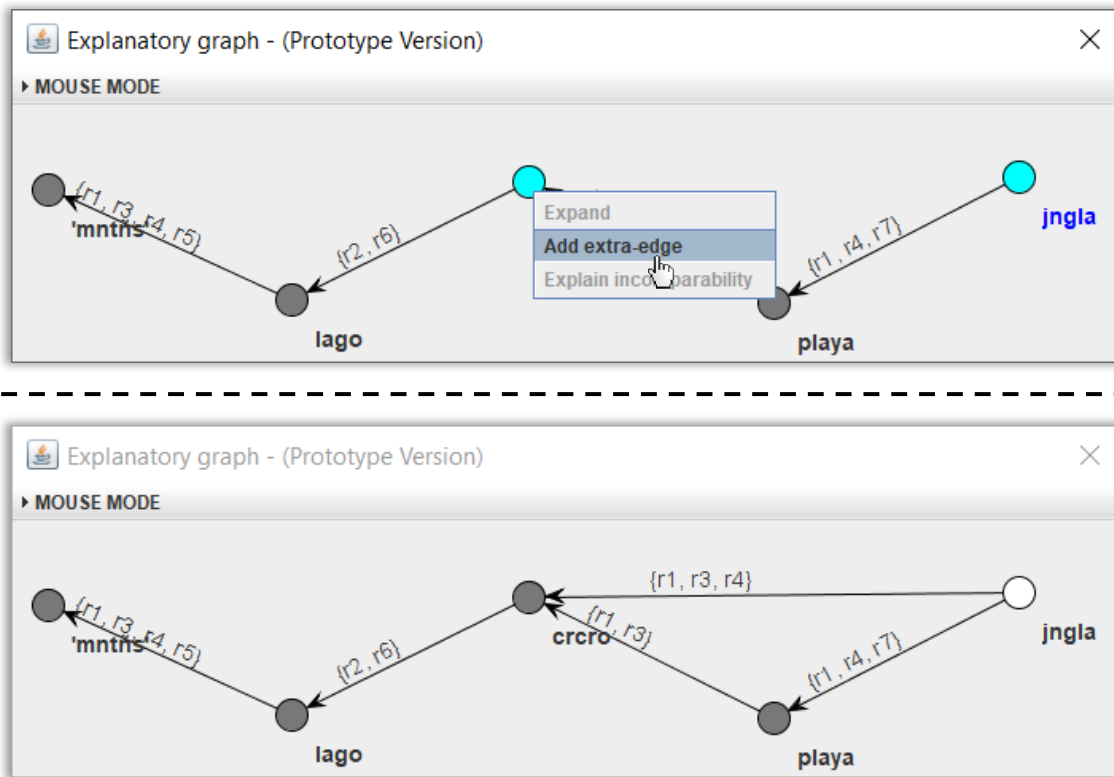


Figura 5.7: Arriba: Captura de pantalla mostrando el menú que aparece al hacer clic derecho en uno de los nodos seleccionados. Abajo: Captura de pantalla mostrando el nuevo arco que aparece entre los nodos *jngla* y *crcro*, luego de solicitar que dicho arco se muestre.

Otra decisión que se tomó en pos de mostrar la menor cantidad de arcos posibles, es la de compactar en un meta-nodo aquellas alternativas que participan en un ciclo, y que por lo tanto son equivalentes (ver Proposición 4.2.1). Luego, de manera similar a lo que ocurre con los arcos ocultos, los usuarios pueden decidir *expandir* los meta-nodos para poder explorar y entender por qué las alternativas dentro del mismo son consideradas equivalentes. La Figura 5.8 muestra el proceso que debe llevarse a cabo para expandir el meta-nodo $[t_1, t_2, t_4]$ que se muestra inicialmente en el grafo explicativo asociado al problema del robot anti-incendios del Ejemplo 4.1.3. Recuerde que para dicho problema, se obtiene como resultado que las alternativas t_1 , t_2 y t_4 son equivalentes. Cabe mencionar, que los arcos que entran y salen de los meta-nodos también están etiquetados con un conjunto de *DePref*-rules. Este conjunto, es igual a la unión de los conjuntos de *DePref*-rules que etiquetan a cada uno de los arcos que entran y salen (respectivamente)

de los nodos internos del meta-nodo, sin contar los arcos entre los nodos internos. Por ejemplo, en la Figura 5.8.1, se puede ver que el conjunto de reglas que etiqueta al arco que sale del meta-nodo es $\{r_9, r_{10}, r_{12}, r_{13}, r_{14}\}$, que es igual a la unión de los conjuntos $\{r_9, r_{10}, r_{12}\}$, $\{r_{10}, r_{12}\}$ y $\{r_9, r_{10}, r_{13}, r_{14}\}$ los cuales etiquetan los arcos que van de t_1 , t_2 y t_4 (respectivamente) a t_3 (ver Figura 5.8.3)

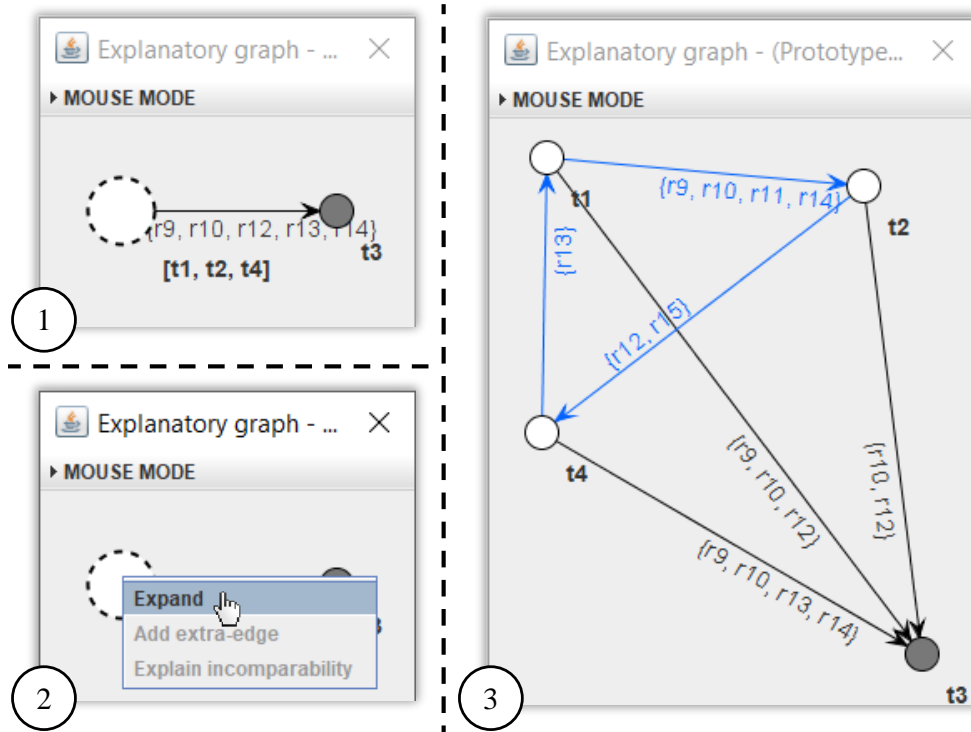


Figura 5.8: 1) Captura de pantalla mostrando un meta-nodo. Los meta-nodos están etiquetados con los nombres de todas las alternativas dentro del mismo, en este caso, serían t_1 , t_2 y t_4 (las alternativas seleccionadas del Ejemplo 4.1.3). 2) Captura de pantalla mostrando el menú que aparece al hacer clic derecho en un meta-nodo. 3) Captura de pantalla mostrando el grafo explicativo, luego de expandir el meta-nodo $[t_1, t_2, t_4]$. Se puede ver como todas las alternativas dentro del meta-nodo ahora son nodos individuales. Los arcos entre estas alternativas están seleccionados para remarcar la existencia del ciclo.

En este punto, uno podría creer que, con las interacciones vistas hasta el momento, los usuarios puedan obtener cualquier explicación disponible acerca de los resultados. Sin embargo, existe una situación particular en la que estas interacciones no bastan. De acuerdo con la Definición 4.3.2, un arco desde una alternativa d_x a otra alternativa d_y

existirá si y sólo si existe una conclusión garantizada que sustente la preferencia entre esas alternativas, es decir, si $pref(d_x, d_y) \in W$. Esto significa, que dadas dos alternativas incomparables (ver Definición 4.1.1), se podrá asegurar que dentro de un grafo explicativo no existirán arcos entre ellas. La incomparabilidad puede surgir por dos motivos: (a) porque no existe una *DePref*-rules en \mathcal{R} que pueda ser satisfecha por la evidencia correspondiente a estas alternativas, o (b) existen *DePref*-rules que sean satisfechas, pero $pref(d_x, d_y)$ y $pref(d_y, d_x)$ no están garantizadas. Para el caso (a), no hay nada que explicar ya que de acuerdo con la evidencia, no hay razones para creer que una de las alternativas debería ser preferida a la otra. No obstante, para el caso (b) las razones existen ya que existen *DePref*-rules que pueden satisfacerse. Esto implica que habrá argumentos que sustentan tanto a $pref(d_x, d_y)$ como a su complemento $pref(d_y, d_x)$, pero ninguno de estos argumentos será aceptado. Por lo tanto, en estos casos, sí existe una justificación que podría ser requerida por lo usuarios para entender por qué dos alternativas son incomparables. A continuación, se presenta un ejemplo que demuestra la existencia de este tipo de situaciones.

Ejemplo 5.2.1. *Una empresa se encuentra en proceso de contratación. Necesita un nuevo empleado para un puesto que quedó vacante. La decisión acerca de quién debe ser contratado deberá ser tomada por un comité formado por otros 3 empleados de la empresa: ag1, ag2 y ag3. Como ag1 y ag2 tienen un puesto de mayor jerarquía que ag3, su opinión es más importante en la decisión. Luego de las entrevistas, los agentes acordaron evaluar a los candidatos de acuerdo a: la remuneración pretendida, su formación, experiencia y qué tan bien estuvo su entrevista. Finalmente, para poder decidir a cuál candidato contratar, los agentes establecen las *DePref*-rules que creen más apropiadas para comparar la aptitud de los candidatos, junto con la importancia que cada uno le asigna a estas *DePref*-rules. Todos los elementos del escenario de elección, las *DePref*-rules y los órdenes de importancia y prioridad que modelan el problema anterior, están detallados en las ventanas de la Figura 5.9.*

The screenshot shows three windows from a software application:

- Agents priority-order**: Contains the text "ag1 > ag3; ag2 > ag3".
- Importance orders**: A table with columns "Agent" and "Importance Order".

Agent	Importance Order
ag1	r1 > r2, r1 > r3, r1 > r4, r9 > r3, r9 > r4
ag2	r4 > r1, r6 > r1, r7 > r1, r8 > r2
ag3	r4 > r1, r5 > r1, r7 > r1, r8 > r9
- Criteria**: A table with columns "Criterion" and "Domain (Values)".

Criterion	Domain (Values)
remuneracion	between(300000,120000)
formacion	[terciario,universitario,especializado]
experiencia	[junior,semi-senior,senior]
entrevista	[pesima,mala,regular,buena,excelente]
- Evidence**: A table with columns "alternative", "remuneracion", "formacion", "experiencia", and "entrevista".

alternative	remuneracion	formacion	experiencia	entrevista
d1	200000	universitario	semi-senior	buena
d2	null	universitario	semi-senior	excelente
d3	230000	especializado	junior	buena
d4	180000	terciario	junior	regular
d5	200000	terciario	senior	buena
d6	240000	especializado	semi-senior	mala
- Preference rules**: A list of rules (r1-r9) defining preference relations between alternatives based on criteria.

Figura 5.9: Captura de pantalla de la aplicación mostrando las ventanas en las que aparecen cargados los datos que modelan la situación planteada en el Ejemplo 5.2.1.

Luego de ejecutar la aplicación, se obtiene que el candidato que debería ser elegido es *d2*, quien no informó remuneración pretendida, posee una formación universitaria, es semi-senior y tuvo una excelente entrevista. La Figura 5.10 (1) muestra el grafo explicativo que permite dilucidar la relación de preferencia entre los candidatos. Note que, en este dominio en particular, contar con un orden de preferencia como parte de la respuesta, y no sólo con el resultado final, les permite a los usuarios tener otras opciones en caso de que el mejor candidato termine rechazando la propuesta.

Como se puede apreciar en la Figura 5.10, los candidatos *d1* y *d5* son incomparables. La Figura 5.10 (2) muestra cómo, a través de la aplicación, los usuarios pueden consultar por qué estas alternativas son incomparables al seleccionar la acción “*Explain incomparability*”, habiendo seleccionado previamente las alternativas involucradas, y haciendo clic derecho en una de ellas para desplegar el menú de interacciones. La Figura 5.10 (3) muestra el grafo de argumento y explicación dialéctica que justifican y explican la incomparabilidad entre las alternativas *d1* y *d5*.

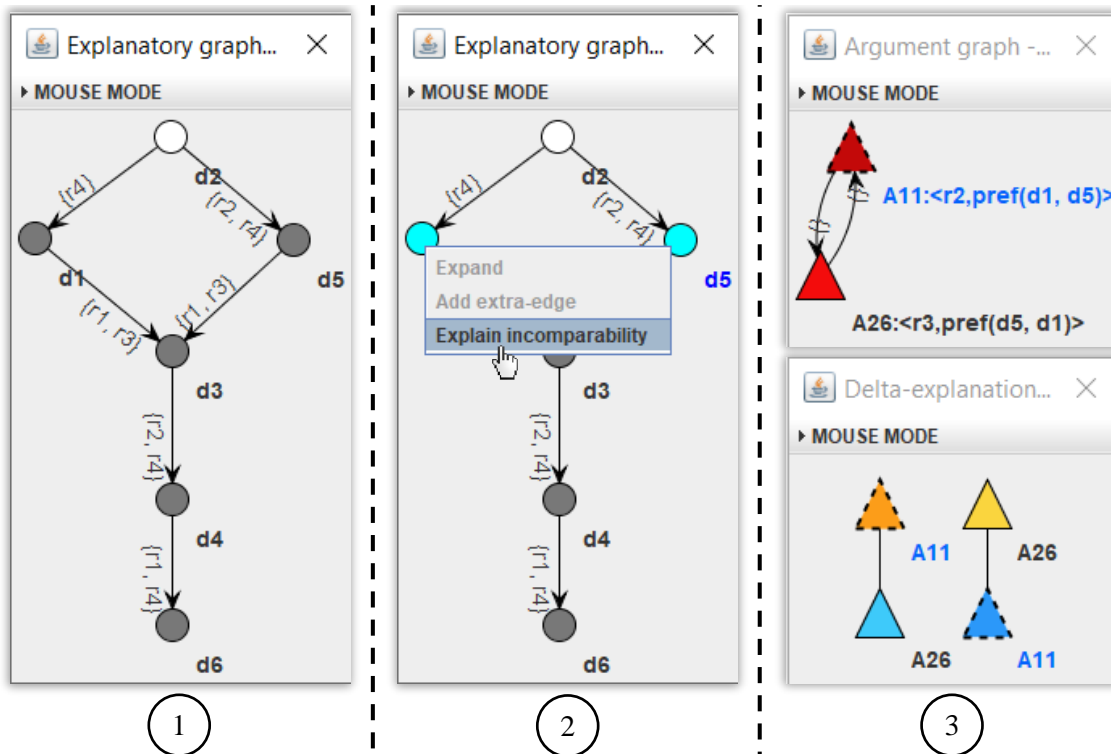


Figura 5.10: 1) Captura de pantalla de la aplicación mostrando el grafo explicativo que justifica los resultados del motor de decisión al ejecutarlo con los datos especificados en la Figura 5.9. 2) Captura de pantalla de la aplicación mostrando el menú que aparece al hacer clic derecho en una de las alternativas, luego de seleccionar dos alternativas incomparables. 3) Grafo de argumentos y explicación dialéctica que justifican y explican la incomparabilidad entre las alternativas seleccionadas.

Observe que los grafos explicativos por sí solos no son suficiente para poder brindarle a los usuarios todas las explicaciones acerca de por qué una alternativa es, o no, preferida a otra. La aplicación desarrollada, soluciona esta deficiencia haciendo uso de una estructura especial que identifica las alternativas incomparables, y de esta forma puede brindar acceso a las explicaciones acerca de la incomparabilidad de dos alternativas, tal y como fue mostrado en el ejemplo anterior.

5.3. Resultados experimentales

En esta sección se presentarán los resultados de un análisis empírico que se llevó a cabo para medir el desempeño de la aplicación antes presentada. Todas las pruebas fueron llevadas a cabo con un procesador AMD Ryzen 5600H @3.3Ghz, con 8GB de RAM DDR4 @3200Mhz. Los factores de desempeño que se tuvieron en cuenta para este análisis fueron el tiempo de ejecución, y la cantidad de argumentos generados, como un estimador del uso de memoria. Todos los datos fueron obtenidos de la pantalla de datos estadísticos de la aplicación (parte derecha de la pantalla principal, ver Figura 5.1). La Figura 5.11 muestra un recorte de esta parte de la pantalla. El tiempo de ejecución fue calculado como la suma del tiempo de razonamiento, o “Reasoning Time”, y el tiempo de selección, o “Selection Time”.

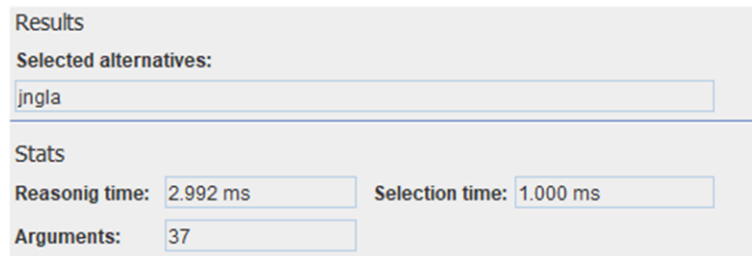


Figura 5.11: Recorte de la pantalla principal de la aplicación mostrando la sección de datos estadísticos. “Reasoning Time” hace referencia al tiempo que la aplicación tarda desde comienza a evaluar las *DePref*-rules para generar los argumentos hasta que todos estos son evaluados para determinar el conjunto de conclusiones garantizadas (W). “Selection Time” hace referencia al tiempo que la aplicación tarda en calcular la clausura transitiva sobre W , y luego obtener el conjunto de alternativas seleccionadas (S). Por último, “Arguments” hace referencia a la cantidad de argumentos que fueron generados durante el proceso.

La Figura 5.12 muestra los valores de rendimiento obtenidos para los problemas de decisión planteados en los ejemplos presentados anteriormente:

- (a) El problema de la familia Jota presentado en el Ejemplo 4.1.2, que cuenta con cinco (5) alternativas y ocho (8) *DePref*-rules.
- (b) El problema del robot anti-incendios presentado en el Ejemplo 4.1.3, que cuenta con cuatro (4) alternativas y ocho (8) *DePref*-rules.

- (c) El problema de contratación de empleados presentado en el Ejemplo 5.2.1, que cuenta con seis (6) alternativas y nueve (9) *DePref*-rules.

El gráfico de la izquierda muestra los valores correspondientes al tiempo de ejecución (en mili-segundos) que consumió la aplicación para resolver cada problema, diferenciando entre el tiempo de razonamiento y el tiempo de selección, mientras que el gráfico de la derecha muestra la cantidad de argumento generados en cada problema.

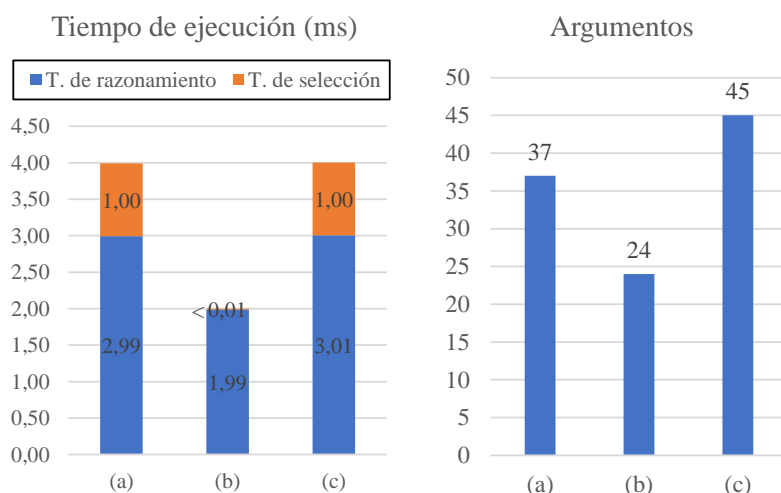


Figura 5.12: Valores obtenidos para el tiempo de ejecución (izquierda) y la cantidad de argumentos (derecha) tras resolver tres de los problemas de decisión presentados en esta tesis: (a) el problema de la Familia Jota (ver Ejemplo 4.1.2), (b) el problema del robot anti-incendios (ver Ejemplo 4.1.3), y (c) el problema de contratación de empleados (ver Ejemplo 5.2.1).

Como la cantidad de alternativas y *DePref*-rules es similar en todos los problemas antes mencionados, el desempeño de la aplicación a la hora de resolver estos problemas, también es similar. Sin embargo, parece que tanto el tiempo de ejecución, como la cantidad de argumentos generados tiende a aumentar cuando los conjuntos de alternativas o de *DePref*-rules crecen.

Para analizar la variación de los factores de rendimiento con respecto a la cantidad de alternativas, se generaron nuevos problemas de decisión considerando los mismos criterios, rango de valores, *DePref*-rules, agentes, y los ordenes de importancia y prioridad

que los utilizados en el problema (c). Esto fue así, ya que este problema considera 4 criterios con distinta cantidad de valores para sus rangos, y es el problema con la mayor cantidad de *DePref*-rules. Luego, a partir de estos datos se generaron nuevos conjuntos de evidencia de manera aleatoria, aumentando paulatinamente la cantidad de alternativas consideradas como parte del escenario de elección. Los conjuntos de evidencia aleatorios fueron generados automáticamente por la aplicación por medio de la opción “Random” (ver Figura 5.13). Los valores asignados a cada alternativa son seleccionados siguiendo una distribución uniforme entre todos los valores dentro del rango de valores legales de cada criterio, con un 5% de probabilidad de asignar un valor nulo. Al seleccionar esta opción, la aplicación solicitará que el usuario ingrese la cantidad de alternativas que deberá tener el escenario de elección, y luego generará valores aleatorios para cada una de las alternativas con respecto a cada uno de los criterios de preferencia (previamente definidos en el paso 1).

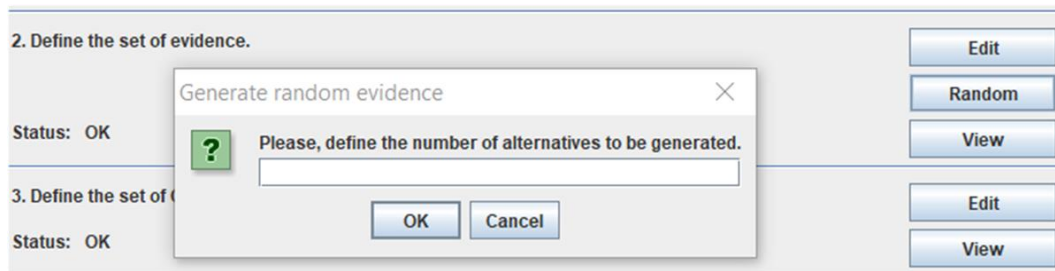


Figura 5.13: Captura de pantalla mostrando el cuadro de diálogo que aparece al hacer clic en la opción “Random” del paso 2.

La Figura 5.14 muestra los valores de los factores de desempeño que se obtuvieron al ir aumentando el número de alternativas que son consideradas como parte del escenario de elección. El eje horizontal muestra la cantidad de alternativas consideradas, el eje vertical izquierdo muestra el tiempo de ejecución, y el eje vertical derecho muestra la cantidad de argumentos. Debajo del gráfico aparecen detallados los valores correspondientes a cada una de las pruebas, por ejemplo, al considerar 75 alternativas, el tiempo de selección fue de 2,6 segundos, el tiempo de razonamiento fue de 21,15 segundos, y se generaron 8.833 argumentos. Por cada configuración se realizaron cinco pruebas, es decir, cinco pruebas considerando 10 alternativas, cinco pruebas considerando 25, cinco pruebas considerando 50, y así siguiendo hasta llegar a las 175 alternativas. Los valores de rendimiento que

se muestran, son el promedio de los valores obtenidos en cada una de estas pruebas. Se puede notar un crecimiento del tiempo de ejecución exponencial, a medida que aumenta la cantidad de alternativas. También se nota un aumento de la cantidad de argumentos, pero este crecimiento es mucho menor en comparación con el del tiempo de ejecución. En el caso extremo, con 175 alternativas, se tiene que el tiempo de ejecución (promedio) fue de 947,83 segundos (109,34 seg + 838,49 seg), aproximadamente 16 minutos. En este caso, la cantidad (promedio) de argumentos generados fue de 48.517, y suponiendo (generosamente) unos 256 bytes por argumento, se tendría un consumo de casi 12 megabytes de memoria. Está claro que la limitación con respecto a la cantidad de alternativas que pueden ser analizadas a la vez viene dada por el consumo de tiempo, y no por un posible exceso del uso de memoria. Dependiendo del contexto en el que se use la aplicación, la velocidad de respuesta puede ser un factor crítico, por lo que debería restringirse el número de alternativas. Por ejemplo, en un problema como el del robot anti-incendios (ver Ejemplo 4.1.3), que requiere acción inmediata, probablemente sea buena idea restringir el número de alternativas a 50, para tener un tiempo de respuesta aproximado de 4 segundos. Por otro lado, en problemas como el de decidir a dónde vacacionar (ver Ejemplo 2.1.1) o a quién contratar (ver Ejemplo 5.2.1), los usuarios podrían tolerar una cierta espera para obtener una solución, lo que los habilitaría a poder comparar un mayor número de alternativas.

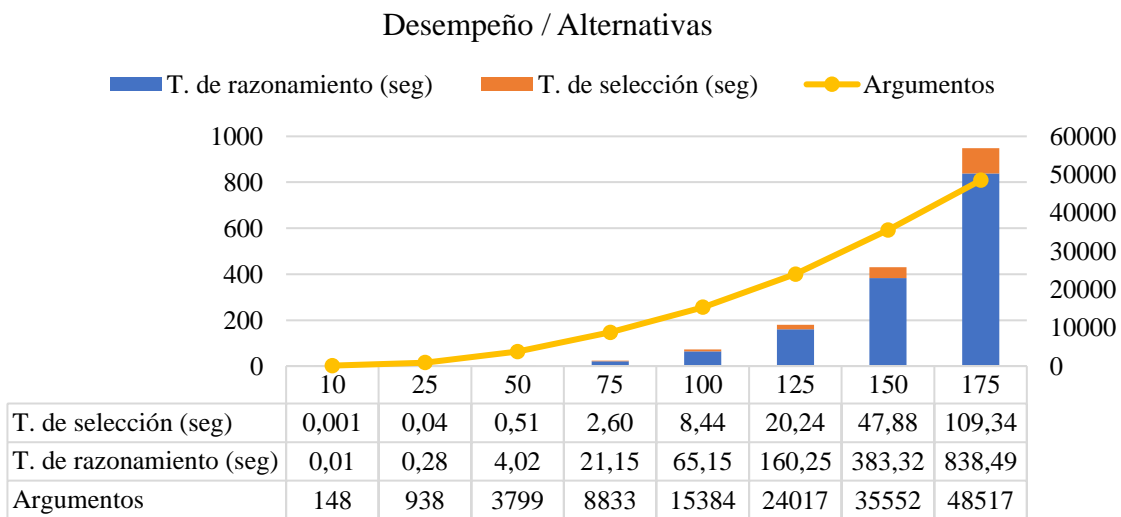


Figura 5.14: Valores de los factores de desempeño obtenidos al aumentar el número de alternativas tomando como base la configuración del problema (c).

Por último, para analizar la variación de los factores de rendimiento con respecto a la cantidad de *DePref*-rules, se tomó nuevamente como base la configuración del problema (c), pero en este caso sólo preservando la configuración con respecto a los criterios y a los agentes. En cada prueba se utilizaron conjuntos de *DePref*-rules de distintos tamaños, y por cada uno de estos conjuntos se realizaron cinco pruebas. El primer conjunto fue construido sólo con las primeras cuatro reglas del Ejemplo 5.2.1 (*r1*, *r2*, *r3* y *r4*), y los demás conjuntos se construyeron agregando de a una regla más por cada uno, hasta llegar a tomar todas las *DePref*-rules de dicho ejemplo. Fueron en total seis conjuntos, uno con cuatro *DePref*-rules, el siguiente con cinco, el siguiente con seis, y así hasta llegar al conjunto con todas las *DePref*-rules del Ejemplo 5.2.1. Todas las pruebas fueron realizadas considerando un escenario de elección con 50 alternativas, y para cada una de ellas se generó un nuevo conjunto de evidencia aleatorio, de la misma forma en la que fueron generados los conjuntos de evidencia para el experimento anterior.

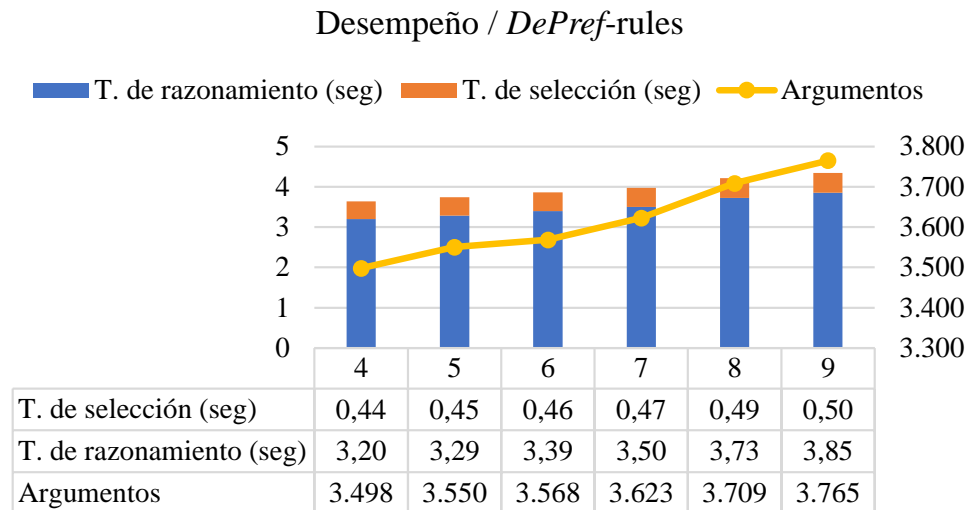


Figura 5.15: Valores de los factores de desempeño obtenidos al aumentar el número de *DePref*-rules tomando como base la configuración del problema (c), y fijando la cantidad de alternativas en 50.

La Figura 5.15 muestra los valores (promedio) de los factores de desempeño obtenidos en cada una de las pruebas antes mencionadas. Se puede observar que el tiempo de ejecución y la cantidad de argumentos crecen a medida que aumenta la cantidad de *DePref*-rules. Sin embargo, a diferencia de lo que sucede con respecto a la cantidad de

alternativas, el incremento para ambos factores de desempeño parecería seguir una relación lineal con respecto a la cantidad *DePref*-rules. Esto indica que el tamaño del conjunto de *DePref*-rules no afecta significativamente el desempeño de la aplicación, en comparación a como lo hace el tamaño del conjunto de alternativas.

5.4. Resumen

En este capítulo se ha descrito el funcionamiento de la aplicación que se desarrolló con el propósito de implementar el formalismo de toma de decisiones grupales que se introdujo en los capítulos anteriores. Se hizo hincapié en describir la interfaz de aplicación, tanto para la etapa de modelado del problema, como en lo que respecta a las interacciones que deben realizar los usuarios para obtener las explicaciones acerca del funcionamiento del sistema.

Esta aplicación no sólo permite el uso del formalismo, ya sea para propósitos personales o de investigación, sino que potencia y extiende los conceptos formales de las explicaciones, como lo son las explicaciones dialécticas (ver Definición 4.3.1) y los grafos explicativos (ver Definición 4.3.2), para que los usuarios puedan sacarle el máximo provecho. Esto, en parte, se debe a que para la implementación de estos conceptos se tuvo como objetivo tratar de reducir la cantidad de información que fuese presentada al usuario para no agobiarlo, sin privarlo de poder consultar por más información. En línea con esto, se tomó la decisión de que, en principio, la aplicación sólo muestre una parte del grafo explicativo total, ocultando arcos que no sean necesarios para poder dilucidar (por transitividad) el orden entre las alternativas. Como fue explicado, los usuarios pueden hacer que estos arcos sean mostrados, en caso de ser necesario (ver Figura 5.7). Otra decisión que se tomó siguiendo el mismo objetivo, fue que las alternativas que resulten ser equivalentes sean agrupadas para reducir aún más la complejidad del orden mostrado, dándole la posibilidad a los usuarios de desagrupar estas alternativas, en caso de querer visualizar las relaciones entre ellas y/o entre las demás alternativas fuera del grupo (ver Figura 5.8).

La aplicación, además, complementa las explicaciones con ayudas visuales que permiten entender mejor las relaciones entre los árboles de dialéctica de las explicaciones dialécticas y los argumentos de los grafos de argumentos, como se muestra en la Figura 5.6). Esta aplicación también subsana una falencia de los grafos explicativos, permitiendo a los usuarios entender por qué dos alternativas son incomparables, o mejor dicho,

por qué no hay una conclusión garantizada que establezca una preferencia entre ellas (ver Figura 5.10).

Además, se mostraron algunos resultados experimentales que permiten delimitar el alcance práctico que tiene esta aplicación. Mediante los experimentos realizados, se pudo determinar (de forma empírica) la influencia del tamaño del conjunto de alternativas y del conjunto de *DePref*-rules, en lo que respecta al tiempo de respuesta y al uso de memoria de la aplicación.

Capítulo 6

Comparación con otros formalismos

En esta sección serán descritos distintos trabajos en los que se plantea utilizar algún tipo de modelo argumentativo para el desarrollo de sistemas de toma de decisiones. No obstante, antes de continuar con esto, se presentarán algunos conceptos generales sobre argumentación abstracta que serán necesarios para comprender los trabajos en cuestión.

6.1. Argumentación abstracta

En esta sección, se presentarán las nociones básicas correspondientes a los marcos argumentativos abstractos introducidos por Dung en [Dun95]. La propuesta de Dung comienza con la definición de los marcos de argumentación abstracta (o AAF, por sus siglas en inglés). Concretamente, un AAF es una tupla $\langle Args, Atts \rangle$ donde el primer elemento es un conjunto de argumentos, y el segundo elemento es una relación binaria $Atts \subseteq Args \times Args$ que representa el ataque (o derrota) entre los argumentos. Dados dos argumentos $\mathcal{A}, \mathcal{B} \in Args$, si $(\mathcal{A}, \mathcal{B}) \in Atts$ se dice que \mathcal{A} ataca (o derrota) a \mathcal{B} . La relación de ataque denota el hecho de que estos argumentos no pueden ser aceptados simultáneamente por hallarse en conflicto.

La Figura 6.1 muestra un AAF representado en forma de grafo, donde cada nodo (triángulo) representa un argumento, y una flecha de un argumento a otro indica que el argumento que se encuentra en el origen de la flecha ataca al argumento que se encuentra apuntado por la misma. Esta representación es comúnmente utilizada para representar los argumentos y ataques de un AAF. El AAF de la

figura es $\mathbf{F}_D = \langle \text{Args}_D, \text{Atts}_D \rangle$, donde $\text{Args}_D = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_4, \mathcal{A}_5\}$ y $\text{Atts}_D = \{(\mathcal{A}_1, \mathcal{A}_2), (\mathcal{A}_2, \mathcal{A}_3), (\mathcal{A}_3, \mathcal{A}_4), (\mathcal{A}_4, \mathcal{A}_5), (\mathcal{A}_5, \mathcal{A}_3), (\mathcal{A}_5, \mathcal{A}_4)\}$.

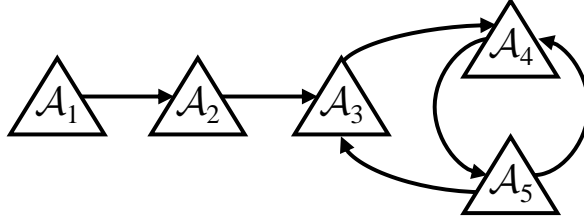


Figura 6.1: Representación gráfica del AAF \mathbf{F}_D .

En [Dun95], se considera que los argumentos son entidades abstractas indivisibles, cuyo rol está determinado por su relación con los otros argumentos. Este formalismo se abstrae del lenguaje de representación de conocimiento, del proceso de construcción de argumentos y de cómo se identifican los conflictos entre ellos, con el propósito de poder concentrarse únicamente en determinar cuáles deberían ser los argumentos aceptados. Para esto, el formalismo de Dung adopta una aproximación declarativa que permite definir una serie de semánticas de aceptabilidad para los argumentos de un AAF. Cada semántica determina los distintos conjuntos de argumentos que podrían llegar a ser considerados como aceptados, a los que se denomina extensiones (diríjase a [BCG11] para una excelente introducción a este tema). En [Dun95], se consideran cuatro semánticas: *completa*, *preferida*, *estable* y *grounded*.

Previo a la definición de las semánticas de aceptabilidad, es necesario definir algunas nociones auxiliares. Sea $\langle \text{Args}, \text{Atts} \rangle$ un AAF:

- Un conjunto $S \subseteq \text{Args}$ es *libre de conflicto* si y sólo si no existen $\mathcal{A}, \mathcal{B} \in S$ tales que $(\mathcal{A}, \mathcal{B}) \in \text{Atts}$.
- Un conjunto $S \subseteq \text{Args}$ *defiende* a un argumento $\mathcal{A} \in \text{Args}$ si y sólo si $\forall \mathcal{B} \in \text{Args}$ tal que $(\mathcal{B}, \mathcal{A}) \in \text{Atts}$, existe $\mathcal{A}' \in S$ tal que $(\mathcal{A}', \mathcal{B}) \in \text{Atts}$.
- Un conjunto $S \subseteq \text{Args}$ es *admisibile* si y sólo si S es libre de conflictos y $\forall \mathcal{A} \in S$, S defiende a \mathcal{A} .

Considerando el AAF de la Figura 6.1, se tiene (por ejemplo) que \emptyset , $\{\mathcal{A}_1\}$, $\{\mathcal{A}_2\}$, $\{\mathcal{A}_1, \mathcal{A}_3\}$, $\{\mathcal{A}_1, \mathcal{A}_4\}$ y $\{\mathcal{A}_1, \mathcal{A}_5\}$ son conjuntos libres de conflicto, mientras que (por ejem-

plo) $\{\mathcal{A}_1, \mathcal{A}_2\}$ o $\{\mathcal{A}_1, \mathcal{A}_3, \mathcal{A}_5\}$ no lo son. Observe que de todos los conjuntos libres de conflictos antes presentados, los únicos admisibles son \emptyset , $\{\mathcal{A}_1\}$ y $\{\mathcal{A}_1, \mathcal{A}_5\}$ (\mathcal{A}_1 no recibe ataques, por lo que cualquier conjunto lo defiende; y \mathcal{A}_5 ataca a todos sus atacantes, por lo que cualquier conjunto que lo contenga lo defiende).

A partir de los conceptos anteriores, Dung define la semántica completa que, partiendo de la noción de admisibilidad, caracteriza a las extensiones completas de un AF. De manera similar, luego se definen las semánticas preferida, estable y grounded, las cuales imponen restricciones adicionales sobre las extensiones completas. Sea $\langle Args, Atts \rangle$ un AAF:

- Un conjunto $S \subseteq Args$ es una extensión completa si y sólo si S es admisible y $\forall \mathcal{A} \in Args$ tal que S defiende a \mathcal{A} , $\mathcal{A} \in S$.
- Un conjunto $S \subseteq Args$ es una extensión preferida si y sólo si S es una extensión completa maximal (con respecto a \subseteq).
- Un conjunto $S \subseteq Args$ es una extensión estable si y sólo si S es una extensión preferida y $\forall \mathcal{B} \in Args \setminus S$, existe $\mathcal{A} \in S$ tal que $(\mathcal{A}, \mathcal{B}) \in Atts$.
- Un conjunto $S \subseteq Args$ es una extensión grounded si y sólo si S es una extensión completa minimal (con respecto a \subseteq).

De esta manera, la semántica que se elija determina el conjunto de argumentos que puede ser aceptado. Siguiendo con el ejemplo antes presentado, se tiene que tanto $\{\mathcal{A}_1\}$ como $\{\mathcal{A}_1, \mathcal{A}_5\}$ son extensiones completas. Observe que \emptyset no es una extensión completa ya que \mathcal{A}_1 , al no poseer atacantes, es defendido por este conjunto, pero no está incluido en el mismo. A su vez, $\{\mathcal{A}_1\}$ es una extensión grounded, mientras que $\{\mathcal{A}_1, \mathcal{A}_5\}$ es, al mismo tiempo, una extensión preferida, y una extensión estable. Así, si se tomara en cuenta la semántica grounded el único argumento que sería aceptado es \mathcal{A}_1 , mientras que si se tomara en cuenta cualquiera de las otras semánticas, $\{\mathcal{A}_1, \mathcal{A}_5\}$ debería ser el conjunto de argumentos aceptados.

Ahora considere un nuevo AAF $\mathbf{F}'_D = \langle Args_D, Atts'_D \rangle$, donde $Atts'_D = (Atts_D \setminus \{(\mathcal{A}_3, \mathcal{A}_4), (\mathcal{A}_5, \mathcal{A}_3)\})$, es decir, \mathbf{F}'_D es el AAF que se obtiene a partir de \mathbf{F}_D al eliminar el ataque de \mathcal{A}_3 a \mathcal{A}_4 y el de \mathcal{A}_5 a \mathcal{A}_3 (ver Figura 6.2).

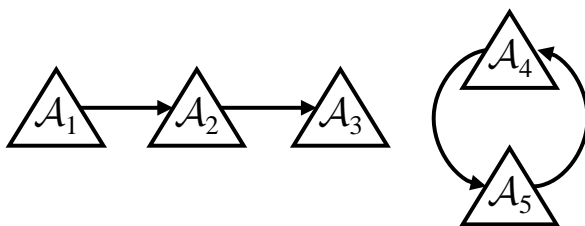


Figura 6.2: Representación gráfica del AAF \mathbf{F}'_D .

En este nuevo AAF existen dos extensiones preferidas (que también, en este caso coinciden, con las estables): $\{\mathcal{A}_1, \mathcal{A}_3, \mathcal{A}_4\}$ y $\{\mathcal{A}_1, \mathcal{A}_3, \mathcal{A}_5\}$. Para determinar, en casos como este, el conjunto de argumentos aceptados, Dung clasifica a los argumentos en tres conjuntos. Sea $\langle Args, Atts \rangle$ un AAF y $x \in \{completa, preferida, estable, grounded\}$ una semántica:

- Un argumento $\mathcal{A} \in Args$ es escépticamente aceptado con respecto a x si y sólo si \mathcal{A} pertenece a todas las extensiones correspondientes a la semántica x .
- Un argumento $\mathcal{A} \in Args$ es crédulamente aceptado con respecto a x si y sólo si \mathcal{A} pertenece a al menos una de las extensiones correspondientes a la semántica x .
- Un argumento $\mathcal{A} \in Args$ es rechazado con respecto a x si y sólo si \mathcal{A} no pertenece a ninguna de las extensiones correspondientes a la semántica x .

Así, considerando \mathbf{F}'_D , se tendría \mathcal{A}_1 y \mathcal{A}_3 son escépticamente aceptados con respecto a las semánticas preferida y estable, mientras que \mathcal{A}_4 y \mathcal{A}_5 son crédulamente aceptados con respecto a estas semánticas. Por otro lado, \mathcal{A}_2 es rechazado bajo cualquier semántica.

6.2. Sistemas argumentativos para toma decisiones individuales

En esta sección se presentarán tres propuestas que definen un sistema de toma de decisiones haciendo uso de un modelo argumentativo. Estas propuestas fueron elegidas entre varios trabajos, por ser lo suficientemente representativas del conjunto total ya que todas utilizan enfoques diferentes. Entre otras cosas, una de las propuestas utiliza marcos de argumentación abstracta, otra utiliza marcos de argumentación abstracta dinámicos

[RMGS10] (también conocidos como DAF, por sus siglas en inglés), y la tercera un marco de argumentación estructurada llamado argumentación basada en suposiciones [DKT09] (también conocido como ABA, por sus siglas en inglés). Por otro lado, dos de ellas calculan la aceptabilidad de los argumentos utilizando las semánticas de Dung, mientras que una de ella lo hace siguiendo un enfoque similar al de DeLP [GS04], y al presentado en esta tesis, mediante el uso líneas de argumentación y árboles de dialéctica (conceptos previamente presentados en el Capítulo 3). Estas y otras diferencias, irán siendo descritas a lo largo de esta sección.

6.2.1. Marco de trabajo para la agregación de evidencia sobre tratamientos médicos [HW12]

Hunter y Williams, en [HW12], proponen un marco de trabajo para la agregación de evidencia sobre tratamientos clínicos, con el objetivo de producir una recomendación acerca de qué tratamiento debería ser aplicado.

Este marco de trabajo recibe como entrada una tabla de evidencia en la que se comparan pares de tratamientos. Cada fila de esta tabla (o pieza de evidencia), contrasta un par de tratamientos en cuanto a algún aspecto (o indicador) específico, como podría ser el promedio de supervivencia de los pacientes o la probabilidad de generar algún efecto colateral. La Tabla 6.1 muestra la evidencia que aparece en [HW12] con respecto a tomar (CP) o no tomar (NP) píldoras anticonceptivas para evitar el embarazo.

	Izquierda	Neto	Derecha	Indicador	Valor	Tipo de estudio	Significativo?
e_{01}	CP	<	NC	Cáncer de mama	1.04	ECA	Sí
e_{02}	CP	>	NC	Cáncer de ovarios	0.99	MA	Sí
e_{03}	CP	>	NC	Embarazo	0.05	ECA	Sí
e_{04}	CP	<	NC	Trombosis	1.02	MA	Sí

Tabla 6.1: Evidencia que aparece en [HW12] con respecto a tomar (CP) o no tomar (NP) píldoras anticonceptiva para evitar el embarazo.

Observe que cada pieza de evidencia indica los dos tratamientos que están siendo comparados (el del lado izquierdo y el del lado derecho), el resultado neto de la comparación, es decir, a cuál de los dos tratamientos favorecen los resultados obtenidos, el indicador por el que están siendo comparados, y un valor que determina que tan bueno o malo es

un tratamiento con respecto al otro. Por ejemplo, para el caso de la pieza de evidencia e_{01} se puede observar que CP es inferior a NC con respecto a la probabilidad de producir cáncer de mama ya que por cada 104 mujeres que contrajeron cáncer de mama y tomaban la píldora anticonceptiva, hay 100 mujeres que contrajeron cáncer de mama y no tomaban la píldora anticonceptiva. Además, cada entrada de la tabla indica el tipo de estudio mediante el cual se obtuvieron resultados (ensayo clínico aleatorio (ECA), meta-análisis (MA), análisis de red (AR), o estudio de cohorte (EC)), y si los resultados son (o no) estadísticamente significativos, datos que les permitirán a los médicos evaluar la robustez de la comparación.

El objetivo de esta propuesta es determinar la superioridad de los tratamientos, mediante un proceso argumentativo que sea capaz de contrastar las conclusiones que se derivan a partir de la evidencia. La salida del sistema será lo que los autores definen como grafo de superioridad. Esto es, un grafo dirigido en el cual sus nodos representan a los tratamientos, y sus arcos denotan la superioridad entre los mismos, siendo que los arcos unidireccionales representan la superioridad de un tratamiento hacia otro, y los arcos bidireccionales representan la equivalencia entre tratamientos. La Figura 6.3 muestra cómo se vería un grafo de superioridad en el que se contrastan los tratamientos CP y NP.

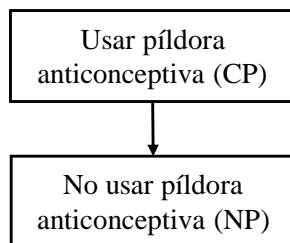


Figura 6.3: Grafo de superioridad para los tratamientos CP y NP.

El proceso argumentativo que toma una tabla de evidencia como entrada y produce un grafo de superioridad como salida consta de cinco pasos:

1. Generación de argumentos inductivos. Para cada par de tratamientos τ_1 y τ_2 se particiona la evidencia correspondiente a estos dos tratamientos en tres conjuntos:

SUP es el subconjunto de evidencia que indica que τ_1 es superior a τ_2 ;

EQU es el subconjunto de evidencia que indica que τ_1 y τ_2 son indiferentes; e

INF es el subconjunto de evidencia que indica que τ_1 es inferior a τ_2 .

Luego, a partir de la evidencia y un conjunto de reglas de inferencia se generan todos los argumentos inductivos. Un argumento inductivo es una tupla $\langle X, \alpha \rangle$, donde X es un subconjunto de evidencia correspondiente a dos tratamientos τ_1 y τ_2 , y α es la conclusión que se deduce a partir de X y una de las siguientes reglas de inferencia:

- a. Si $X \subseteq SUP$, entonces τ_1 es superior a τ_2 , denotado $\tau_1 > \tau_2$.
- b. Si $X \subseteq EQU$, entonces τ_1 y τ_2 son indiferentes, denotado $\tau_1 \sim \tau_2$.
- c. Si $X \subseteq INF$, entonces τ_1 es inferior a τ_2 , denotado $\tau_1 < \tau_2$.

Por ejemplo, de acuerdo con la evidencia de la Tabla 6.1, los argumentos que se podrían generar con respecto a CP y NP son:

$$\begin{aligned} \mathcal{A}_1 &= \langle \{e_{01}\}, CP < NP \rangle & \mathcal{A}_4 &= \langle \{e_{02}\}, CP > NP \rangle \\ \mathcal{A}_2 &= \langle \{e_{04}\}, CP < NP \rangle & \mathcal{A}_5 &= \langle \{e_{03}\}, CP > NP \rangle \\ \mathcal{A}_3 &= \langle \{e_{01}, e_{04}\}, CP < NP \rangle & \mathcal{A}_6 &= \langle \{e_{02}, e_{03}\}, CP > NP \rangle \end{aligned}$$

Vale la pena notar, que sea n la cantidad de piezas de evidencia que compare a dos tratamientos τ_1 y τ_2 dentro de uno de los conjuntos $SUP/EQU/INF$ van a existir $2^n - 1$ argumentos para concluir que τ_1 es superior/indiferente/inferior a τ_2 . Es decir, un argumento por cada forma posible de agrupar las distintas piezas de evidencia que lleven a la misma conclusión. Esto debe ser tenido en cuenta a la hora de implementar el sistema ya que el número de argumentos crecería exponencialmente en relación a la cantidad de piezas de evidencia.

2. Identificar las preferencias entre los argumentos inductivos. En línea con [AC02], para poder determinar cuando un argumento es más fuerte (o preferido) que otro, y así, poder resolver los conflictos entre los mismos, es necesario establecer la forma de compararlos. Para esto, los usuarios del sistema deben definir un orden de preferencia entre los beneficios que ofrecen los tratamientos, para que entonces, los argumentos puedan ser comparados en base a los beneficios en los que se sustenta su conclusión.

Los beneficios vinculados a cada argumento son representados por un conjunto de pares $(Indicador, Valor)$ asociados a los indicadores y valores normalizados correspondientes a las piezas de evidencia en las que se sustenta el argumento. Siguiendo con el ejemplo anterior, se tendría que:

$$Beneficios(\mathcal{A}_1) = \{(c\u00e1ncer_de_mama, 0.96)\}$$

$$\text{Beneficios}(\mathcal{A}_2) = \{(trombosis, 0.98)\}$$

$$\text{Beneficios}(\mathcal{A}_3) = \{(\text{cáncer_de_mama}, 0.96), (trombosis, 0.98)\}$$

$$\text{Beneficios}(\mathcal{A}_4) = \{(\text{cáncer_de_ovarios}, 0.99)\}$$

$$\text{Beneficios}(\mathcal{A}_5) = \{(\text{embarazo}, 0.05)\}$$

$$\text{Beneficios}(\mathcal{A}_6) = \{(\text{cáncer_de_ovarios}, 0.99), (\text{embarazo}, 0.05)\}$$

Note que el valor asociado a la probabilidad de contraer cáncer de mama para el argumento \mathcal{A}_1 es 0,96 y no 1,04 como se mostraba en la evidencia de la Tabla 6.1. Esto es debido a que, en este caso, la evidencia está destacando la inferioridad de CP con respecto a NP, pero al verlo en términos de beneficios, lo que en realidad hay que destacar es la superioridad de NP con respecto a CP, por lo que lo más adecuado es normalizar el valor, calculando la inversa del valor que aparece en la evidencia. Lo mismo ocurre con los beneficios de \mathcal{A}_2 y el valor asociado a la probabilidad de contraer trombosis.

A continuación, la Figura 6.4 muestra mediante un grafo dirigido, las preferencias que un usuario del sistema podría tener con respecto a los beneficios antes mencionados. En este grafo, cada nodo representa un conjunto de beneficios, y un arco de un nodo a otro indica que los beneficios del primero son al menos tan preferidos como los del segundo.

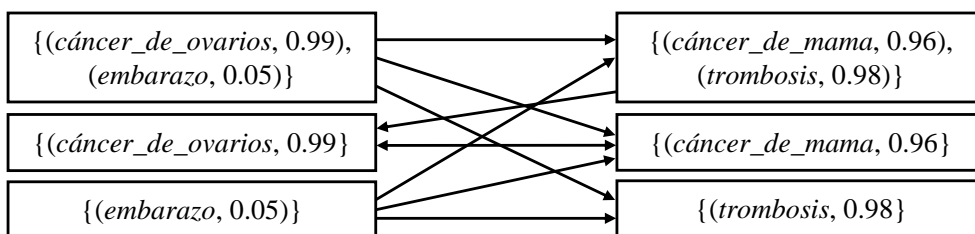


Figura 6.4: Preferencias de un posible usuario sobre los beneficios de los argumentos generados a partir de la evidencia acerca de los tratamientos CP y NP.

Cabe mencionar, que los autores advierten la dificultad que tendría establecer las preferencias sobre conjuntos de beneficios en ocasiones donde exista una gran cantidad de los mismos. Por tal motivo, los autores definen ciertas propiedades generales que debería cumplir la relación de preferencia entre beneficios (como por ejemplo, que sea reflexiva y transitiva), y proponen una relación de preferencia general que determina la fuerza de un argumento basada en la cantidad de beneficios en los que se sustenta. Además,

ellos mismos remarcan la necesidad de definir reglas que permitan generar la relación de preferencia de manera automática. Sin embargo, sólo muestran un ejemplo de cómo podrían ser estas reglas para evaluar los tratamiento contra el glaucoma, pero no definen un mecanismo o lenguaje formal que permita hacerlo de manera general.

3. Generar meta-argumentos. En este punto los usuarios deben realizar un análisis de la evidencia con el objetivo de identificar puntos débiles sobre las distintas piezas de evidencia, como por ejemplo, que los resultados no sean estadísticamente significativos, o que la evaluación clínica no haya sido llevada a cabo sobre sujetos con características aleatorias, y por ende, exista un sesgo en los resultados. Si este es el caso, los usuarios podrán emitir su juicio generando un meta-argumento que desestime la validez de la evidencia con la que hayan sido generados ciertos argumentos inductivos. Los meta-argumentos no poseen una estructura como es el caso de los argumentos inductivos, sino que solo representan la incorformidad del usuario con respecto a la calidad de la evidencia.

Si bien Hunter y Williams definen ciertos patrones para generar meta-argumentos, a diferencia de los argumentos inductivos, los meta-argumentos deberán ser (en su mayoría) generados a mano ya que en la mayoría de los casos requieren que los usuarios evalúen por sí mismos la calidad de la evidencia.

4. Generar el grafo de argumentos. Se disponen los argumentos inductivos y los meta-argumentos en un mismo grafo de argumentos. En este grafo, los ataques entre argumentos inductivos se establecen a partir de la relación de conflicto entre ellos y la relación de preferencia entre sus beneficios. Sean τ_1 y τ_2 un par de tratamientos, y sean \mathcal{A} y \mathcal{A}' dos argumentos inductivos cuyas conclusiones son α y α' , respectivamente. Se dice que \mathcal{A} y \mathcal{A}' están en conflicto si:

$$\alpha = \tau_1 > \tau_2, \text{ y } \alpha' = \tau_1 < \tau_2 \text{ o } \alpha' = \tau_1 \sim \tau_2, \text{ o}$$

$$\alpha = \tau_1 \sim \tau_2, \text{ y } \alpha' = \tau_1 > \tau_2 \text{ o } \alpha' = \tau_1 < \tau_2, \text{ o}$$

$$\alpha = \tau_1 < \tau_2, \text{ y } \alpha' = \tau_1 > \tau_2 \text{ o } \alpha' = \tau_1 \sim \tau_2.$$

Luego, se tiene que un argumento inductivo \mathcal{A} atacará a otro argumento inductivo \mathcal{A}' si y sólo si están en conflicto y no es el caso que los beneficios de \mathcal{A}' sean estrictamente preferidos a los de \mathcal{A} . En cuanto a los meta-argumentos, estos sólo pueden atacar a los argumentos inductivos, y no pueden ser atacados por ningún argumento. Si un meta-argumento es generado para desestimar la validez de una pieza de evidencia e , entonces

este meta-argumento deberá atacar a todo argumento inductivo cuyo conjunto de evidencia contenga a e .

La Figura 6.5 muestra el grafo de argumentos que se obtiene al considerar los argumentos para los tratamientos CP y NP, y las preferencias de la Figura 6.4. Note que en este caso ningún meta-argumento fue generado.

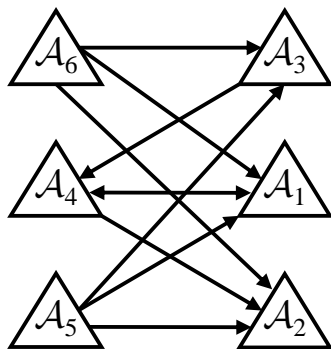


Figura 6.5: Grafo de argumentos que se obtiene al considerar los argumentos para los tratamientos CP y NP, y las preferencias de la Figura 6.4.

5. Generar el grafo de superioridad. Una vez definido el conjunto de argumentos y sus ataques, la aceptabilidad de los argumentos es evaluada utilizando la semántica grounded de Dung, y en caso que la extensión grounded sea vacía, se utiliza la semántica preferida, y se selecciona una de las múltiples extensiones preferidas. Observe que para el caso del grafo de argumentos de la Figura 6.5, los argumentos aceptados son: \mathcal{A}_4 , \mathcal{A}_5 y \mathcal{A}_6 . Luego, las conclusiones de los argumentos aceptados son las que se utilizarán para determinar la superioridad de los tratamientos. Note que para que la extensión grounded sea vacía, por cada par de tratamientos, deberían existir al menos dos argumentos inductivos cuyas conclusiones estén en conflicto, y no debería existir preferencia entre los beneficios de estos tratamientos (asumiendo que no hay argumentos que desestimen la validez de la evidencia). Si esto sucede, todos los tratamientos deberían ser considerados incomparables ya que no se podría determinar una preferencia entre sus beneficios. Sin embargo, al seleccionar alguna de las extensiones preferidas lo que se estaría haciendo es considerar que los beneficios que sustentan los argumentos que pertenecen a la extensión seleccionada, son preferidos a los beneficios de los argumentos que no pertenecen a la extensión. Pudiendo así, opacar la transparencia del funcionamiento del sistema ya que al elegir una extensión,

indirectamente, se establecen preferencias inexistentes, que no fueron definidas de forma explícita por el usuario.

Finalmente, las conclusiones de los argumentos aceptados son utilizadas para construir el grafo de superioridad. Este grafo es lo que se les presenta a los médicos como resultado del proceso de agregación, permitiéndoles entender y visualizar el orden resultante entre los distintos tratamientos. La Figura 6.3 muestra el grafo de superioridad obtenido a partir de las conclusiones de los argumentos aceptados del grafo de argumentos de la Figura 6.5.

Existen varios puntos en los que la propuesta de Hunter y Williams difiere con respecto al trabajo presentado en esta tesis. El primero de ellos es la forma en la que se comparan las alternativas (o tratamientos para el caso de [HW12]). De acuerdo con estos autores, cada pieza de evidencia es utilizada para comparar las alternativas y generar argumentos, incluso cuando la pieza de evidencia en cuestión sugiere que las alternativas son indiferentes con respecto al factor o criterio que está siendo evaluado. En cambio, mediante el uso de *DePref*-rules uno puede especificar distintas sentencias que establecen como uno desea comparar las alternativas en base a varios criterios de una sola vez, teniendo en cuenta sus distintos valores, pudiendo establecer comparaciones complejas en términos de hasta cuánto uno está dispuesto a perder en lo que respecta a uno o varios criterios si es que obtiene algún beneficio con respecto a otros criterios. Por otro lado, en la propuesta de Hunter y Williams, las preferencias se establecen sobre los distintos conjuntos de beneficios asociados a los argumentos, y teniendo en cuenta la gran cantidad de argumentos que podrían llegar a existir, establecer estas preferencias podría ser una tarea ardua. A diferencia de esto, en el trabajo presentado en los capítulos anteriores, si bien existe un trabajo similar a la hora de determinar el orden de importancia entre las *DePref*-rules o de determinar el orden de prioridad entre los agentes, la cantidad de elementos a comparar es mucho menor en relación a la cantidad de argumentos que se generan (ver Sección 5.3). Además, las propuestas también difieren en cuanto al proceso de evaluación de argumentos que se utiliza. Los árboles de dialéctica presentan una ventaja ya que (sin costo adicional) mejoran la transparencia del sistema por ser estructuras que resultan más intuitivas y fáciles de entender que el solo hecho de presentar un grafo de argumentos junto con las extensiones. De hecho, según [NJ17], esto último no debería ser considerado como una explicación ya que sólo puede ser interpretado por alguien que entienda cómo se definen las extensiones.

6.2.2. El uso de argumentación para explicar las mejores decisiones [ZFLT14]

En este trabajo, los autores presentan un modelo argumentativo para resolver problemas de toma de decisiones multi-criterio que tiene la habilidad, además de resolver estos problemas, de brindar explicaciones en lenguaje natural. Más precisamente, Zhong et al. describen cómo generar una instancia concreta de un marco de argumentación basado en suposiciones (o ABA, por assumption-based argumentation [DKT09]), en el cual las conclusiones que puedan derivarse con respecto a si una decisión debe ser considerada como una de las mejores, estén alineadas con las mejores decisiones que se obtendrían al modelar el problema mediante el uso de un marco de decisión de Fan et. al [FT14]. Luego, haciendo uso de los argumentos y sus relaciones, los autores muestran como construir explicaciones argumentativas (en lenguaje natural) acerca de por qué una cierta decisión debe (o no) ser considerada como una de las mejores. De hecho, el procedimiento que estos autores utilizan para construir el marco ABA es muy similar al que se plantea en [FT14], con la diferencia de que las mejores alternativas no solo deben ser dominantes ([MTV09]), sino que además deben poseer la cualidad de mínima desviación. Las siguientes propuestas también siguen una estrategia similar: [FCS⁺13, ZFM⁺18, ZST⁺20].

A continuación, se presenta las nociones de ABA que son necesarias para comprender la propuesta de Zhong.

Argumentación basada en suposiciones [DKT09]

La argumentación basada en suposiciones es uno de los enfoques provenientes del campo de la argumentación estructurada, como también es el caso de DeLP [GS14] o ASPIC+ [MP14]. A diferencia de la argumentación abstracta, los enfoques de argumentación estructurada ofrecen herramientas para representar el conocimiento, especifican cómo se constituyen los argumento (su estructura), y cómo se identifican los conflictos entre los mismos.

En particular, un marco ABA es una tupla $\langle L, R, A, C \rangle$, donde:

- $\langle L, R \rangle$ es un sistema deductivo con lenguaje L y reglas $R \subseteq \{\beta_0 \leftarrow \beta_1, \dots, \beta_m \mid \beta_i \in L, 0 \leq i \leq m\}$. Cabe aclarar que L no suele especificarse ya que puede ser reconstruido a partir de los literales que aparecen en R y en A ;

- $A \subseteq L$ es un conjunto no vacío de literales, conocido como las suposiciones; y
- C es un mapeo total de A a $2^L \setminus \{\emptyset\}$. Cada elemento de $C(\alpha)$ es conocido como uno de los contrarios de α .

Para cada regla $\rho = \beta_0 \leftarrow \beta_1, \dots, \beta_m$, β_0 es la cabeza de ρ (identificada como $Head(\rho) = \beta_0$), y β_1, \dots, β_m constituyen el cuerpo de ρ (identificado como $Body(\rho) = \{\beta_1, \dots, \beta_m\}$). Si $m = 0$, ρ es representada como $\beta_0 \leftarrow$ y $Body(\rho) = \emptyset$. Si en un marco ABA, ninguna suposición aparece como cabeza de una regla, se dice que este marco ABA es plano (en [ZFLT14] todos los marcos ABA que serán presentados serán planos).

En un marco ABA, los argumentos representan la deducción de una conclusión a partir de las reglas y la suposiciones, y los ataques están dirigidos a las suposiciones. Concretamente, dado un marco ABA $\langle L, R, A, C \rangle$,

- Un argumento para la conclusión $\beta \in L$ sustentado por $A \subseteq A$ (denotado $A \vdash \beta$) es un árbol finito cuyos nodos están etiquetados por las premisas en L o por el símbolo τ , tal que la raíz está etiquetada con β , las hojas están etiquetadas con τ o con una suposición de A , y los nodos que no son hojas (incluida la raíz) están etiquetados con $\beta' \in L$ y tienen tantos hijos como elementos en el cuerpo de una regla en R con cabeza β' ; y
- Un argumento $A_1 \vdash \beta_1$ ataca a un argumento $A_2 \vdash \beta_2$ si y sólo si $\beta_1 \in C(\alpha)$ para algún $\alpha \in A_2$.

El siguiente ejemplo servirá para ilustrar los conceptos antes mencionados. Sea $\langle L_X, R_X, A_X, C_X \rangle$ un marco ABA tal que:

$$R_X = \{(p \leftarrow q, a), (q \leftarrow), (r \leftarrow b, c), (t \leftarrow)\}, \quad A_X = \{a, b, c\}, \text{ y}$$

$$C_X(a) = \{r\}; \quad C_X(b) = \{s\}; \quad C_X(c) = \{t\}.$$

A partir de este marco ABA se pueden construir los siguientes argumentos: $\emptyset \vdash q$, $\{a\} \vdash p$, $\{b, c\} \vdash r$ y $\emptyset \vdash t$. Observe que $\{b, c\} \vdash r$ ataca a $\{a\} \vdash p$ ya que $r \in C_X(a)$, y que $\emptyset \vdash t$ ataca a $\{b, c\} \vdash r$ ya que $t \in C_X(c)$.

Una vez contruidos los argumentos, e identificados los ataques (o conflictos) entre los mismos, es posible utilizar las semánticas de Dung para determinar cuáles de estos argumentos deberían aceptarse, y cuáles no. En particular, para [ZFLT14], los autores utilizan la noción de admisibilidad para determinar cuándo un argumento es aceptado. Un argumento \mathcal{A} será aceptado si y sólo si este es admisible, es decir, si existe un conjunto admisible S tal que $\mathcal{A} \in S$.

Teniendo en cuenta el ejemplo anterior, se tiene que $S = \{\emptyset \vdash q, \{a\} \vdash p, \emptyset \vdash t\}$ es un conjunto de argumentos admisible ya que defiende a $\{a\} \vdash p$ del ataque de $\{b, c\} \vdash r$, y ningún otro argumento es atacado en S . Por lo tanto, todos los argumentos en S son admisibles. Por otro lado, como $\{b, c\} \vdash r$ es atacado por $\emptyset \vdash t$ y este argumento no es atacado por ningún otro, $\{b, c\} \vdash r$ no puede ser defendido, y por ende, no es un argumento admisible.

Como se mencionó anteriormente, Zhong et al. describen cómo generar una instancia concreta de un marco ABA en el cual las conclusiones que puedan derivarse con respecto a si una decisión debe ser considerada, estén alineadas con las conclusiones que se obtendrían al modelar el problema mediante el uso de un marco de decisión de Fan et. al [FT14]. A continuación, para comprender la propuesta de Zhong, se explican brevemente los marcos de decisión de Fan.

Marco de decisión de Fan et al. [FT14]

Para Fan et al., un marco de decisión se define a partir de un conjunto de alternativas, un conjunto de atributos mediante los cuales se caracterizan las alternativas, y un conjunto de metas que determinan los atributos que se busca que las alternativas posean. Formalmente, un marco de decisión es una tupla $\langle \mathcal{D}, \mathcal{F}, \mathcal{G}, \mathcal{DF}, \mathcal{GF} \rangle$, donde:

- $\mathcal{D} = \{d_1, \dots, d_n\}$ ($n > 0$), es el conjunto de alternativas,
- $\mathcal{F} = \{f_1, \dots, f_m\}$ ($m > 0$), es el conjunto de atributos,
- $\mathcal{G} = \{g_1, \dots, g_l\}$ ($l > 0$), es el conjunto de metas,
- \mathcal{DF} es una matriz de $(n \times m)$ que representa la relación entre las alternativas y los atributos, tal que $\mathcal{DF}[i, j] = 1$, si d_i posee el atributo f_j , y $\mathcal{DF}[i, j] = 0$ en caso contrario,
- \mathcal{GF} es una matriz de $(l \times m)$ que representa la relación entre las metas y los atributos, tal que $\mathcal{GF}[i, j] = 1$, si g_i es satisfecha por el atributo f_j , y $\mathcal{GF}[i, j] = 0$ en caso contrario.

Se dice que una alternativa $d_i \in \mathcal{D}$ cumple con una meta $g_j \in \mathcal{G}$ si y sólo si existe un atributo $f_k \in \mathcal{F}$ tal que $\mathcal{DF}[i, k] = 1$ y $\mathcal{GF}[j, k] = 1$. La función $\gamma(d)$ retorna todas las metas que son satisfechas por la alternativa d .

Zhong et al. toman como dominio de aplicación para su propuesta el problema de determinar la condena de un hecho delictivo en base a la semejanza con otros delitos y las condenas aplicadas a los mismos. La Tabla 6.2 resume la información que se tiene

acerca de una serie de delitos (x_1, \dots, x_{10}), cada uno con sus respectivos atributos y condenas aplicadas. Los atributos f_1, \dots, f_9 representan “mayor de 18 años”, “entre 16 y 18 años”, “robo”, “reincidencia”, “botín de gran valor”, “botín muy gran valor”, “botín de extremadamente gran valor”, “botín encontrado”, “accesorio”, respectivamente. Se puede ver que en esta tabla, cada delito tiene asociado un valor para cada uno de los atributos, 1 cuando el delito posee el atributo y 0 en caso contrario. Por ejemplo, para el caso x_2 , se tiene que el acusado era mayor de 18 años (f_1), que el valor del botín fue extremadamente grande (f_7), y la condena que se le dio fue de 10 años cárcel, 3 años de privación de la libertad de los derechos políticos y \$10.000 de multa.

Caso	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	Condena
x_1	0	1	0	1	0	1	0	0	0	1año+\$1.000
x_2	1	0	0	0	0	0	1	0	0	10años+3años+\$10.000
x_3	1	0	0	1	1	0	0	1	0	6meses+\$1.000
x_4	1	0	0	0	0	1	0	1	0	3años+\$3.000
x_5	1	0	0	0	1	0	0	1	0	6meses+\$1.000
x_6	1	0	0	0	0	1	0	1	0	5años+\$5.000
x_7	1	0	0	1	0	0	0	0	0	5meses+\$1.000
x_8	1	0	0	0	0	0	1	0	1	7meses+\$7.000
x_9	1	0	0	1	1	0	0	0	0	4meses+\$1.000
x_{10}	1	0	1	0	1	0	0	1	0	6meses+\$1.000

Tabla 6.2: Información que se tiene con respecto a las características y condenas de los 10 casos presentados en [ZFLT14].

A partir de esta información, se podría desear conocer cuál sería la condena más adecuada para un nuevo caso en el que (por ejemplo) el acusado es mayor de 18 años (f_1), el valor del botín es extremadamente grande (f_7), pero este fue hallado (f_8). Para conseguir esto, Zhong et al. plantean modelar el problema mediante un marco de decisión de Fan et al. de la siguiente manera: cada una de las condenas representará una decisión, los atributos de cada delito estarán asociados a su correspondiente condena, y las metas serán los atributos asociados al nuevo caso. La intuición detrás de esta idea es que una dada condena anterior es adecuada para un nuevo delito si se cumple que ambos delitos posean los mismos atributos. Siguiendo con el ejemplo, el marco de decisión instanciado para determinar la condena del nuevo caso se define como $\mathbf{M}_Z = \langle \mathcal{D}_Z, \mathcal{F}_Z, \mathcal{G}_Z, \mathcal{DF}_Z, \mathcal{GF}_Z \rangle$, donde $\mathcal{D}_Z = \{d_1, \dots, d_{10}\}$ (d_i representa a la condena aplicada al caso x_i), $\mathcal{F}_Z = \{f_1, \dots, f_9\}$, $\mathcal{G}_Z = \{g_1, g_2, g_3\}$, \mathcal{DF}_Z contiene la información de la Tabla 6.2 con respecto a las conde-

nas y sus respectivos atributos, y \mathcal{GF}_Z contiene la información mostrada por la Tabla 6.3. Note que existe una meta por cada atributo del nuevo caso, y que cada meta es satisfecha sólo por el atributo al que está asociada.

	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9
g_1	1	0	0	0	0	0	0	0	0
g_2	0	0	0	0	0	0	1	0	0
g_3	0	0	0	0	0	0	0	1	0

Tabla 6.3: Tabla \mathcal{GF}_Z del marco de trabajo instanciado para determinar la condena de un nuevo caso que posee los atributos f_1 , f_7 y f_8 .

Una forma de determinar cuáles son las mejores alternativas dentro de un marco de decisión es a partir de la noción de dominancia. Dado un marco de decisión $\langle \mathcal{D}, \mathcal{F}, \mathcal{G}, \mathcal{DF}, \mathcal{GF} \rangle$, se dice que una alternativa $d \in \mathcal{D}$ es fuertemente dominante si y sólo si $\gamma(d) = \mathcal{G}$, y se dice que d es débilmente dominante si y sólo si no existe otra alternativa $d' \in \mathcal{D}$ tal que $\gamma(d) \subset \gamma(d')$. En lo que respecta a \mathbf{M}_Z , se tiene que ninguna de las alternativas son fuertemente dominantes, aunque d_2 y d_8 son débilmente dominantes porque cumplen con las metas g_1 y g_2 , al igual que d_3 , d_4 , d_5 , d_6 y d_{10} que cumplen con las metas g_1 y g_3 .

Existen ocasiones, como sucede en este dominio de aplicación, en las que la noción de dominancia no es lo suficientemente específica ya que, por ejemplo, ignora otros factores de potencial interés como la presencia de atributos “redundantes”, es decir, atributos que no son requeridos para cumplir ninguna meta. De aquí surge el concepto de desviación mínima. Una alternativa con desviación mínima es aquella que su conjunto de atributos redundantes es minimal (con respecto a la inclusión de conjuntos) si se compara con los conjuntos de atributos redundantes de todas las demás alternativas. Formalmente, sea λ la función que retorna el conjunto de atributos redundantes de una alternativa, se dice que una alternativa $d_x \in \mathcal{D}$ se desvía mínimamente si y sólo si no existe $d' \in \mathcal{D}$ tal que $\lambda(d_x) \subset \lambda(d')$. Zhong et al., proponen combinar el concepto de dominancia con el concepto de desviación mínima, haciendo que las mejores alternativas sean aquellas que, además de ser dominantes, posean la menor cantidad de atributos redundantes (se desvíen mínimamente). Entre todas estas alternativas débilmente dominantes, la única que posee la cualidad de desviación mínima es d_2 ya que $\lambda(d_2) = \{\}$, y todas las demás poseen al menos un atributo redundante ($\lambda(d_3) = \{f_4, f_5\}$, $\lambda(d_4) = \{f_6\}$, $\lambda(d_5) = \{f_5\}$,

$\lambda(d_6) = \{f_6\}$, $\lambda(d_8) = \{f_9\}$, $\lambda(d_{10}) = \{f_5\}$). Esto quiere decir que d_2 sería la sentencia más adecuada ya que x_2 es el caso que más se parece al nuevo, y por lo tanto (en base a los precedentes), la condena de este nuevo caso debería ser similar a la de x_2 .

Cabe destacar que la utilidad del concepto de desviación mínima para identificar las mejores alternativas está muy ligada al dominio de aplicación presentado en [ZFLT14]. En otros dominios, poseer atributos redundantes podría no ser un problema. Más aún, pareciera que los autores sólo se enfocaron en problemas en los que cada meta puede ser cumplida por un sólo atributo, y no por varios, ya que en dicho caso podría existir otra forma de redundancia entre los atributos. Suponga que existe una meta que es cumplida por una alternativa si esta posee cualquiera de dos atributos, si existiera una alternativa que posea los dos atributos a la vez, entonces existiría una especie de redundancia ya que en realidad con poseer uno sólo de ellos sería suficiente para cumplir la meta.

Contra-parte argumentativa del marco decisión y generación de explicaciones

La principal contribución de esta propuesta es la definición constructiva de dos marcos ABA que permiten identificar, mediante un proceso argumentativo, las alternativas fuertemente (o débilmente) dominantes con desviación mínima correspondientes a un marco de decisión. La idea es traducir el problema de encontrar las mejores alternativas en un marco de decisión, al problema de determinar la admisibilidad de un conjunto de argumentos, ya que a partir de las conclusiones de estos argumentos será posible determinar cuáles son las mejores alternativas.

Dado un marco de decisión $\mathbf{M} = \langle \mathcal{D}, \mathcal{F}, \mathcal{G}, \mathcal{DF}, \mathcal{GF} \rangle$, el marco ABA $\langle \mathbf{L}, \mathbf{R}, \mathbf{A}, \mathbf{C} \rangle$ para alternativas fuertemente dominantes con desviación mínima (o MDSD ABA por sus siglas en ingles) correspondiente a \mathbf{M} , se construye de la siguiente manera:

- \mathbf{R} consiste de las siguientes reglas (y ninguna más)¹:

$\forall d, f \in \mathcal{F}$ tal que d posea a f , $hasAttr(d, f) \leftarrow$;

$\forall g \in \mathcal{G}, f \in \mathcal{F}$ tal que g es satisfecha por f , $satBy(g, f) \leftarrow$;

$deviate(D, F) \leftarrow hasAttr(D, F), redundant(F)$;

$notRedundant(F) \leftarrow satBy(G, F)$;

¹Note que para representar de forma compacta todas las reglas, se utilizaron esquemas con variables, que luego deben ser instanciadas de la siguiente manera: D y D' con alternativas, F con atributos y G con metas.

$$notMSD(D) \leftarrow sd(D'), worse(D, D'), nWorse(D', D);$$

$$nSD(D) \leftarrow nMet(D, G);$$

$$met(D, G) \leftarrow hasAttr(D, F), satBy(G, F);$$

$$worse(D, D') \leftarrow deviate(D, F), nAttr(D', F);$$

- $\mathbf{A} = \{ms(d), sd(d), redundant(f), nAttr(d, f) \mid d \in \mathcal{D}, f \in \mathcal{F}\} \cup \{nWorse(d, d') \mid d, d' \in \mathcal{D}\} \cup \{nMet(d, g) \mid d \in \mathcal{D}, g \in \mathcal{G}\}$.

- Para cada $d, d' \in \mathcal{D} (d \neq d')$, $f \in \mathcal{F}$, y $g \in \mathcal{G}$ se tiene que:

$$\mathbf{C}(ms(d)) = \{notMSD(d), nSD(d)\}; \quad \mathbf{C}(sd(d)) = \{nSD(d)\};$$

$$\mathbf{C}(redundant(f)) = \{notRedundant(f)\}; \quad \mathbf{C}(nAttr(d, f)) = \{hasAttr(d, f)\};$$

$$\mathbf{C}(nWorse(d, d')) = \{worse(d, d')\}; \quad \mathbf{C}(nMet(d, g)) = \{met(d, g)\}.$$

La intuición detrás de la definición anterior es la siguiente. Una alternativa d siempre se puede asumir fuertemente dominante con desviación mínima, esto es representado por la suposición $ms(d)$. Esta suposición puede ser refutada si se presentan argumentos para creer que no es fuertemente dominante ($nSD(d)$), o para creer que no posee la cualidad de desviación mínima ($notMSD(d)$). Recuerde que una alternativa d no es fuertemente dominante si no cumple con alguna meta g ($nMet(d, g)$), y no posee la cualidad de desviación mínima si existe otra alternativa d' que sea fuertemente dominante ($sd(d')$) tal que d posea un atributo redundante que d' no ($worse(d, d')$), y que d' no contenga más atributos redundantes que d ($nWorse(d', d)$). Como $sd(d)$ es una suposición, d se asume fuertemente dominante hasta que se demuestre lo contrario ($nSD(d)$). La suposición $nMet(d, g)$ podrá ser refutada si es posible construir un argumento para su contrario ($met(d, g)$), si es que d posee el atributo f ($hasAttr(d, f)$) que es necesario para satisfacer la meta g ($satBy(g, f)$). Las otras reglas, suposiciones y contrarios en el MDSD ABA pueden ser interpretados siguiendo esta misma lógica.

De forma similar, El marco ABA $\langle \mathbf{L}', \mathbf{R}', \mathbf{A}', \mathbf{C}' \rangle$ para alternativas débilmente dominantes con desviación mínima (o MDWD ABA por sus siglas en ingles) correspondiente a \mathbf{M} , se construye de la siguiente manera:

- \mathbf{R}' consiste de las siguientes reglas (y ninguna más):

$$\forall d \in \mathcal{D}, f \in \mathcal{F} \text{ tal que } d \text{ posea a } f, hasAttr(d, f) \leftarrow;$$

$$\forall g \in \mathcal{G}, f \in \mathcal{F} \text{ tal que } g \text{ es satisfecha por } f, satBy(g, f) \leftarrow;$$

$$deviate(D, F) \leftarrow hasAttr(D, F), redundant(F);$$

$$notRedundant(F) \leftarrow satBy(G, F);$$

$$\begin{aligned}
notMWD(D) &\leftarrow wd(D'), worse(D, D'), nWorse(D', D); \\
nWD(D) &\leftarrow met(D', G), nMet(D, G), nMore(D, D'); \\
more(D, D') &\leftarrow met(D, G), nMet(D', G); \\
met(D, G) &\leftarrow hasAttr(D, F), satBy(G, F); \\
worse(D, D') &\leftarrow deviate(D, F), nAttr(D', F);
\end{aligned}$$

$$- A' = \{mw(d), wd(d), redundant(f), nAttr(d, f) \mid d \in \mathcal{D}, f \in \mathcal{F}\} \cup \{nWorse(d, d') \mid d, d' \in \mathcal{D}\} \cup \{nMet(d, g) \mid d \in \mathcal{D}, g \in \mathcal{G}\} \cup \{nMore(d, d') \mid d, d' \in \mathcal{D}\}.$$

- Para cada $d, d' \in \mathcal{D}$ ($d \neq d'$), $f \in \mathcal{F}$, y $g \in \mathcal{G}$ se tiene que:

$$\begin{aligned}
C'(mw(d)) &= \{notMWD(d), nWD(d)\}; & C'(wd(d)) &= \{nWD(d)\}; \\
C'(redundant(f)) &= \{notRedundant(f)\}; & C'(nAttr(d, f)) &= \{hasAttr(d, f)\}; \\
C'(nWorse(d, d')) &= \{worse(d, d')\}; & C'(nMet(d, g)) &= \{met(d, g)\}; \\
C'(nMore(d, d')) &= \{more(d, d')\}.
\end{aligned}$$

Los marcos MDSD ABA y MDWD ABA pueden ser usados para determinar si una alternativa d debe ser seleccionada verificando si existe un argumento admisible cuya conclusión sea $ms(d)$ o $mw(d)$, respectivamente. Una vez instanciado el marco ABA correspondiente, una manera de obtener el conjunto de argumentos admisibles es a partir de la construcción de árboles de disputa. Un árbol de disputa (o dispute tree [DKT06, Ton13]) es un árbol (posiblemente infinito) cuyos nodos están identificados por niveles como nodos proponentes (nodos-P) y nodos oponentes (nodos-O), siendo la raíz un nodo-P. Cada nodo está etiquetado con el argumento que ataca al argumento con el que está etiquetado su padre, excepto por el nodo raíz que está etiquetado con el argumento por el que se inicia la disputa. Todo nodo-P posee un nodo hijo por cada uno de los argumentos que atacan al argumento en este nodo, y todo nodo-O posee exactamente un hijo el cual debe estar etiquetado con uno de los argumentos que atacan al argumento en este nodo. Note que las hojas del árbol deben ser nodos-P ya que todo nodo-O debe tener un hijo. Si se puede construir un árbol de disputa tal que no exista ningún argumento que etiquete tanto a un nodo-P como a un nodo-O, entonces ese árbol es considerado admisible, y el conjunto de todos los argumentos que etiquetan a los nodos-P de dicho árbol es admisible. La Figura 6.6 muestra un fragmento del árbol de disputa que se obtendría al consultar por la admisibilidad de la conclusión $mw(d_2)$ del MSWD ABA correspondiente al marco de decisión \mathbf{M}_Z .

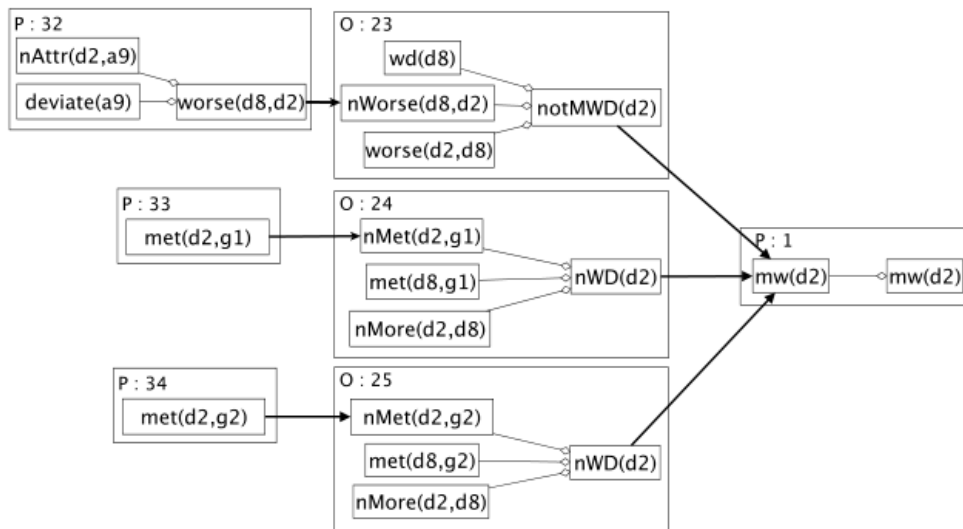


Figura 6.6: Fragmento de un árbol de disputa construido a partir de un argumento para $mw(d_2)$. Cada argumento está delimitado por un rectángulo, las conclusiones de los argumentos aparecen en las cajas internas del lado derecho, y las suposiciones y hechos en las cajas internas del lado izquierdo. Note que cada argumento está identificado como P u O. Si dentro de un argumento sólo existe una única caja interior, entonces este argumento está soportado por un conjunto vacío de suposiciones. Las flechas entre argumentos (rectángulos exteriores) representan el ataque de uno a otro dentro del árbol de disputa.

Al igual que los árboles de dialéctica presentados en el Capítulo 3, los árboles de disputa no solo constituyen un procedimiento de prueba para determinar la admisibilidad de los argumentos, sino que facilitan la comprensión de los resultados, permitiendo visualizar más fácilmente las relaciones de ataque y defensa entre los argumentos. Cabe aclarar que Zhong et al., para sacar aún más provecho de las explicaciones que pueden brindar los árboles de disputa, en [ZFLT14] también definen un algoritmo que toma un árbol de disputa como entrada y produce una explicación en lenguaje natural que se construye a partir de las conclusiones y premisas de los argumentos que etiquetan los nodos del árbol.

Más recientemente, en [ZFLT19], los mismos autores realizan un análisis teórico más profundo de su propuesta anterior, refinan aún más el algoritmo de generación de explicaciones en lenguaje natural, y realizan un estudio empírico para determinar el grado de aceptación de su formalismo al aplicarlo para determinar de forma automática la condena de diversos casos legales.

Una de las diferencias más significativas entre el trabajo de Zhong et al. y la propuesta presentada en esta tesis es que la solución propuesta por dichos autores, si bien se presenta de manera general, en realidad está muy ligada a su aplicación concreta para determinar la condena a los culpables de haber cometido un delito, la cual se base en las condenas aplicadas a casos anteriores. Otra diferencia importante que se puede observar es que la propuesta de Zhong et al. necesita de la existencia de información previa (casos anteriores) para poder funcionar correctamente. Además, para esta propuesta, las preferencias o más bien los criterios por lo que son comparadas las alternativas se encuentran predefinidos ya que lo que posiciona a una alternativa por sobre otra es la similitud entre los atributos del nuevo caso y los casos anteriores, siendo esto codificado en el sistema por medio de las nociones de dominancia y desviación mínima. Como fue explicado en el Capítulo 2 y en Capítulo 3, las *DePref*-rules permiten modelar las preferencias de los agentes de una manera más flexible, permitiendo definir cómo comparar las alternativas, y luego además el MEPM considera los ordenes de importancia entre las reglas de cada agente y la prioridad entre los mismos a la hora de determinar las derrotas entre los argumentos, que a su vez determinan las conclusiones garantizadas que dan lugar al orden entre las alternativas. Y aunque en la propuesta de esta tesis no se provee de explicaciones en lenguaje natural, complementa la falta de ello dando la posibilidad a los usuarios de analizar e interactuar con las explicaciones por medio de herramientas visuales, como se explica en el Capítulo 5.

6.2.3. Marco para toma de decisiones basado en sistemas de argumentación dinámicos [FTG⁺17]

En este trabajo, Ferreti et al., presentan un marco de decisión abstracto basado en reglas de decisión y argumentación dinámica. Este marco de decisión consta de cinco elementos: un conjunto de alternativas disponibles (\mathcal{D}), un conjunto de literales distinguidos con los cuales se identifican los criterios por los que serán comparadas las alternativas (\mathcal{C}), un orden total y estricto sobre los criterios ($>_{\mathcal{C}}$) que representa el orden de preferencia o importancia entre los criterios, una componente epistémica (\mathcal{K}) que será usada para modelar relaciones de preferencia y conflictos entre las alternativas disponibles, y una componente de decisión (Γ) que implementa la política de toma de decisiones del agente basada en un conjunto de reglas de decisión.

A continuación, se presenta el ejemplo de aplicación introducido en [FTG⁺17] para ilustrar los diferentes conceptos de su propuesta. La Figura 6.7 muestra una escena en la que se puede ver un agente robótico cuya tarea es llevar las cajas a la región identificada como “store” (o almacenamiento). La decisión que debe tomar este agente es la de cuál de estas cajas mover primero. Se puede ver que este agente tiene tres alternativas (cajas), las cuáles serán identificadas como $\mathcal{D}_F = \{box_1, box_2, box_3\}$ (box_i representa a la caja i). Estas alternativas serán comparadas según su tamaño, su distancia al robot, y su distancia al almacenamiento, siendo $\mathcal{C}_F = \{smaller, nearer_robot, nearer_store\}$ los literales que identificarán a dichos criterios de comparación, con $nearer_robot >_{\mathcal{C}_F} nearer_store >_{\mathcal{C}_F} smaller$. A partir de estos literales, será posible determinar cuándo una alternativa es más chica que otra (*smaller*), cuándo está más cerca del robot que otra (*nearer_robot*), o cuándo está más cerca del store que otra (*nearer_store*). Esta información es la que (en parte) contiene la componente epistémica \mathcal{K}_F , y la que será conocida como *la evidencia*. Por ejemplo, para el caso del escenario de la Figura 6.7, se tiene que la caja 1 es más chica que la caja 3, lo que será representado como $smaller(box_1, box_3)$; que la caja 1 está mas cerca del robot que la caja 2, lo que será representado como $nearer_robot(box_1, box_2)$; y que la caja 2 está más cerca del store que la caja 3, lo que será representado como $nearer_store(box_2, box_3)$. Como caso especial, si dos alternativas d_x y d_y fuesen iguales con respecto a todos los criterios de comparación, esto se representaría como $same_att(d_x, d_y)$, note que esto no sucede en el ejemplo de la Figura 6.7.

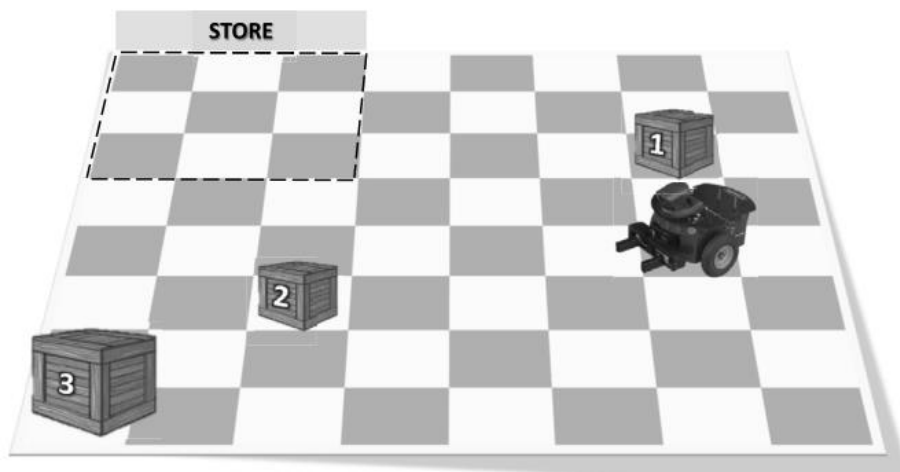


Figura 6.7: Ejemplo de aplicación de [FTG⁺17] - Un agente robótico que debe decidir cuál caja mover primero.

La componente epistémica es, en realidad, un marco de argumentación dinámico (o DAF por su siglas en inglés) instanciado de una forma especial. Los DAF fueron introducidos en [RMGS10], y proveen una extensión a los marcos de argumentación abstracta de Dung que es capaz de operar en ambientes dinámicos. En un DAF existe un conjunto de evidencia el cual determina los argumentos activos dentro de un conjunto de argumentos de trabajo. Al verse modificado el conjunto de evidencia, también se verá modificado el conjunto de argumentos activos, evitando así, la necesidad de instanciar un nuevo marco argumentativo cada vez que la evidencia cambie. Formalmente, un DAF es un tupla $\langle E, Args, \bowtie, \mathbf{pref} \rangle$, donde E es el conjunto de evidencia, conformado por las premisas que representan el estado actual del mundo; $Args$ es el conjunto de argumentos de trabajo, que representa todos los argumentos que podrían llegar a estar activos; \bowtie es una relación binaria (y simétrica) entre los argumentos de $Args$, la cual representa la relación de conflicto entre los argumentos; y \mathbf{pref} es la función de preferencia que determina cuándo un argumento es preferido (o es más fuerte) que otro. En un DAF, un argumento es una tupla $\langle \{\beta_1, \dots, \beta_n\}, \alpha \rangle$ que representa el paso de razonamiento que se necesita para llegar a la conclusión α a partir del conjunto de premisas $\{\beta_1, \dots, \beta_n\}$. Sea $\mathcal{A} \in Args$ un argumento, $cl(\mathcal{A})$ representa la conclusión de \mathcal{A} y $\overline{cl(\mathcal{A})}$ es el complemento de dicha conclusión con respecto a la negación (\neg). Un argumento estará activo cuando todas sus premisas sean satisfechas. Dado un conjunto de evidencia E , una premisa β es satisfecha si $\beta \in E$, o bien, existe un argumento activo cuya conclusión es β . Cuando dos argumentos \mathcal{A} y \mathcal{B} están en conflicto, es decir cuando $(\mathcal{A}, \mathcal{B}) \in \bowtie$, se utiliza la función \mathbf{pref} para determinar cuál derrota² a cuál. Si $\mathbf{pref}(\mathcal{A}, \mathcal{B}) = \mathcal{A}$, entonces \mathcal{A} derrota a \mathcal{B} ; si $\mathbf{pref}(\mathcal{A}, \mathcal{B}) = \mathcal{B}$, entonces \mathcal{B} derrota a \mathcal{A} ; y si $\mathbf{pref}(\mathcal{A}, \mathcal{B}) = \epsilon$, entonces ambos se derrotan mutuamente. Luego, el conjunto de argumentos activos aceptados (o garantizados) de un DAF se obtienen a partir de un proceso de prueba dialéctico similar al que ocurre en DeLP y en el marco propuesto en esta tesis. Es decir, para cada argumento se construye un árbol de dialéctica (en el que intervienen sólo argumentos activos) y si la marca de la raíz es **U**, entonces ese argumento será aceptado y su conclusión estará garantizada; caso contrario (si la marca de la raíz es **D**), el argumento será rechazado. Al igual que en DeLP, para que una conclusión esté garantizada, basta con que exista un sólo argumento aceptado que sustente

²En realidad, el concepto de derrota en los DAF (y todo lo relacionado con el mismo) se define en términos de estructuras argumentativas. Una estructura argumentativa agrupa todos los argumentos que son necesarios para llegar a una conclusión. Como en este caso, todas las estructuras argumentativas estarían compuestas por un sólo argumento, hablar de argumentos o estructuras argumentativas es equivalente.

dicha conclusión. Tenga en cuenta que esta descripción de los DAF no es exhaustiva, por lo que en caso de requerir más información debe dirigirse a [MRF⁺10, RMGS10].

Volviendo a la propuesta de Ferreti et al., la componente epistémica \mathcal{K} de un marco de decisión $\langle \mathcal{D}, \mathcal{C}, >_{\mathcal{C}}, \mathcal{K}, \Gamma \rangle$ es un DAF $\langle E, \text{Args}, \bowtie, \text{pref} \rangle$, donde:

- E es un conjunto consistente de evidencia cuyas premisas son de la forma $c(d_x, d_y)$, o $\text{same_att}(d_x, d_y)$, con $c \in \mathcal{C}$, y $\{d_x, d_y\} \subset \mathcal{D}$. Se dice que E es consistente si y sólo si $\{c(d_x, d_y), c(d_y, d_x)\} \not\subseteq E$, $\{c(d_x, d_y), \text{same_att}(d_x, d_y)\} \not\subseteq E$, y $\{c(d_y, d_x), \text{same_att}(d_x, d_y)\} \not\subseteq E$.

- Args es tal que si $c \in \mathcal{C}$, $\{d_x, d_y\} \subseteq \mathcal{D}$ ($d_x \neq d_y$) y $\text{better} \notin \mathcal{C}$, entonces:

$$\begin{aligned} &\langle \{c(d_x, d_y)\}, \text{better}(d_x, d_y) \rangle \in \text{Args} \\ &\langle \{c(d_x, d_y)\}, \neg \text{better}(d_y, d_x) \rangle \in \text{Args} \\ &\langle \{c(d_y, d_x)\}, \text{better}(d_y, d_x) \rangle \in \text{Args} \\ &\langle \{c(d_y, d_x)\}, \neg \text{better}(d_x, d_y) \rangle \in \text{Args} \\ &\langle \{\text{same_att}(d_x, d_y)\}, \neg \text{better}(d_x, d_y) \rangle \in \text{Args} \\ &\langle \{\text{same_att}(d_x, d_y)\}, \neg \text{better}(d_y, d_x) \rangle \in \text{Args} \end{aligned}$$

- $\bowtie = \{(\mathcal{A}, \mathcal{B}) \mid \{\mathcal{A}, \mathcal{B}\} \subseteq \text{Args}, \text{cl}(\mathcal{A}) = \overline{\text{cl}(\mathcal{B})}\}$.

- Sean $\mathcal{A}, \mathcal{B} \in \text{Args}$ dos argumentos, y $\text{dlit} : \text{Args} \rightarrow \mathcal{C}$ una función que retorna el literal distinguido presente en la premisa de un argumento³:

$$\text{pref}(\mathcal{A}, \mathcal{B}) = \begin{cases} \mathcal{A} & \text{si } \text{dlit}(\mathcal{A}) >_{\mathcal{C}} \text{dlit}(\mathcal{B}), \\ \mathcal{B} & \text{si } \text{dlit}(\mathcal{B}) >_{\mathcal{C}} \text{dlit}(\mathcal{A}), \\ \epsilon & \text{en caso contrario.} \end{cases}$$

De acuerdo con el escenario representado en la Figura 6.7, $\mathcal{K}_F = \langle E_F, \text{Args}_F, \bowtie_F, \text{pref}_F \rangle$ es la componente epistémica del marco de decisión $\langle \mathcal{D}_F, \mathcal{C}_F, >_{\mathcal{C}_F}, \mathcal{K}_F, \Gamma_F \rangle$, donde:

$$E_F = \left\{ \begin{array}{ll} \text{smaller}(\text{box}_1, \text{box}_2), & \text{nearer_store}(\text{box}_2, \text{box}_3), \\ \text{smaller}(\text{box}_2, \text{box}_3), & \text{nearer_robot}(\text{box}_1, \text{box}_2), \\ \text{nearer_store}(\text{box}_1, \text{box}_3), & \text{nearer_robot}(\text{box}_1, \text{box}_3), \\ \text{nearer_store}(\text{box}_2, \text{box}_1), & \text{nearer_robot}(\text{box}_2, \text{box}_3) \end{array} \right\}$$

³La definición de esta función es diferente de la presentada en [FTG⁺17] ya que dicha propuesta fue definida en términos de estructuras argumentativas.

El conjunto de argumentos de trabajo $Args_F$ consta de cuarenta y dos argumentos, los cuales están representados en la Figura 6.8. Los triángulos blancos representan los argumentos activos con respecto a E_F , y los triángulos negros los argumentos inactivos. El texto arriba de los triángulos muestra la conclusión de cada argumento, y el texto debajo la premisa. La etiqueta dentro de cada triángulo blanco será usada para referenciar el argumento que está representando. La derrota entre los argumentos está representada mediante una flecha sólida. El argumento al inicio de la flecha derrota al que se encuentra al final de la misma. Note que en esta figura *same_att*, *smaller*, *nearer_store*, *nearer_robot*, *better*, \neg *better*, *box*₁, *box*₂, y *box*₃ están abreviados como *sa*, *sm*, *ns*, *nr*, *b*, \neg *b*, 1, 2, y 3, respectivamente. Como se puede ver en la Figura 6.8, los argumentos activos de $Args_F$ con respecto a E_F son:

$$\begin{aligned}
\mathcal{A}_1 &= \langle \{smaller(box_1, box_3)\}, better(box_1, box_3) \rangle \\
\mathcal{A}_3 &= \langle \{smaller(box_2, box_3)\}, better(box_2, box_3) \rangle \\
\mathcal{A}_5 &= \langle \{nearer_store(box_2, box_1)\}, better(box_2, box_1) \rangle \\
\mathcal{A}_7 &= \langle \{nearer_store(box_1, box_3)\}, better(box_1, box_3) \rangle \\
\mathcal{A}_9 &= \langle \{nearer_store(box_2, box_3)\}, better(box_2, box_3) \rangle \\
\mathcal{A}_{11} &= \langle \{nearer_robot(box_1, box_2)\}, better(box_1, box_2) \rangle \\
\mathcal{A}_{13} &= \langle \{nearer_robot(box_1, box_3)\}, better(box_1, box_3) \rangle \\
\mathcal{A}_{15} &= \langle \{nearer_robot(box_2, box_3)\}, better(box_2, box_3) \rangle \\
\mathcal{A}_2 &= \langle \{smaller(box_1, box_3)\}, \neg better(box_3, box_1) \rangle \\
\mathcal{A}_4 &= \langle \{smaller(box_2, box_3)\}, \neg better(box_3, box_2) \rangle \\
\mathcal{A}_6 &= \langle \{nearer_store(box_2, box_1)\}, \neg better(box_1, box_2) \rangle \\
\mathcal{A}_8 &= \langle \{nearer_store(box_1, box_3)\}, \neg better(box_3, box_1) \rangle \\
\mathcal{A}_{10} &= \langle \{nearer_store(box_2, box_3)\}, \neg better(box_3, box_2) \rangle \\
\mathcal{A}_{12} &= \langle \{nearer_robot(box_1, box_2)\}, \neg better(box_2, box_1) \rangle \\
\mathcal{A}_{14} &= \langle \{nearer_robot(box_1, box_3)\}, \neg better(box_3, box_1) \rangle \\
\mathcal{A}_{16} &= \langle \{nearer_robot(box_2, box_3)\}, \neg better(box_3, box_2) \rangle
\end{aligned}$$

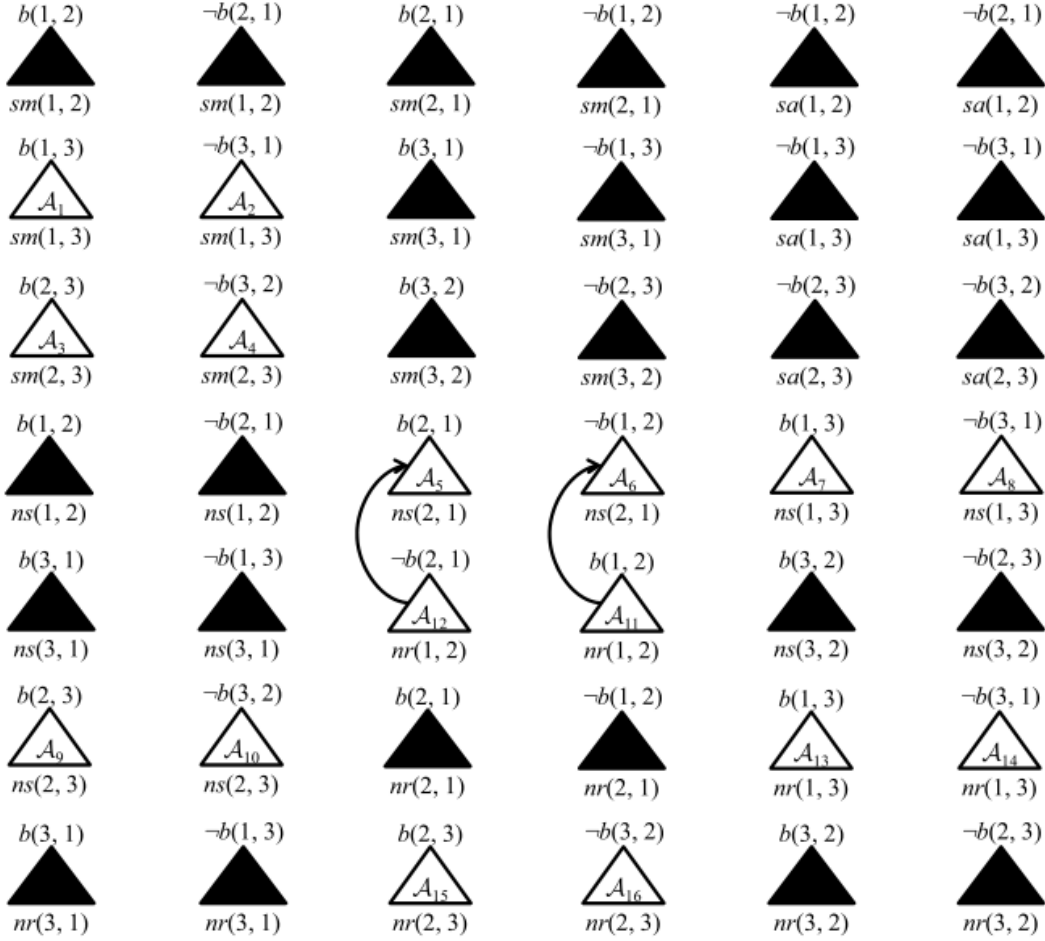


Figura 6.8: Conjunto de argumentos de trabajo $Args_F$ y derrotas correspondientes al ejemplo de aplicación de [FTG⁺17].

En la lista de arriba, aquellos argumentos activos que sustentan que una alternativa es mejor que otra (*better*) aparecen primero, y los que sustentan que una alternativa no es mejor que otra (*-better*) aparecen al final. Note que los conflictos entre los argumentos son $\bowtie_F = \{(\mathcal{A}_6, \mathcal{A}_{11}), (\mathcal{A}_{11}, \mathcal{A}_6), (\mathcal{A}_5, \mathcal{A}_{12}), (\mathcal{A}_{12}, \mathcal{A}_5)\}$. Observe que el argumento \mathcal{A}_{11} sustenta que box_1 es mejor que box_2 , basado en la premisa que box_1 está más cerca del robot que box_2 ; mientras que \mathcal{A}_6 sustenta que box_1 no es mejor que box_2 basado en que box_2 está mas cerca del store que box_1 . Luego, como $nearer_robot >_{c_F} nearer_store$, $\text{pref}(\mathcal{A}_{11}, \mathcal{A}_6) = \mathcal{A}_{11}$, y por lo tanto, \mathcal{A}_{11} derrota a \mathcal{A}_6 , y \mathcal{A}_6 no derrota a \mathcal{A}_{11} . Al construir el árbol de dialéctica para \mathcal{A}_{11} se obtendría un árbol con un único nodo ya que no existe ningún argumento activo que derrote a \mathcal{A}_{11} , y por ende, la raíz de ese

árbol estaría marcada como **U**, lo que significa que \mathcal{A}_{11} estaría aceptado, y su conclusión ($better(box_1, box_2)$) estaría garantizada. Note que la raíz del árbol de dialéctica para \mathcal{A}_6 estaría marcada como **D** ya que \mathcal{A}_{11} derrota a \mathcal{A}_6 , y no existe un derrotador para \mathcal{A}_{11} . Algo similar ocurre entre \mathcal{A}_{12} y \mathcal{A}_5 , por lo que la conclusión de \mathcal{A}_{12} ($\neg better(box_2, box_1)$) también está garantizada. Además, note que el resto de los argumentos activos no están en conflicto con ningún otro argumento, por lo que sus conclusiones también están garantizadas. Esto quiere decir que el conjunto de conclusiones garantizadas en \mathcal{K}_F es $W_F = \{better(box_1, box_3), better(box_2, box_3), better(box_1, box_2), \neg better(box_3, box_1), \neg better(box_3, box_2), \neg better(box_2, box_1)\}$.

Las conclusiones garantizadas son procesadas por el marco de decisión para determinar cuáles son las mejores alternativas. Para esta tarea se utiliza la componente de decisión (Γ), la cual se especifica un conjunto de reglas de decisión ([FEGS08]). Dado un subconjunto de alternativas $B \subseteq \mathcal{D}$ ($B \neq \emptyset$), también conocido como experimento de elección, una regla de decisión ($D \stackrel{B}{\Leftarrow} P, T$) debe interpretarse como si cumplen todas las pre-condiciones del conjunto P , y no se cumple ninguna de las restricciones del conjunto T , entonces de B debe seleccionarse el subconjunto de alternativas D . Estos autores, definen la componente de decisión a partir de un conjunto especial de reglas de decisión. Dado un marco de decisión $\mathbf{F} = \langle \mathcal{D}, \mathcal{C}, >_{\mathcal{C}}, \mathcal{K}, \Gamma \rangle$, un experimento de elección $B \subseteq \mathcal{D}$, y sean W, Y y Z tres variables esquemáticas. La componente de decisión de \mathbf{F} es $\Gamma = DR_1 \cup DR_2$, donde:

- DR_1 es el conjunto de reglas de decisión obtenidas al instanciar las variables W, Y y Z con las alternativas de B en la regla de decisión

$$\{W\} \stackrel{B}{\Leftarrow} \{better(W, Y)\}, \{better(Z, W)\}$$

- DR_2 es el conjunto de reglas de decisión obtenidas al instanciar las variables W, Y y Z con las alternativas de B en la regla de decisión

$$\{W, Y\} \stackrel{B}{\Leftarrow} \{\neg better(W, Y), \neg better(Y, W)\}, \{better(Z, W), better(Z, Y)\}$$

Por un lado, las reglas de decisión del conjunto DR_1 establecen que una alternativa $W \in B$ será seleccionada si W es mejor que otra alternativa Y , y no existe una alternativa Z que sea mejor que W . Por otro lado, las reglas del conjunto DR_2 establecen que dos alternativas X e Y contenidas en B serán seleccionadas si se puede asegurar que W no es mejor que Y , que Y no es mejor que W , y además, no existe una alternativa Z que sea mejor que ambas.

Luego, para determinar cuáles son las reglas aplicables, y así determinar cuales son las mejores alternativas, se utiliza el conjunto de conclusiones garantizadas. Dado un marco de decisión $\mathbf{F} = \langle \mathcal{D}, \mathcal{C}, >_{\mathcal{C}}, \mathcal{K}, \Gamma \rangle$, una de las reglas de decisión $D \stackrel{B}{\Leftarrow} P, T$ en Γ es aplicable con respecto a \mathcal{K} , si todas las pre-condiciones P pertenecen al conjunto de conclusiones garantizadas W de \mathcal{K} , y ninguna de las restricciones T pertenece a W . Finalmente, si una regla es aplicable, las alternativas en D son seleccionadas.

Para finalizar con el ejemplo de aplicación de [FTG⁺17], a partir del conjunto de conclusiones garantizadas $W_F = \{better(box_1, box_3), better(box_2, box_3), better(box_1, box_2), \neg better(box_3, box_1), \neg better(box_3, box_2), \neg better(box_2, box_1)\}$, y considerando a $\mathcal{D}_F = \{box_1, box_2, box_3\}$ como el experimento de elección, se obtiene que las únicas reglas de decisión aplicables deben ser de la forma $\{box_1\} \stackrel{B}{\Leftarrow} \{better(box_1, Y)\}, \{better(Z, box_1)\}$. Por lo tanto, la única alternativa seleccionada sería box_1 . Note que como $\{better(box_1, box_2), better(box_2, box_3)\} \subset W_F$, ni box_2 , ni box_3 pueden ser seleccionadas ya que se violarían las restricciones de la forma $better(Z, box_2)$ y $better(Z, box_3)$.

Observe que en esta propuesta las preferencias del agente son modeladas por medio del conjunto de criterios \mathcal{C} , y el orden total estricto entre los mismo $>_{\mathcal{C}}$. Esto hace que las preferencias que se pueden modelar siguiendo este enfoque sean bastante limitadas. Por lo tanto, en el marco propuesto por Ferreti et al. no se permite modelar preferencias más complejas que traten de balancear la importancia que se le da a cada criterio, como por ejemplo, no se podría modelar algo como: “aunque la distancia al robot sea mayor, si esta es menor a dos, entonces sería preferible escoger la caja más pequeña”, o “mientras la diferencia de tamaño no sea demasiada, es preferible escoger la caja que esté mas cerca del almacenamiento”. De hecho, sólo a través de \mathcal{C} y $>_{\mathcal{C}}$ uno podría determinar cuál es la mejor alternativas sin tener que realizar todo el proceso argumentativo antes mencionado. Las mejores alternativas son aquellas que poseen el mejor valor con respecto al criterio más importante según $>_{\mathcal{C}}$ y, si todas las alternativas son igual de buenas con respecto a ese criterio, se pasa a evaluar el segundo más importante, y así siguiendo. Si todas las alternativas son igual de buenas en todos los criterios, entonces se seleccionan todas las alternativas.

6.3. Sistemas argumentativos para toma de decisiones con múltiples agentes

Al considerar varios agentes, el problema de toma de decisiones se vuelve más complejo ya que surgen otras problemáticas que deben resolverse, como por ejemplo: ¿qué sucede cuando dos (o más) agentes poseen intereses en conflicto?, ¿qué sucede si dos (o más) agentes no coinciden en las características de una alternativa?, o ¿qué sucede si los agentes poseen información que no quieren revelar ante los demás?. A continuación, se resumen dos propuestas que atacan el problema desde diferentes perspectivas.

6.3.1. Toma de decisiones entre dos agentes mediante diálogos ABA [FTMW14]

En esta propuesta, los autores presentan una solución al problema de toma de decisiones entre dos agentes, la cual considera un protocolo de diálogo argumentativo basado en suposiciones. La idea principal detrás de esta propuesta, es que a partir del diálogo, los dos agentes puedan acordar cuáles son las mejores alternativas, de acuerdo con la noción de dominancia débil en el contexto de un marco de decisión de Fan et al. [FTMW14] (tal y como se explicó en la Sección 6.2.2), y considerando los conocimientos y preferencias de ambos agentes.

En línea con el trabajo de [ZFLT14], estos autores consideran que el conocimiento (y preferencias) de cada agente deberá estar representado mediante un ABA tal que las conclusiones de los argumentos admisibles estén alineadas con el conjunto de alternativas dominantes que son identificadas para un dado marco de decisión de Fan et al. [FT14], pero con la posibilidad de que el conocimiento de los agentes no sea completo, en el sentido de que los agentes pueden desconocer si una alternativa posee cierto atributo, o que si un atributo está (o no) asociado a la satisfacción de una meta. La Figura 6.9 muestra las tablas \mathcal{DF}_x y \mathcal{GF}_x ($x \in \{1, 2\}$) pertenecientes a los marcos de decisión que modelan el conocimiento (y preferencias) de dos agentes \mathbf{a}_1 y \mathbf{a}_2 acerca de posibles alojamientos en Londres (ejemplo tomado de [FTMW14]). Los departamentos que los agentes conocen son identificados como $depA$, $depB$ y $depC$, mientras que los atributos que pueden tener asociados son: $\$5k$ (si el alquiler ronda los \$5.000), $\$7k$ (si el alquiler ronda los \$7.000),

$\$10k$ (si el alquiler ronda los $\$10.000$), $enSK$ (si se encuentra en South Kenington), $enPic$ (si se encuentra en Piccadilly) y $cPriv$ (si la entrada es por una calle privada).

\mathcal{DF}_1	$\$5k$	$\$7k$	$\$10k$	$enSK$	$enPic$	$cPriv$
$depA$	1	0	0	u	u	0
$depB$	0	0	1	0	1	0

\mathcal{GF}_1	$\$5k$	$\$7k$	$\$10k$	$enSK$	$enPic$	$cPriv$
$económico$	1	0	0	0	0	0
$cercano$	0	0	0	0	1	0
$tranquilo$	0	0	0	0	0	1

\mathcal{DF}_2	$\$5k$	$\$7k$	$\$10k$	$enSK$	$enPic$
$depA$	1	0	0	1	0
$depC$	0	1	0	1	0

\mathcal{GF}_2	$\$5k$	$\$7k$	$\$10k$	$enSK$	$enPic$
$económico$	1	0	0	0	0
$cercano$	0	0	0	1	0

Figura 6.9: Tablas \mathcal{DF}_x y \mathcal{GF}_x ($x \in \{1, 2\}$) pertenecientes a los marcos de decisión de los agentes \mathbf{a}_1 y \mathbf{a}_2 del ejemplo presentado en [FTMW14].

Note que el conjunto de alternativas que cada agente conoce no es el mismo, pero ambos conocen la alternativa $depA$. Además, note que \mathbf{a}_1 no sabe (marcado como u) si $depA$ posee el atributo $enSK$ ni si posee el atributo $enPic$, aunque \mathbf{a}_2 sí posee esta información, pero \mathbf{a}_2 ignora la existencia del atributo $cPriv$. Más aún, note que \mathbf{a}_1 posee una meta ($tranquilo$) que \mathbf{a}_2 no.

Luego, a la hora de combinar el conocimiento y preferencias de ambos agentes, los autores de esta propuesta consideran que uno de los agentes debe ser más confiable que el otro. Luego, el conocimiento y preferencias del agente más confiable tendrá prioridad por sobre el del otro, siendo posible eliminar las inconsistencias que existen entre las creencias de estos agentes. Así, considerando que \mathbf{a}_1 es más confiable que \mathbf{a}_2 , y sean

$\mathbf{M}_1 = \langle \mathcal{D}_1, \mathcal{F}_1, \mathcal{G}_1, \mathcal{DF}_1, \mathcal{GF}_1 \rangle$ y $\mathbf{M}_2 = \langle \mathcal{D}_2, \mathcal{F}_2, \mathcal{G}_2, \mathcal{DF}_2, \mathcal{GF}_2 \rangle$ los marcos de decisión de cada uno de ellos (respectivamente), el marco de decisión que representa los conocimientos y preferencias de ambos agentes se define como $\mathbf{M}_J = \langle \mathcal{D}_J, \mathcal{F}_J, \mathcal{G}_J, \mathcal{DF}_J, \mathcal{GF}_J \rangle$, donde:

- $\mathcal{D}_J = \mathcal{D}_1 \cup \mathcal{D}_2$; • $\mathcal{F}_J = \mathcal{F}_1 \cup \mathcal{F}_2$; • $\mathcal{G}_J = \mathcal{G}_1 \cup \mathcal{G}_2$;
- $\mathcal{DF}_J[i, j] = \begin{cases} \mathcal{DF}_1[i, j] & \text{si } \mathcal{DF}_1[i, j] = 1 \text{ o } \mathcal{DF}_1[i, j] = 0, \\ \mathcal{DF}_2[i, j] & \text{si } \mathcal{DF}_1[i, j] = u, \text{ y } \mathcal{DF}_2[i, j] = 1 \text{ o } \mathcal{DF}_2[i, j] = 0, \\ u & \text{en caso contrario.} \end{cases}$
- $\mathcal{GF}_J[k, j] = \begin{cases} \mathcal{GF}_1[k, j] & \text{si } \mathcal{GF}_1[k, j] = 1 \text{ o } \mathcal{GF}_1[k, j] = 0, \\ \mathcal{GF}_2[k, j] & \text{si } \mathcal{GF}_1[k, j] = u, \text{ y } \mathcal{GF}_2[k, j] = 1 \text{ o } \mathcal{GF}_2[k, j] = 0, \\ u & \text{en caso contrario.} \end{cases}$

con $1 \leq i \leq |\mathcal{D}_J|$, $1 \leq j \leq |\mathcal{F}_J|$, y $1 \leq k \leq |\mathcal{G}_J|$.

\mathcal{DF}_J	\$5k	\$7k	\$10k	<i>enSK</i>	<i>enPic</i>	<i>cPriv</i>
<i>depA</i>	1	0	0	1	0	0
<i>depB</i>	0	0	1	0	1	0
<i>depC</i>	0	1	0	1	0	<i>u</i>

\mathcal{GF}_J	\$5k	\$7k	\$10k	<i>enSK</i>	<i>enPic</i>	<i>cPriv</i>
<i>económico</i>	1	0	0	0	0	0
<i>cercano</i>	0	0	0	0	1	0
<i>tranquilo</i>	0	0	0	0	0	1

Figura 6.10: Tablas \mathcal{DF}_J y \mathcal{GF}_J pertenecientes al marco de decisión que se obtiene al combinar los marcos de decisión de los agentes \mathbf{a}_1 y \mathbf{a}_2 del ejemplo presentado en [FTMW14].

La Figura 6.10 muestra las tablas \mathcal{DF}_J y \mathcal{GF}_J . Observe que al combinar los marcos de decisión, la información que era desconocida por \mathbf{a}_1 , pero conocida por \mathbf{a}_2 pudo completarse. Sin embargo, como \mathbf{a}_1 no conoce la alternativa *depC*, y \mathbf{a}_2 no considera la existencia del atributo *cPriv*, es imposible saber si *depC* posee dicho atributo. Más allá de esto, note que la única alternativa débilmente dominante con respecto a \mathbf{M}_J es *depA* ya que cumple con las metas *económico* y *cercano*, y que ninguna alternativa cumple con la

meta *tranquilo*. Por lo tanto, de acuerdo con el conocimiento y preferencias de los dos agentes, *depA* debería ser la alternativa seleccionada. A continuación, se explicará como los agentes llegarán a esta conclusión a partir de un diálogo. Sin embargo, para que esto sea posible, los autores requieren que el conocimiento de los dos agentes sea representado a partir de un marco ABA cuyas conclusiones admisibles permitan determinar cuándo una alternativa es débilmente dominante con respecto al marco de decisión que modela el conocimiento de cada agente.

Sean \mathbf{a}_x y \mathbf{a}_y dos agentes ($x, y \in \{1, 2\}$), y $\mathbf{M}_x = \langle \mathcal{D}_x, \mathcal{F}_x, \mathcal{G}_x, \mathcal{DF}_x, \mathcal{GF}_x \rangle$ el marco de decisión de \mathbf{a}_x , el ABA multi-agente para alternativas (débilmente) dominantes $\langle \mathbf{L}, \mathbf{R}_x, \mathbf{A}_x, \mathbf{C}_x \rangle$ correspondiente a \mathbf{M}_x se compone de la siguiente manera:

- \mathbf{R}_x consiste de las siguientes reglas (y ninguna más):

- $\mathbf{a}_x > \mathbf{a}_y \leftarrow$, si \mathbf{a}_x es más confiable que \mathbf{a}_y ;
- $isD(d) \leftarrow, \forall d \in \mathcal{D}_x; \quad isF(d) \leftarrow, \forall f \in \mathcal{F}_x; \quad isG(g) \leftarrow, \forall g \in \mathcal{G}_x;$
- $nHasAttr(d, f, \mathbf{a}_x) \leftarrow, \forall d \in \mathcal{D}, f \in \mathcal{F}$ tal que \mathbf{a}_1 considere que d no posea a f ;
- $nSatBy(g, f, \mathbf{a}_x) \leftarrow, \forall g \in \mathcal{G}, f \in \mathcal{F}$ tal que \mathbf{a}_1 considere que f no satisface a g ;
- $\neg hasAttr(D, F, Z) \leftarrow nHasAttr(D, F, Z'), Z' > Z;$
- $\neg satBy(D, F, Z) \leftarrow nSat(D, F, Z'), Z' > Z;$
- $hasAttr(D, F) \leftarrow hasA(D, F, Z), isD(D), isF(F);$
- $satBy(G, F) \leftarrow satBy(G, F, Z), isD(D), isF(F);$
- $met(D, G) \leftarrow hasA(D, F), satBy(G, F);$
- $notSel(D) \leftarrow nMet(D, G), isD(D), isG(G);$
- $othersMet(D, G) \leftarrow met(D', G), D \neq D'.$

- \mathbf{A}_x contiene los siguientes elementos (y ningún otro):

- $hasAttr(d, f, \mathbf{a}_x), \forall d \in \mathcal{D}, f \in \mathcal{F}$ tal que \mathbf{a}_x considere que d posea a f ;
- $satBy(g, f, \mathbf{a}_x), \forall g \in \mathcal{G}, f \in \mathcal{F}$ tal que \mathbf{a}_x considere que f satisface a g ;
- $nMet(d, g), \forall d \in \mathcal{D}, g \in \mathcal{G};$
- $none(d, g), \forall d \in \mathcal{D}, g \in \mathcal{G};$
- $sel(d), \forall d \in \mathcal{D}.$

- \mathbf{C}_x es tal que:

- $\mathbf{C}_x(hasAttr(D, F, Z)) = \{\neg hasAttr(D, F, Z)\};$
- $\mathbf{C}_x(satBy(D, F, Z)) = \{\neg satBy(D, F, Z)\};$
- $\mathbf{C}_x(sel(D)) = \{notSel(D)\};$

$$C_x(nMet(D, G)) = \{met(D, G), none(D, G)\};$$

$$C_x(none(D, G)) = \{othersMet(D, G)\}.$$

Por como está diseñado el marco ABA que representa el conocimiento de los dos agentes, ante la ausencia de información, no se agrega ninguna regla de la forma $(nHasAttr(D, F, A) \leftarrow)$ o $(nSat(G, F, A) \leftarrow)$, ni tampoco ninguna suposición de la forma $hasAttr(D, F, A)$ o $satBy(G, F, A)$. Esto permite que durante el diálogo los agentes complementen su conocimiento, haciendo que cada uno pueda completar el conocimiento que el otro no posee, sin caer en contradicciones. Por otro lado, en caso de que los agentes tengan creencias u opiniones contradictorias (como por ejemplo, que uno crea que cierta alternativa posee un atributo, que el otro cree que no lo posee), los autores consideran que los agentes tendrán asignado un valor de confianza, y se le dará prioridad a la información del agente más confiable, determinado por la regla $(\mathbf{a}_x > \mathbf{a}_y \leftarrow)$. De este modo, si se combinaran⁴ los dos marcos ABA, suponiendo que el agente \mathbf{a}_y (el menos confiable) cree que una alternativa d posee el atributo f ($hasA(d, f, \mathbf{a}_y)$), mientras que \mathbf{a}_x crea que esto no es así ($nHasAttr(d, f, \mathbf{a}_x)$), se podrá construir un argumento para refutar $hasAttr(d, f, \mathbf{a}_y)$, a partir de la regla $\neg hasAttr(D, F, Z) \leftarrow nHasAttr(D, F, Z'), Z' > Z$ instanciada como $\neg hasAttr(d, f, \mathbf{a}_y) \leftarrow nHasAttr(d, f, \mathbf{a}_x), \mathbf{a}_x > \mathbf{a}_y$. Note que esto es análogo a lo que ocurre con $satBy(D, F, Z)$ y $\neg satBy(D, F, Z)$.

En cuanto a determinar si una alternativa es (o no) débilmente dominante, en principio se asume que toda alternativa d puede ser seleccionada ($sel(d)$) (es decir, que en principio todas son consideradas dominantes). Esto puede ser refutado si d no cumple con alguna meta g siguiendo la regla $notSel(D) \leftarrow nMet(D, G), isD(D), isG(G)$. Sin embargo, si se tiene que d tiene f y que f satisface a g , entonces se podrá construir un argumento para refutar a $nMet(d, g)$ a partir de la regla $(met(D, G) \leftarrow hasA(D, F), satBy(G, F))$. Por otro lado, también se asume que salvo por d , ninguna otra alternativa cumple la meta g ($none(d, g)$), lo cual también contradice a $nMet(d, g)$. No obstante, si existe una alternativa distinta de d que cumpla con g , entonces se podrá construir un argumento para refutar $none(d, g)$ a partir de la regla $(othersMet(D, G) \leftarrow met(D', G), D \neq D')$.

Luego, el computo de las alternativas dominantes puede ser llevado a cabo mediante un diálogo-ABA [FT11, FT12] entre los dos agentes \mathbf{a}_x y \mathbf{a}_y cuyos marcos ABA son $\langle \mathbf{L}, \mathbf{R}_x, \mathbf{A}_x, \mathbf{C}_x \rangle$ y $\langle \mathbf{L}, \mathbf{R}_y, \mathbf{A}_y, \mathbf{C}_y \rangle$, respectivamente (note que ambos agentes deben compartir

⁴La combinación de dos marcos ABA $\mathbf{F}_1 = \langle \mathbf{L}, \mathbf{R}_1, \mathbf{A}_1, \mathbf{C}_1 \rangle$ y $\mathbf{F}_2 = \langle \mathbf{L}, \mathbf{R}_2, \mathbf{A}_2, \mathbf{C}_2 \rangle$ es el marco ABA $\mathbf{F}_J = \mathbf{F}_1 \uplus \mathbf{F}_2 = \langle \mathbf{L}, \mathbf{R}_1 \cup \mathbf{R}_2, \mathbf{A}_1 \cup \mathbf{A}_2, \mathbf{C}_J \rangle$, donde $\mathbf{C}_J(\alpha) = \mathbf{C}_1 \cup \mathbf{C}_2$ para todo $\alpha \in \mathbf{A}_1 \cup \mathbf{A}_2$.

el mismo lenguaje). Un diálogo-ABA está compuesto por declaraciones (o utterances, en inglés) de la forma $\langle \mathbf{a}_x, \mathbf{a}_y, T, C, ID \rangle$ ($x, y = 1, 2, x \neq y$), donde \mathbf{a}_x es el remitente, \mathbf{a}_y es el destinatario, C es el contenido de la declaración, siendo una de las siguientes: a) $claim(\beta)$ para alguna conclusión $\beta \in \mathbf{L}$; b) $rl(\beta_0 \leftarrow \beta_1, \dots, \beta_m)$ para alguna regla con $\beta_k \in \mathbf{L}$ ($0 \leq k \leq m$); c) $asm(\alpha)$ para alguna suposición $\alpha \in \mathbf{L}$; d) $ctr(\alpha, \beta)$ para identificar que $\alpha \in \mathbf{L}$ y $\beta \in \mathbf{L}$ son contrarios; o bien, e) la sentencia de paso $\pi \notin \mathbf{L}$; $ID \in \mathbb{N}$ es el identificador de la declaración; y $T \in \mathbb{N} \cup \{0\}$ ($T < ID$) es el identificador objetivo. El identificador de una declaración representa la posición de la declaración en el diálogo (comenzando en 1), y su identificador objetivo debe ser igual al identificador de una declaración previa cuyo contenido se relaciona con el de la declaración actual, salvo para la primer declaración cuyo objetivo debe ser 0.

Formalmente, un diálogo-ABA entre dos agentes \mathbf{a}_x y \mathbf{a}_y acerca de un literal $\chi \in \mathbf{L}$, denotado $D_{\mathbf{a}_y}^{\mathbf{a}_x}(\chi)$, es una secuencia finita de declaraciones $\delta = [u_1, \dots, u_n]$ ($n \geq 0$) entre \mathbf{a}_x y \mathbf{a}_y , tal que \mathbf{a}_x es el remitente de u_1 y su contenido es $claim(\chi)$; el objetivo de cada declaración de paso (aquellas cuyo contenido es π) es 0; para cada declaración $u_j = \langle -, -, T, -, - \rangle^5$ ($j > 1$), con $T \neq 0$, existe una declaración $u_i = \langle -, -, -, C, T \rangle$ ($i < j$), con $C \neq \pi$; y no existen dos declaraciones de paso seguidas, excepto por las últimas dos.

En esta propuesta, Fan et al., demostraron que si el diálogo-ABA que es llevado a cabo por los dos agentes cumple con ciertas restricciones, el literal a partir del cual surge el diálogo (χ) se corresponde con la conclusión de un argumento admisible del marco ABA que se obtiene al combinar los marcos ABA de los dos agentes. Sea $\delta = [u_1, \dots, u_n]$ un diálogo-ABA y sea $\mathbf{M}_\delta = \langle \mathbf{L}, \mathbf{R}_\delta, \mathbf{A}_\delta, \mathbf{C}_\delta \rangle$ el marco ABA que se extrae de δ , donde:

- $\mathbf{R}_\delta = \{\rho \mid rl(\rho) \text{ es el contenido de algún } u_i \text{ en } \delta\}$;
- $\mathbf{A}_\delta = \{\alpha \mid asm(\alpha) \text{ es el contenido de algún } u_i \text{ en } \delta\}$;
- $\mathbf{C}_\delta(\alpha) = \{\beta \mid ctr(\alpha, \beta) \text{ es el contenido de algún } u_i \text{ en } \delta\}$.

Si se cumple que:

1. El marco ABA extraído del diálogo debe ser plano, es decir, ninguna regla tiene una suposición en su cabeza, y toda suposición tiene un conjunto no vacío de contrarios;
2. Salvo por la primer declaración, todas las demás son relativas a otra. Una declaración $u_j = \langle -, -, T, C_j, - \rangle$ es relativa a otra declaración $u_i = \langle -, -, -, C_i, ID \rangle$ ($i < j$) si y sólo si $T = ID$ y se cumple uno de los siguiente casos:

⁵Por $-$ se entiende una variable anónima como las de Prolog.

- a) u_j declara la existencia de una regla cuya cabeza aparece dentro del contenido de u_i . Esto es, $C_j = rl(\rho_j)$ con $Head(\rho_j) = \beta$ y $C_i = rl(\rho_i)$ con $\beta \in Body(\rho_i)$, o $C_i = ctr(\alpha, \beta)$, o bien $C_i = claim(\beta)$.
 - b) u_j declara la existencia de una suposición que aparece dentro del contenido de u_i . Esto es, $C_j = asm(\alpha)$ y $C_i = rl(\rho)$ con $\alpha \in Body(\rho)$, o $C_i = ctr(\alpha', \alpha)$, o bien $C_i = claim(\alpha)$.
 - c) u_j indica que una contradicción para la suposición indicada en u_j . Esto es, $C_j = ctr(\alpha, \beta)$ y $C_i = asm(\alpha)$.
3. El diálogo es exhaustivo, es decir, no puede quedar fuera del dialogo ninguna declaración que pueda ser construida por un agente y sea relativa a otra declaración del diálogo, siempre y cuando no existan dos declaraciones cuyo objetivo y contenido sean el mismo.

entonces, se dice que el diálogo es coherente, y por lo tanto, el literal a partir del cual surge el diálogo se corresponde con la conclusión de un argumento admisible del marco ABA que se obtiene al combinar los marcos ABA de los dos agentes. Cabe aclarar, que los autores de esta propuesta consideran que los agentes deben construir los diálogos haciendo uso de una función que a partir de su conocimiento y el dialogo llevado hasta el momento, retorne la siguiente declaración que debería ser introducida en el diálogo. Sin embargo, los autores no especifican formalmente cómo construir dicha función, ni tampoco cómo se arbitra la participación de los agentes en el diálogo (es decir, cuándo es el turno de cada agente).

De esta forma, es posible determinar cuáles son las alternativas débilmente dominantes de acuerdo con el marco de decisión que se obtendría al combinar los marcos de decisión que representan el conocimiento de los agentes. Los autores demostraron que una alternativa d es débilmente dominante con respecto a un marco de decisión combinado \mathbf{M}_J si y sólo si $sel(d)$ es la conclusión de un argumento admisible en el marco ABA que se obtiene al combinar los marcos ABA multi-agente para alternativas débilmente dominantes de los agentes. Por lo tanto, dada una alternativa d si $D_{\mathbf{a}_y}^{\mathbf{a}_x}(sel(d))$ es coherente, entonces $sel(d)$ es una conclusión admisible, lo que significa que d es una alternativa dominante con respecto al marco de decisión combinado de los agentes \mathbf{a}_x y \mathbf{a}_y .

A continuación, se muestra un fragmento del diálogo $D_{a_2}^{a_1}(sel(depA))$ extraído de [FT11]. Las metas *económico* y *cercano* están abreviada como *ec* y *ce*, respectivamente.

$\langle a_1, a_2, 0, claim(sel(depA)), 1 \rangle$
 $\langle a_2, a_1, 1, asm(sel(depA)), 2 \rangle$
 $\langle a_1, a_2, 2, ctr(sel(depA), notSel(depA)), 3 \rangle$
 $\langle a_2, a_1, 3, rl(notSel(depA) \leftarrow nMet(depA, ec), isD(depA), isG(ec)), 4 \rangle$
 $\langle a_1, a_2, 4, asm(nMet(depA, ec)), 5 \rangle$
 $\langle a_2, a_1, 4, rl(isD(depA) \leftarrow), 6 \rangle$
 $\langle a_1, a_2, 4, rl(isG(ec) \leftarrow), 7 \rangle$
 $\langle a_1, a_2, 5, ctr(nMet(depA, ec), met(depA, ec)), 8 \rangle$
 $\langle a_1, a_2, 8, rl(met(depA, ec) \leftarrow hasAttr(depA, \$5k), satBy(ec, \$5k)), 9 \rangle$
 $\langle a_2, a_1, 9, rl(hasAttr(depA, \$5k) \leftarrow hasAttr(depA, \$5k, a_2)), 10 \rangle$
 $\langle a_2, a_1, 10, asm(hasAttr(depA, \$5k, a_2)), 11 \rangle$
 $\langle a_2, a_1, 9, rl(satBy(ec, \$5k) \leftarrow satBy(ec, \$5k, a_2)), 12 \rangle$
 $\langle a_2, a_1, 12, asm(satBy(ec, \$5k, a_2)), 13 \rangle$
 $\langle a_2, a_1, 11, ctr(hasAttr(depA, \$5k, a_2), \neg hasAttr(depA, \$5k, a_2)), 14 \rangle$
 $\langle a_2, a_1, 12, ctr(satBy(ec, \$5k, a_2), \neg satBy(ec, \$5k, a_2)), 15 \rangle$
 $\langle a_2, a_1, 14, rl(\neg hasAttr(depA, \$5k, a_2) \leftarrow nHasAttr(depA, \$5k, a_1), a_1 > a_2), 16 \rangle$
 $\langle a_1, a_2, 16, rl(a_1 > a_2 \leftarrow), 17 \rangle$
 $\langle a_2, a_1, 14, rl(\neg satBy(ec, \$5k, a_2) \leftarrow nSatBy(depA, \$5k, a_1), a_1 > a_2), 18 \rangle$
 $\langle a_1, a_2, 18, rl(a_1 > a_2 \leftarrow), 19 \rangle$
 $\langle a_2, a_1, 3, rl(notSel(depA) \leftarrow nMet(depA, ce), isD(depA), isG(ce)), 20 \rangle$
 ...

En este diálogo se pueden ver las primeras veinte declaraciones en las que los dos agentes discuten para determinar si *depA* cumple con la meta *económico*. Ellos verifican que el costo de *depA* es $\$5k$ y que *económico* es satisfecho al dejar que a_2 introduzca las dos suposiciones $hasAttr(depA, \$5k, a_2)$ y $satBy(económico, \$5k, a_2)$ sin recibir ninguna objeción por parte de a_1 . Luego de esto, los agentes seguirían discutiendo y confirmarían que *depA* cumple con *cercano* y que ninguna alternativa cumple con *tranquilo*, por lo que *depA* es la única alternativa débilmente dominante.

Una primera observación con respecto a la comparación entre el trabajo presentado en esta tesis y la propuesta de Fan et al. es que esta última sólo contempla la existencia de

dos agentes, mientras que en la otra no existe ninguna restricción con respecto al número de agentes. Más allá de esto, la principal diferencia entre las dos propuestas subyace en la forma en la que se resuelve el problema ya que en Fan et al. plantean resolverlo por medio de un diálogo. Sin embargo, a la hora de implementar la propuesta habría que evaluar la conveniencia entre esto y hacer que los agentes compartan los marcos de decisión que representan sus creencias, que luego estos marcos sean combinados, y que a partir del marco de decisión combinado se obtenga el marco-ABA equivalente, y finalmente obtener el conjunto de alternativas dominantes de las conclusiones de los argumentos admisibles (sin la necesidad de que los agentes entablen un diálogo). Además de ofrecer las mismas ventajas con respecto al uso de un marco-argumentativo para obtener los resultados, esta otra alternativa parece ser más sencilla a la hora de implementar, ya que no es necesario definir una función que determine cual debe ser la próxima declaración de un agente dentro de un diálogo (teniendo en cuenta que esta función no fue definida en [FTMW14]), ni tampoco es necesario definir la manera en la que el diálogo será coordinado. Por lo tanto, y sin ánimo de menospreciar el esfuerzo de los autores por adaptar el uso de los diálogos-ABA para resolver problemas de toma de decisiones entre dos agentes, uno podría concluir que quizás resolver este tipo de problemas mediante un diálogo podría no ser tan ventajoso frente a la idea de que los agentes compartan cierta información con la que luego se pueda instanciar un sistema argumentativo que resuelva el problema, tal y como sucede en la propuesta presentada en esta tesis con respecto al conjunto de evidencia o al conjunto de *DePref*-rules. Por otro lado, como la propuesta de [FTMW14] esta basada en los marcos de Fan et al. ([FT14]), al igual que la propuesta presentada en [ZFLT14] (ver Sección 6.2.2), la forma en la que los agentes plantean sus preferencias difiere mucho con respecto a como podrían hacerlo mediante el uso de *DePref*-rules. Al igual que la propuesta presentada en esta tesis, la propuesta de Fan et al. también contempla la posibilidad de que cierta parte de evidencia o información que se tenga con respecto a las alternativas no sea conocida por ninguno de los agentes, es decir, es posible trabajar con evidencia incompleta. Otro aspecto a tener en cuenta de la propuesta de Fan et al., es que consideran que uno de los agentes debe ser más confiable que el otro, y le otorga prioridad al conocimiento del más confiable. Dicha limitación fue contemplada por lo autores, por lo que en su trabajo incorporan una segunda versión de los marcos argumentativos en la que se considera que los agentes son igual de confiables, y en la cuál se terminan considerando sólo los hechos en los que ambos agentes están de acuerdo. En lo que respecta al marco de trabajo presentado en esta tesis, la prioridad de los agentes

es definida por medio de un orden parcial, por lo que pueden existir agentes de igual prioridad.

6.3.2. Toma de decisiones basada en argumentación entre múltiples agentes con información privada [GTWX16]

En este artículo los autores presentan un protocolo de cooperación basado en argumentación abstracta, inspirado en los algoritmos de resolución de problemas de satisfacción de restricciones distribuidos (o DisCSP, por sus siglas en inglés)[YH00].

Los DisCSP considerados por estos autores se describen como una tupla $\langle Ag, Act, Con \rangle$, donde $Ag = \{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ es el conjunto de agentes ($n \geq 2$), $Act = \{\alpha_1, \dots, \alpha_m\}$ ($m \geq 1$) es el conjunto de acciones disponibles, y $Con : Act \rightarrow 2^{\{0, \dots, n\}}$ es la función de restricciones. Dado una acción $\alpha \in Act$, $Con(\alpha) \subseteq \{0, \dots, n\}$ es el número de agentes requeridos para llevar a cabo dicha acción. Suponga que una acción α debe ser llevada a cabo por 0 o 2 agentes simultáneamente (por ejemplo, porque se desea modelar que dos agentes deben concurrir juntos, o no concurrir a un cierto espectáculo), entonces $Con(\alpha) = \{0, 2\}$.

Un DisCSP se resuelve al encontrar (de forma distribuida) una estrategia con la cual asignar tareas a los agentes de forma tal que se cumplan las restricciones del problema. Una estrategia es una lista $[\alpha_1, \dots, \alpha_n]$, donde $\alpha_i \in Act$ es la acción que será llevada a cabo por el agente $\mathbf{a}_i \in Ag$. Dado un DisCSP $\langle Ag, Act, Con \rangle$, y sea $N(S, \alpha)$ el número de agentes asignados a α de acuerdo con una estrategia S , se dice que S es aceptable si $\forall \alpha \in Act, N(S, \alpha) \in Con(\alpha)$.

En [GTWX16], se presenta un ejemplo en el que dos agentes (Alice y Bob) deben decidir entre ir a ver un partido de fútbol, o ir al ballet. La única restricción es que, cualquiera sea la decisión, deben ir juntos. Esto puede modelarse por medio de un DisCSP con dos agentes (Alice y Bob), dos acciones (ir al partido de fútbol e ir al ballet) y $Con(partido) = Con(ballet) = \{0, 2\}$. Las únicas dos estrategias aceptables que existen para resolver este problema son aquellas que asignan a Alice y a Bob simultáneamente a ir al partido, o bien, a ir al ballet.

La idea detrás de esta propuesta es que, mediante un protocolo de cooperación basado en argumentos, los agentes puedan generar una estrategia capaz de resolver un DisCSP, de

forma tal que se traten de satisfacer las preferencias de los agentes, sin revelar información que consideren privada. Para esto, cada agente $\mathbf{a}_i \in Ag$ debe contar con un orden de preferencia entre las acciones \geq_i , y un AAF $\mathbf{F}_i = \langle Arg_{s_i}, Att_{s_i} \rangle$ con el que se modela su conocimiento. El conjunto de argumentos Arg_{s_i} se divide en tres conjuntos disjuntos: los argumentos prácticos privados $pPra_i$, los argumentos epistémicos privados $pEpi_i$, y los argumentos epistémicos revelables (o públicos) $rEpi_i$. Al igual que en [AP09], los argumentos prácticos representan las acciones o decisiones consideradas por el agente, y los argumentos epistémicos representan el conocimiento del agente. También en línea con [AP09] sólo los argumentos epistémicos pueden atacar a otros argumentos. Gao et al. consideran que los argumentos prácticos deben ser privados ya que cada agente es libre de considerar sólo las acciones que este desee y los demás agentes no tienen por qué conocer estas acciones. Para esta propuesta se considera, además, que todo argumento epistémico privado debe ser atacado por algún argumento epistémico revelable y que ningún argumento epistémico revelable puede ser atacado por uno epistémico privado. Esto obliga a que los agentes están abiertos a cambiar de opinión con respecto a la aceptación de sus creencias privadas. De aquí en adelante, dado un agente $\mathbf{a}_i \in Ag$ tal que $\mathbf{F}_i = \langle Arg_{s_i}, Att_{s_i} \rangle$ es su AAF, $\mathcal{A}_{ij} \in Arg_{s_{ij}}$ identifica el argumento práctico de \mathbf{a}_i que representa a la acción $\alpha_j \in Act$.

La Figura 6.11 muestra los AAF de Alice y Bob. Cada argumento está rodeado por un rectángulo, y una flecha de un argumento a otro representa el ataque del primero al segundo. Las etiquetas de los argumentos prácticos están subrayadas, mientras que las etiquetas de los argumentos epistémicos privados están escritas en negrita y finalizan con un signo de pregunta. El resto de las etiquetas identifican a los argumentos epistémicos revelables.

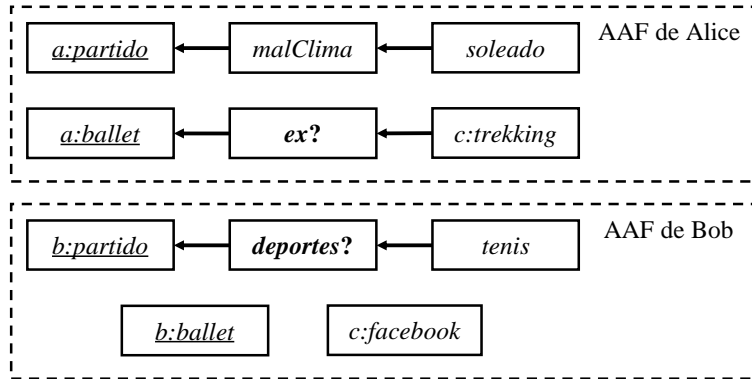


Figura 6.11: Marcos de argumentación de Alice y Bob presentados en el ejemplo de aplicación de [GTWX16].

Note que tanto Alice como Bob consideran todas las alternativas, esto es representado por los argumentos prácticos *a:partido*, *a:ballet*, *b:partido* y *b:ballet*. Alice cree que si hay mal clima (*malClima*), no es buena idea ir al partido, aunque en el noticiero dijeron que iba a estar soleado (*soleado*). Por otro lado, Alice sabe que la ex de Bob suele ir al ballet y prefiere evitar cruzarse con ella (*ex?*). Alice no quiere que Bob se entere de su preocupación por lo que esta información es privada. Más allá de eso, Carol (la hija de Bob con su ex pareja) comentó que tal vez irían a hacer trekking con su madre (*c:trekking*), por lo que Alice asume que ella no va a asistir al ballet. Por su parte, Bob quiere ir a ver el partido, pero no recuerda si a Alice le gustan los deportes (*deportes?*), y tampoco quiere preguntárselo, aunque sí recuerda que ella disfrutó ir a ver un partido de tenis (*tenis*). Además, Bob vió en el Facebook de Carol que ella y su madre tenían entradas para el ballet (*c:facebook*). Observe que todos los argumentos prácticos son admisibles de acuerdo con los AAF de cada agente. Sin embargo, si Bob compartiría con Alice el argumento *c:facebook*, este atacaría al argumento *c:trekking* y entonces *a:ballet* dejaría de ser admisible (ya que no existe un argumento que ataque a *c:facebook*), por lo que ir al ballet dejaría de ser una opción factible para Alice.

Antes de continuar, hace falta introducir algunas nociones. En primer lugar la noción de conjunto relativo-admisible. Dado un AAF $\langle Args, Atts \rangle$, un conjunto de argumentos $E \subseteq Args$ es relativo-admisible si $\exists \mathcal{A} \in E$, denominado un tópico de E , tal que cada argumento $\mathcal{B} \in E$ defiende a \mathcal{A} . En este caso, se entiende que un argumento \mathcal{B} defiende a otro \mathcal{A} si y sólo si:

- $\mathcal{A} = \mathcal{B}$; o
- $\exists \mathcal{B}' \in \text{Args}$ tal que \mathcal{B}' ataca a \mathcal{A} y \mathcal{B} ataca a \mathcal{B}' ; o
- $\exists \mathcal{B}' \in \text{Args}$ tal que \mathcal{B}' defiende a \mathcal{A} y \mathcal{B} defiende a \mathcal{B}' .

Un conjunto relativo-admisibile E es una explicación compacta para \mathcal{A} si y sólo si \mathcal{A} es un tópico de E y E es el conjunto relativo-admisibile más pequeño (con respecto a \subseteq) entre todos los conjuntos relativo-admisibles cuyos tópicos incluyan a \mathcal{A} . Por las restricciones que existen para los ataques entre los argumentos de los AAF de los agentes, se puede asegurar que si E es una explicación compacta para un argumento práctico \mathcal{A}_{ij} , entonces \mathcal{A}_{ij} es el único tópico de E , y se dice que $R = E \setminus \{\mathcal{A}_{ij}\}$ es una razón compacta del agente \mathbf{a}_i para la acción α_j . Considerando el AAF de Alice, se tendría que $\{\text{soleado}\}$ y $\{c:\text{trekking}\}$ son las razones compactas de Alice para “ir al partido” y para “ir al ballet”, respectivamente.

En el protocolo de cooperación presentado por Gao et al., cada agente es responsable de asignarse la acción que va a realizar. Para esto, si por ejemplo un agente \mathbf{a}_i desea asignarse la acción α_j , (antes de hacerlo) este agente deberá verificar que α_j sea factible con respecto a su conocimiento, y al conocimiento público de los demás agentes. Esto se logra, mediante un intercambio de argumentos entre los agentes orquestado por medio de una variante de las “two-parties immediate-response disputes” (o TPI-disputes) [DBC03], a la que los autores denominaron TPI-dispute de razón compacta. Sea R una razón compacta de un agente \mathbf{a}_i para realizar una acción α_j , la TPI-dispute de R es una discusión entre dos agentes (o jugadores), denominados el defensor y el desafiante, que se lleva a cabo para verificar la admisibilidad del conjunto de argumentos revelables $rR \subset R$ con respecto al conocimiento de todos los agentes. Tanto el defensor como el desafiante, son agentes diferentes al resto de los agentes del DisCSP ya que sólo actuarán como intermediarios que preguntarán a los otros agentes acerca de sus argumentos revelables, para poder presentarlos en la discusión. El objetivo del desafiante es presentar argumentos que ataquen a los argumentos en rR , y el objetivo del defensor es presentar argumentos que los defiendan a los argumentos en rR de estos ataques. En el primer paso el defensor presenta todos los argumentos en rR . Si $rR = \emptyset$, la discusión termina y el defensor gana. Luego, de esto (si $rR \neq \emptyset$) ambos jugadores toman turnos para preguntarle a todos los agentes menos a \mathbf{a}_i si poseen algún argumento revelable que ataque a alguno de los argumentos presentados previamente por su contrincante; en caso de que el jugador del turno actual encuentre un argumento, este lo presenta en la discusión siempre y cuando respete las reglas que se enuncian a continuación:

- Ambos jugadores pueden hacer “backtracking”, es decir, presentar un argumento que ataque a otro argumento que fue presentado anteriormente por su contrincante en un turno anterior al previo.
- Ambos jugadores pueden repetir argumentos presentados previamente por el defensor.
- El desafiante tiene permitido repetir argumentos presentados previamente por el mismo, si y sólo si, el argumento que quiere presentar se encuentra en una línea de disputa diferente a la línea en la que quiere introducirlo actualmente. Una línea de disputa es una secuencia de argumentos (presentados en la discusión) en la que cada uno, a excepción del primero, ataca a su predecesor.
- Ningún otro movimiento es válido.

El desafiante gana la discusión si presenta un argumento presentado previamente por el defensor, o si el defensor se queda sin movimientos; mientras que el defensor gana si el desafiante no puede ganar. Si gana el jugador defensor, entonces los argumentos en rR son considerados admisibles con respecto al conocimiento de \mathbf{a}_i y el conocimiento público de todos los demás agentes, lo que implica que la acción α_j es factible para \mathbf{a}_i .

Siguiendo con el ejemplo, para verificar si es factible para Alice ir al ballet, se lleva a cabo la TPI-Dispute de $\{c:trekking\}$. Como $c:trekking$ es revelable, el primer paso del defensor es presentar el conjunto de argumentos $\{c:trekking\}$. Luego, el desafiante presenta el argumento $c:facebook$ de Bob ya que ambos agentes están de acuerdo en que $c:facebook$ derrota a $c:trekking$. Como no existe un derrotador para $c:facebook$, el defensor se queda sin movimientos, y por lo tanto, pierde la disputa. De esta manera, Alice puede comprobar que “ir al ballet” no es factible para ella. En cambio, al llevar cabo la TPI-dispute de $\{soleado\}$ (para corroborar si “ir al partido” es factible), dado que $soleado$ no tiene derrotadores, el desafiante se quedaría sin movimientos en su primer turno, por lo que el defensor ganaría la disputa. Cabe aclarar que para que la propuesta de Gao et al. funcione, todos los agentes deben estar de acuerdo en la relación de derrota que existe entre los argumentos revelables. Sin embargo, para que esto sea así, los agentes deberían compartir una relación de ataque global (y un conjunto de argumentos globales), o especificar alguna otra forma para identificar las derrotas. Como por ejemplo, a través de una relación de conflicto y una función de preferencia, como en el caso de los DAF ([RMGS10]), que como fue explicado en la Sección 6.2.3, los mismos cuentan con una

relación de conflicto \bowtie , y una función de preferencia \mathbf{pref} para determinar las derrotas entre los argumentos.

Protocolo de comunicación (perspectiva del agente \mathbf{a}_i)	Check(acción α , estrategia S , agente \mathbf{a}_i)
<pre> Inicializar S como una estrategia vacía si $i = 1$ $mi_a \leftarrow null$ if tengo el token then for $\alpha \leftarrow \text{GetNextBestAction}(mi_a)$ do if $\text{Check}(\alpha, S, \mathbf{a}_i) == [\text{consistente}, \text{factible}]$ then actualizar S y mi_a if $i < n$, then enviar $\mathbf{ok?}(S)$ y el token a \mathbf{a}_{i+1} else return S end if end for end if if $i > 1$ then enviar \mathbf{ngd} y el token a \mathbf{a}_{i-1} else return $null$ end if end if </pre>	<pre> $status \leftarrow [no_consistente, no_factible]$ if asignar α a \mathbf{a}_i es consistente con S then $status[0] \leftarrow consistente$ end if obtener todas las razones compactas E_1, \dots, E_K para α for each E_k, con $k = 1, \dots, K$ do if la discusión sugiere que E_k tiene éxito then $status[1] \leftarrow factible$ end if end for return $status$ </pre>

Figura 6.12: Protocolo de cooperación presentado en [GTWX16].

La Figura 6.12 muestra el algoritmo que implementa el protocolo de cooperación antes mencionado. El protocolo se activa cuando al agente \mathbf{a}_1 le llega el mensaje $\mathbf{ok?}(S)$ junto con el token. En el primer paso, \mathbf{a}_1 inicializa la estrategia S como vacía (es decir, ningún agente tiene asignada ninguna acción). En la variable mi_a los agentes almacenarán la acción que se asignen, esta es inicializada como $null$ para representar que en principio los agentes no tienen asignada ninguna acción. Cuando el agente recibe el token, este itera sobre todas las acciones que son igual o menos preferidas que su acción actual mi_a (recuerde que cada agente \mathbf{a}_i cuenta con la relación de preferencia \geq_i sobre el conjunto de acciones). Esto lo consigue invocando la función GetNextBesAction en cada paso de la iteración. Cabe aclarar que cuando la entrada de GetNextBesAction es $null$ esta función retorna la acción más preferida por el agente, o una de ella en caso de ser varias. Por cada acción α obtenida de esta manera, el agente chequea si es consistente con respecto a S y si es factible. Esto lo hace por medio de la función Check . Como se puede ver en la Figura 6.12, el primer paso de esta función es inicializar la variable $status$ con una tupla de dos elementos mediante los cuales se indica el resultado del chequeo. Luego, se verifica la consistencia de la asignación en cuestión, y en caso de ser consistente, se asigna a la

primer componente de *status* el valor *consistente*. Si \mathbf{a}_i no es el último agente (es decir, $i < n$), asignar α a \mathbf{a}_i es consistente con S si $N(\alpha, S) < \maxCon(\alpha)$ (la cantidad de agentes asignados a α no supera la restricción máxima), en caso contrario ($i = n$), esta asignación es consistente si luego de agregarla a S se obtiene una estrategia aceptable. A continuación, \mathbf{a}_i obtiene todas las razones compactas para α , y por cada una de ellas se efectúa una TPI-dispute de razón compacta. Si para alguna de estas la discusión resulta exitosa (es decir, el defensor gana), entonces se asigna a la segunda componente de *status* el valor *factible*. Si el chequeo de la acción actual α determina que esta es consistente y factible, se actualiza S y mi_α para que quede reflejado que \mathbf{a}_i se asignó la acción α . Luego, si \mathbf{a}_i no es el último agente ($i < n$), este envía el mensaje **ok?**(S) y el token al agente \mathbf{a}_{i+1} . Si \mathbf{a}_i es el último agente ($i = n$), entonces S es una estrategia aceptable y el agente la retorna como solución al problema. Si tras haber analizado todas las acciones, \mathbf{a}_i no encuentra una que sea consistente y factible, este debe activar el proceso de backtracking enviando el mensaje **ngd** (por “not good”) al agente \mathbf{a}_{i-1} . Todos los agentes pueden hacer backtracking, excepto por \mathbf{a}_1 (el primer agente). Si el primer agente no puede encontrar una acción consistente y factible, entonces el protocolo falla y se retorna *null*.

Considerando que $\mathbf{a}_1 = \text{Alice}$, con “ir al ballet” $>_i$ “ir al partido”, y que $\mathbf{a}_2 = \text{Bob}$, con “ir al partido” $>_i$ “ir al ballet”. Al ejecutar el protocolo, en primer lugar Alice trataría de verificar si “ir al ballet” es consistente y factible, pero fallaría ya que como se explicó antes la TPI-dispute de $\{c: \text{facebook}\}$ no tendría éxito, por lo que “ir al ballet” no es factible. Luego, al verificar lo mismo, pero para “ir al partido”, obtendría que esta acción es consistente y factible. A continuación, Alice actualiza S asignándose la acción “ir al partido”, y envía el mensaje **ok?**(S) a Bob. Al recibir este mensaje Bob verifica si “ir al partido” es consistente y factible, lo cual es cierto. Por lo tanto, Bob actualiza S asignándose la acción “ir al partido”, y finaliza el protocolo retornando S , mediante la cual se informa que tanto Alice como Bob irán a ver el partido.

Observe que el orden asignado a los agentes dentro del protocolo puede afectar el resultado obtenido. Además, cuanto más adelante esté un agente, más prioridad tendrán sus preferencias. Por ejemplo, suponga que existen dos agentes (\mathbf{a}_1 y \mathbf{a}_2) y dos acciones (α_1 y α_2) tales que $(\alpha_1 >_1 \alpha_2)$, $(\alpha_1 >_2 \alpha_2)$, y $Con(\alpha_1) = Con(\alpha_2) = \{1\}$. Luego, (asumiendo que ambas acciones son factibles para ambos agentes) \mathbf{a}_1 se asignaría α_1 (su acción preferida), pero \mathbf{a}_2 tendría que asignarse α_2 (su acción menos preferida) ya que de otra manera su elección no sería consistente. Mas aún, si bien estos autores demuestran

que su protocolo es sensato, también admiten que no es completo. Es decir, siempre que el protocolo retorne una estrategia, esta será una estrategia aceptable, aunque puede darse el caso que el protocolo retorne *null*, cuando en realidad existe una estrategia aceptable que resuelve el problema.

En comparación con la propuesta presentada en esta tesis, el trabajo de Gao et al. fue concebida teniendo en cuenta un mayor nivel abstracción, lo que le permite omitir detalles que a la hora de implementar o utilizar el formalismo pueden resultar en problemas no triviales de resolver. Por ejemplo, en cuanto a los argumentos, no siempre puede ser sencillo identificar las ideas del mundo real para traducirlas en el sistema como argumentos abstractos, y menos aún, identificar los conflictos y/o ataques que pueden existir entre ellos. En este sentido, poseer una sintaxis para construir reglas que estructuren la forma en la que los agentes desean comparar las alternativas, como es el caso de las *DePref*-rules, facilita la tarea de la/s persona/s que configure/n el sistema. Relacionado con esto, en [GTWX16], las alternativas tampoco poseen descripción o atributos con las que puedan ser comparadas, y en consecuencia, los agentes deben establecer de forma directa un orden entre las alternativas, lo cual no es una tarea nada trivial. De hecho se podría utilizar la propuesta planteada en esta tesis para resolver ese problema a nivel de un sólo agente.

6.4. Resumen

En este capítulo se presentaron, de forma resumida, una serie de trabajos relacionados que proponen el uso de argumentación para el desarrollo de sistemas de toma de decisiones. Para cada uno de los trabajos presentados en este capítulo, se explicó la forma en la que abordan la problemática, y se realizó un análisis de sus debilidades, en contraste con la propuesta presentada en esta tesis. En particular, para cada trabajo se analizó la forma en la que se representa la información del problema y las preferencias de los agentes, qué tipo de modelo argumentativo utilizan, y (si lo hacen) de qué manera brindan explicaciones a los usuarios del sistema. Cabe mencionar, que si bien se distinguió entre aquellos trabajos en los que la toma de decisiones es individual, y aquellos en los que intervienen múltiples agentes, la propuesta presentada en esta tesis está relacionada con todos ya que puede ser aplicada en dominios de ambos tipos.

	Individual			Multi-agente		
	[HW12]	[ZFLT14] [ZFLT19]	[FTG+17]	[FTMW14]	[GTWX16]	TESIS
Representación de información	Tabla de evidencia	Atributos y metas	Conjunto de evidencia	Atributos y metas	Grafo de argumentos	Conjunto de evidencia
Atributos con múltiples valores	Sí	No	Sí*	No	-	Sí
Representación de preferencias	Orden entre beneficios	Metas	Orden entre criterios	Metas	Orden entre acciones	<i>DePref</i> -rules y órdenes de importancia
Prioridad entre agentes	-	-	-	Orden parcial*	Orden de participación	Orden parcial
Sistema argumentativo	AAF*	ABAF	DAF	ABAF	AAF	MEPM
Evaluación de argumentos	Semánticas de Dung	Árbol de disputa	Árbol de dialéctica	Diálogo-ABA	TPI-Dispute	Árbol de dialéctica
Tipo de explicaciones	Grafo de superioridad	Lenguaje natural	-	-	-	Grafo explicativo
Explicaciones interactivas	No	No	No	No	No	Sí
Implementación	Sí	Sí	No	No	No	Sí
Link de descarga	No	No	No	No	No	Sí

Tabla 6.4: Resumen de las características de cada trabajo relacionado y el trabajo propuesto en esta tesis.

La Tabla 6.4 resume las características de los formalismos presentados en cada uno de los trabajos analizado previamente, y del propuesto en esta tesis. Con respecto a la representación de información, el único de estos trabajos que no caracteriza las alternativas en base a sus atributos es el de [GTWX16]. El formalismo que proponen estos autores posee un enfoque más abstracto, por lo que la información que tienen los agentes acerca del problema de decisión se representa mediante un argumentos dispuestos dentro de un marco de argumentación abstracto, o abstract argumentation frameowrk (AAF). Si bien es cierto que un mayor nivel de abstracción suele brindar mayor flexibilidad, también es cierto que aumenta la complejidad a la hora de modelar los problemas por no poseer una estructura bien definida de cómo deben ser los argumentos, o de cómo debe ser la relación de conflicto y/o derrota entre ellos. El resto de los trabajos, de alguna u otra forma, requieren que la información esté representada de forma estructurada, ya sea por medio de tablas, como es el caso de [HW12, ZFLT14, FTMW14], o bien por medio de predicados (o conjuntos de evidencia), como en [FTG⁺17], o en el marco de trabajo propuesto en esta tesis. En relación con esto, los únicos trabajos que genuinamente admiten rangos de valores con respecto a los criterios mediante los cuales se comparan las alternativas son [HW12] y el presentado en esta tesis. No obstante, en [FTG⁺17] también existe un rango de valores, aunque es implícito, ya que en esta propuesta no se consideran valores sino que se indica si una alternativa es mejor que otra (o si son exactamente iguales) con respecto a cada uno de los criterios. En [ZFLT14] y en [FTMW14], la relación entre alternativas, atributos y metas es cualitativa, es decir, no es gradual. Una alternativa puede (o no) poseer un atributo, y una meta puede (o no) ser satisfecha por un atributo.

En cuanto a la representación de preferencias, entre los trabajos analizados existen dos grandes grupos: aquellos que usan alguna clase de relación de orden ([HW12, FTG⁺17, GTWX16]), y aquellos basados en metas ([ZFLT14, FTMW14]). La ventaja de los enfoques basados en un orden es que las comparaciones se hacen de manera sencilla, el problema está en que si los elementos a comparar son muchos, definir este orden puede ser una tarea compleja como podría pasar en el caso de [HW12], o de [GTWX16]. Además, los ordenes no permiten balancear el peso o la importancia que se le da a cada elemento, solo es posible indicar cuándo uno es más importante que otro como ocurre con [FTG⁺17]. Por otro lado, los enfoque basados en metas pueden ser más flexibles en este aspecto, ya que al poder indicar varias metas, o que una meta sea satisfecha por varios atributos, se puede balancear la importancia que se le da estos últimos. Sin embargo, en el caso de [ZFLT14] y [FTMW14], se necesita definir un forma de comparar las alternativas

en base a las metas que estas satisfacen, por lo que los autores deben recaer en uso de nociones como la de dominancia (que puede dejar muchas alternativas sin comparar) o la de desviación mínima (que puede restringir demasiado los dominios de aplicación). La propuesta presentada en esta tesis, podría pensarse entre medio de estos dos enfoques. Por un lado, mediante el uso de *DePref*-rules los agentes pueden, con cierto nivel de flexibilidad, definir la forma de comparar las alternativas, y cuánto peso tiene cada criterio en esta comparación (cuánto los agentes están dispuestos a sacrificar de un dado criterio en pos de obtener un beneficio con respecto a otro), lo que debería estar alineado con las metas o intereses de los agentes, y en donde recae la mayor dificultad del proceso. Además, cada agente puede definir su propio orden entre reglas para determinar cuánta importancia asignarle a cada una. Suponiendo que la cantidad de reglas no es muy grande, o en general, menor al número de alternativas (y al de argumentos), definir los órdenes de importancia no debería ser difícil, y al poseer un orden, la tarea de comparar las reglas (y finalmente los argumentos) se hace sencilla.

Contemplando únicamente los trabajos que consideran múltiples agentes, otro aspecto a comparar es si consideran algún tipo de prioridad entre los agentes. En el caso de [FTMW14], lo primero que hay que tener en cuenta es que sólo puede aplicarse con dos agentes, no así el trabajo de [GTWX16]. Además, la propuesta original de Fan et al. asume que uno de los agentes debe ser más confiable que otro, otorgándole prioridad al conocimiento de este mismo, lo que implicaría el uso de un orden (estricto) total para definir la prioridad entre los agentes. Sin embargo, esta limitación es atacada por los autores, definiendo una segunda propuesta para tratar los casos en los que los agentes tienen el mismo nivel de confiabilidad, por lo que se puede decir este trabajo considera un orden de prioridad parcial. En cuanto a [GTWX16], el orden de prioridad está dado por el orden (total) en el que los agentes participan durante la ejecución del protocolo de comunicación. Los agentes que van primero, tienen más acciones disponibles, y por ende, es más probable que puedan elegir la acción que más prefieran. Cabe recordar, que además de determinar la prioridad de los agentes, este orden también afecta el resultado del proceso, ya que dependiendo del mismo (y de las preferencias de los agentes) el protocolo podría terminar sin llegar a una solución, a pesar de que esta exista. Es decir, el protocolo no es completo. Por su parte, el marco de trabajo propuesto en esta tesis considera cualquier número de agentes, y el orden de prioridad entre ellos puede ser parcial.

Como se dijo en principio, todos los formalismos analizados utilizan alguna clase de

sistema argumentativo para cumplir con su tarea. En [HW12], se generan argumentos a partir de la evidencia y un conjunto de reglas y se los dispone en un abstract argumentation framework (AAF). La relación de derrota entre estos argumentos queda definida a partir de sus conclusiones y la relación de preferencia. Luego, los argumentos son evaluados siguiendo la semántica grounded o preferida [Dun95]. En [GTWX16], también se utiliza argumentación abstracta, pero en este caso no hay reglas para definir los argumentos ni las derrotas entre los mismos, y los argumentos son evaluados mediante TPI-Disputes. Tanto en [ZFLT14], como en [FTMW14] se utiliza assumption-based argumentation framework (o marco ABA) para resolver el problema. En ambos trabajos los marcos son definidos en base a un marco de decisión de Fan [FT11], con el que se modela la información que se tiene del problema. En cuanto al proceso de evaluación de argumentos, en [ZFLT14] se utilizan árboles de disputa, mientras que en [FTMW14] se utilizan Diálogos-ABA [FT11]. Por último, en [FTG⁺17] se utiliza un marco de argumentación dinámico (o DAF, por sus siglas en inglés). Los autores de este trabajo definen claramente el conjunto de argumentos que debe ser generado, los conflictos, y la relación de preferencia entre argumentos, permitiendo que el DAF se genere automáticamente una vez definidos el conjunto de alternativas y criterios. Dentro de los trabajos analizados, esta es la única propuesta que, al igual que el marco de trabajo propuesto en los capítulos anteriores, utiliza árboles de dialéctica para evaluar los argumentos. Sin embargo, la propuesta presentada en esta tesis utiliza un marco argumentativo, denominado marco de evaluación de preferencias multi-agente (MEPM), el cual fue diseñado para poder trabajar con los argumentos generados a partir de las *DePref*-rules, y los ordenes de importancia de varios agente con distintas prioridades.

Sólo en [HW12] y en [ZFLT14], además de en esta tesis, se describe de qué forma brindar explicaciones de los resultados obtenidos. Llamativamente, también son sólo estos trabajos los que documentan algún tipo de implementación de su propuesta. En [HW12] se propone el uso de los grafos de superioridad para mostrar el orden de las alternativas obtenido. En estos grafos, los nodos representan las alternativas, y la relación de orden está representada por medio de flechas entre ellos. La relación de orden se explica a partir del grafo de argumentos y las extensiones que se obtienen siguiendo la semántica de Dung seleccionada. Por otro lado, en [ZFLT14], los autores proponen brindar explicaciones en lenguaje natural. Estas explicaciones se obtienen a partir de un algoritmo que toma como entrada un árbol de disputa y va generando las explicaciones a partir de los argumentos dentro del mismo, teniendo en cuenta sus conclusiones y premisas. En esta tesis se

definieron los grafos explicativos como herramienta visual e interactiva para brindar explicaciones. Estos grafos, inspirados en los grafos de superioridad de [HW12], permiten observar el orden entre las alternativas. Sin embargo, a los grafos de superioridad brindan información acerca de cuáles fueron las reglas que fueron usadas para llegar a las conclusiones que construyen el orden. Además, los grafos explicativos dan acceso a los grafos de argumentos y a las explicaciones dialécticas, que en conjunto justifican por qué un argumento es aceptado o rechazado.

Capítulo 7

Conclusiones

En esta tesis se presentó un marco de trabajo para toma de decisiones individuales y colectivas en sistemas multi-agentes el cual hace uso de un sistema argumentativo. Este marco de trabajo permite modelar diferentes escenarios en los que un grupo de uno o varios agentes debe tomar una decisión para resolver un problema en común teniendo en cuenta múltiples criterios de evaluación.

En el contexto de este marco de trabajo, las preferencias de los agentes son representadas mediante un tipo particular de reglas, especialmente diseñadas para esta propuesta, denominadas reglas de preferencia rebatibles (o *DePref*-rules), cuya sintaxis y semántica fueron presentadas en el Capítulo 2. Una *DePref*-rule es una regla esquemática en término de dos variables X e Y , que indica las condiciones que deben satisfacerse sobre las valoraciones de dos alternativas cualquiera (X e Y) para tener una razón para creer que una de ellas es preferida a la otra. Algo que hace a las *DePref*-rules diferentes a otros enfoques de representación de preferencias, es su capacidad para indicar de forma explícita cómo se compensan las virtudes y las falencias de las alternativa. Estas reglas pueden ser utilizadas para representar preferencias muy simples como: “Si una alternativa X es mejor que una alternativa Y en cuanto a la calidad de servicio, entonces existe una razón para preferir a X sobre Y ”; o también más complejas como: “Si X ofrece al menos 5 días más de paseo que los que ofrece Y , y el servicio que ofrece X es excelente y mejor que el de Y , a pesar de que Y ofrezca un mejor entretenimiento, mientras que el entretenimiento de X no sea pésimo (es decir, sea al menos malo) y mientras que el de Y no sea excelente (es decir, sea a lo sumo bueno), entonces existe una razón para preferir a X sobre Y ”.

Al instanciar las *DePref*-rules con alternativas concretas, es posible inferir una preferencia entre dos alternativas específicas, siempre y cuando, la información acerca de las alternativas, denominada evidencia, respalde el cumplimiento de todas las condiciones establecidas en el cuerpo de las *DePref*-rules. En el Capítulo 2 se presentaron propiedades que permiten caracterizar las preferencias que pueden ser inferidas a partir de las *DePref*-rules. Como se mostró a través de los distintos ejemplos de aplicación introducidos a lo largo de este documento, las preferencias inferidas a partir de un conjunto de *DePref*-rules pueden ser contradictorias. Esto es, se puede llegar a la conclusión de que hay razones para creer que una alternativa d_x es preferida a otra alternativa d_y ($pref(d_x, d_y)$), pero también podrían existir otras razones diferentes para creer que d_y es preferida d_x ($pref(d_y, d_x)$). Para decidir entre conclusiones contradictorias, se utilizó un marco argumentativo mediante el cual es posible determinar cuáles de estas preferencias deben ser consideradas, y cuáles deben ser descartadas. Tal y como se describió por medio de un algoritmo, los argumentos de este sistema son construidos instanciando las *DePref*-rules con cada uno de los posibles pares de alternativas, y evaluando si la evidencia satisface las condiciones establecidas en el cuerpo de estas reglas.

Dado que nuestro enfoque se basa en un sistema multi-agente colaborativo, se asume que cada uno de los agentes debe aceptar las razones de los demás para considerar que una alternativa es preferida a otra. Sin embargo, el formalismo propuesto, permite modelar situaciones en las que no todos los agentes comparten los mismos intereses, y por lo tanto, pueden considerar que una *DePref*-rule sea más importante que otra, así como también, es posible modelar situaciones en las que los intereses de ciertos agentes tengan prioridad sobre los intereses de otros. Esto es modelado por medio de una lista, que contiene un orden parcial estricto por cada uno de los agentes que define (según ese agente) cuándo una *DePref*-rule es más importante que otra, y por medio de un orden parcial estricto que determina la prioridad entre los agentes. En el Capítulo 3, fue explicado cómo estos elementos son considerados por el marco de trabajo argumentativo a la hora de determinar cuándo un argumento es más fuerte que otro, y en base a esto, determinar cuándo un argumento derrota a otro. Una vez identificadas las derrotas entre los argumentos, se propuso un algoritmo de evaluación dialéctico, mediante el cual se determinan los argumentos que deben ser aceptados. Este proceso, construye el árbol de dialéctica de cada uno de los argumentos, y marca sus nodos siguiendo el procedimiento de marcado **m**. En la Sección 3.3 se presentó un teorema y un corolario que aseguran el correcto funcionamiento del sistema argumentativo.

En el Capítulo 4, se indicó la forma en la que el marco de trabajo utiliza las conclusiones de los argumentos aceptados, es decir, las conclusiones garantizadas, para construir un orden de preferencia entre las alternativas. Luego, se explicó cómo este orden es utilizado por el proceso de selección de alternativas (implementado por un algoritmo) para determinar cuáles son las mejores alternativas. El comportamiento del proceso de selección fue formalmente caracterizado.

Además, como fue comentado en la Sección 4.3, la explicación de los resultados es tan importante como los resultados en sí. Dada la cantidad de elementos que podrían estar involucradas en una decisión, disponer de una explicación que sustente el resultado obtenido por el sistema de decisión propuesto, permite que se comprendan con sumo detalle: cuáles son los elementos que fueron tenidos en cuenta, de qué forma interactúan y por qué algunos fueron considerados mejores que otros. Esto permite brindar una respuesta informada de la recomendación otorgada a los tomadores de decisiones. Además, esto fortalece la confianza de los usuarios en el sistema, y permite cumplir con las regulaciones de aquellos países en donde ya es obligatorio que los sistemas capaces de ser utilizados para automatizar el proceso de toma de decisiones brinden explicaciones acerca de su funcionamiento. Para esta propuesta, se diseñaron explicaciones interactivas que se presentan por medio de un grafo, llamado grafo explicativo, mediante el cual se puede visualizar el orden de preferencia entre las alternativas. Cada uno de los nodos de este grafo representa una alternativa, y los arcos del grafo representan la relación de preferencia entre las alternativas. Además, cada arco de un grafo explicativo está asociado a tres elementos: i) el conjunto de *DePref*-rules con las que se construyeron los argumentos aceptados cuya conclusión sustenta la preferencia que este arco representa, ii) la explicación dialéctica que justifica la garantía de dicha conclusión, y iii) un grafo de argumentos en el que se muestra la relación de conflicto y derrota entre los argumentos que aparecen en la explicación dialéctica. Todos estos elementos, permiten que los usuarios puedan obtener una explicación exhaustiva acerca de por qué una alternativa es seleccionada, por qué una alternativa es preferida a otra, o cómo fue construido el orden de preferencia. Además, las explicaciones les pueden servir a los usuarios para detectar errores que hayan ocurrido al establecer los datos del problema, o al especificar las *DePref*-rules.

El marco de trabajo presentado en esta tesis es capaz de funcionar correctamente aún cuando la información que se tenga de las alternativas no sea completa, es decir, cuando existan alternativas para las que no se conozca su valoración con respecto a alguno de los

criterios. Esto fue explicado en la Sección 4.5. Por otro lado, en la Sección 4.6, se estudió formalmente cómo afectan los cambios en la evidencia o en el conjunto de *DePref*-rules al conjunto de argumentos generados. A partir de este estudio, se obtiene que no hace falta re-computar por completo el conjunto de argumentos, ante un cambio de esta índole ya que es posible identificar cuáles son los argumentos que deben ser eliminados, y además, cuáles son las *DePref*-rules y alternativas que deben ser evaluadas para generar los nuevos argumentos.

El marco de trabajo fue implementado, y la aplicación resultante fue detallada en el Capítulo 5. A lo largo de dicho capítulo, se explicó como hacer uso de la interfaz gráfica para sacar el máximo provecho a las explicaciones interactivas. Esta aplicación permitió poner a prueba el formalismo y experimentar con el mismo. Al final del capítulo, se muestran los resultados obtenidos a partir de una serie de experimentos que se llevaron a cabo con la aplicación con el objetivo de tener una estimación del desempeño de la aplicación a medida que aumenta el número de alternativas y de *DePref*-rules.

Finalmente, en el Capítulo 6, se realizó una descripción y análisis de la literatura relacionada, que al igual que este, utilizan argumentación para construir un sistema de toma de decisiones, y se los comparó con el mismo. En primer lugar, fueron analizados sistemas que consideran un sólo agente, y luego, otros diseñados para trabajar con varios agentes. La forma en la que estos trabajos representan el conocimiento, o más precisamente, las alternativas y sus características es diferente entre sí, como así también la forma de representar las preferencias. La principal diferencia del trabajo presentado en esta tesis, con los otros radica en que este es el único que aborda explícitamente nociones como requisitos (valores mínimos) y tolerancia (valores máximos) a la hora de comparar las alternativas, lo que le otorga a los usuarios mayor control y flexibilidad sobre este aspecto. Otra de las ventajas de este trabajo, frente a varios de los otros es que cuenta con una implementación, la cual permite a los usuarios interactuar con los grafos explicativos, siendo este tipo de explicaciones también algo propio de esta propuesta.

Esta tesis comprende el trabajo de investigación realizado en los últimos cinco años. Durante este periodo, han surgido algunas ideas y preguntas que aún no han podido ser desarrolladas o estudiadas. Debido a esto, se dejan planteadas algunas líneas de trabajo a futuro.

Uno de los primeros pasos que se considera relevante para continuar, es estudiar la manera de modificar el formalismo propuesto para contemplar casos en los que cada pieza

de evidencia tenga asociada cierta incertidumbre, pudiendo el conjunto de evidencia ser contradictorio. Otro de los posibles pasos de interés para continuar, radica en estudiar la manera de modificar la sintaxis y semántica de las *DePref*-rules para considerar explícitamente los casos en los que no hay información acerca de los valores de una alternativa con respecto a algún criterio, y así poder comparar aún mejor las alternativas en estos casos. Por último, otro aspecto interesante de explorar, es la factibilidad de utilizar algún mecanismo de aprendizaje para automatizar o simplificar el modelado de preferencias a través de las *DePref*-rules.

Bibliografía

- [ABCM06] ATKINSON, K., BENCH-CAPON, T., AND MODGIL, S. Argumentation for decision support. In *Database and Expert Systems Applications* (2006), S. Bressan, J. Küng, and R. Wagner, Eds., Springer Berlin Heidelberg, pp. 822–831.
- [ABP05] AMGOUD, L., BONNEFON, J.-F., AND PRADE, H. An argumentation-based approach to multiple criteria decision. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty* (2005), Springer Berlin Heidelberg, pp. 269–280.
- [AC02] AMGOUD, L., AND CAYROL, C. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence* 34, 1?3 (2002), 197?215.
- [AGG19] AGIS, R. A., GOTTIFREDI, S., AND GARCÍA, A. J. An approach for distributed discussion and collaborative knowledge sharing: Theoretical and empirical analysis. *Expert Systems with Applications* 116 (2019), 377–395.
- [Amg09] AMGOUD, L. *Argumentation for Decision Making*. Springer US, 2009, pp. 301–320.
- [AP09] AMGOUD, L., AND PRADE, H. Using arguments for making and explaining decisions. *Artificial Intelligence* 173, 3 (2009), 413–436.
- [BBD⁺14] BRIGUEZ, C. E., BUDÁN, M. C., DEAGUSTINI, C., MAGUITMAN, A. G., CAPOBIANCO, M., AND SIMARI, G. R. Argument-based mixed recommenders and their application to movie suggestion. *Expert Systems with Applications* 41, 14 (2014), 6467 – 6482.

- [BBHP99] BOUTILIER, C., BRAFMAN, R. I., HOOS, H. H., AND POOLE, D. Reasoning with conditional ceteris paribus preference statements. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence (1999)*, UAI'99, Morgan Kaufmann Publishers Inc., pp. 71–80.
- [BBTG19] BURON BRARDA, M. E., TAMARGO, L. H., AND GARCÍA, A. J. An approach to enhance argument-based multi-criteria decision systems with conditional preferences and explainable answers. *Expert Systems with Applications* 126 (2019), 171–186.
- [BBTG20] BURON BRARDA, M. E., TAMARGO, L. H., AND GARCÍA, A. J. Using argumentation to obtain and explain results in a decision support system. *IEEE Intelligent Systems* (2020).
- [BBTG23] BURON BRARDA, M. E., TAMARGO, L. H., AND GARCÍA, A. J. Multi-agent multi-criteria decision framework with interactive explanation tools. *Expert Systems with Applications* 228 (2023), 120360.
- [BCG11] BARONI, P., CAMINADA, M., AND GIACOMIN, M. An introduction to argumentation semantics. *Knowledge Engineering Review* 26 (2011), 365–410.
- [BCHS18] BISTARELLI, S., CEBERIO, M., HENDERSON, J. A., AND SANTINI, F. Abstract argumentation frameworks to promote fairness and rationality in multi-experts multi-criteria decision making. In *Constraint Programming and Decision Making: Theory and Applications (2018)*, Springer International Publishing, pp. 7–19.
- [BCKK19] BISQUERT, P., CROITORU, M., KAKLAMANIS, C., AND KARANIKOLAS, N. A decision-making approach where argumentation added value tackles social choice deficiencies. *Progress in Artificial Intelligence* 8, 2 (Jun 2019), 229–239.
- [BD02] BRAFMAN, R. I., AND DOMSHLAK, C. Introducing variable importance tradeoffs into cp-nets. In *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence (2002)*, UAI'02, Morgan Kaufmann Publishers Inc., pp. 69–76.

- [BG96] BONET, B., AND GEFFNER, H. Arguing for decisions: A qualitative model of decision making. In *Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence* (1996), Morgan Kaufmann, pp. 98–105.
- [BLW10] BIENVENU, M., LANG, J., AND WILSON, N. From preference logics to preference languages, and back. In *Proceedings of the Twelfth International Conference on Principles of Knowledge Representation and Reasoning* (2010), KR'10, pp. 414–424.
- [CMG09] CHESÑEVAR, C. I., MAGUITMAN, A. G., AND GONZÁLEZ, M. P. *Empowering Recommendation Technologies Through Argumentation*. Springer US, Boston, MA, 2009, pp. 403–422.
- [CO19] CYRAS, K., AND OLIVEIRA, T. Resolving conflicts in clinical guidelines using argumentation. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems* (2019), AAMAS-19, pp. 1731–1739.
- [CRA⁺21] CYRAS, K., RAGO, A., ALBINI, E., BARONI, P., AND TONI, F. Argumentative xai: A survey. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21* (8 2021), International Joint Conferences on Artificial Intelligence Organization, pp. 4392–4399.
- [DBC03] DUNNE, P. E., AND BENCH-CAPON, T. Two party immediate response disputes: Properties and efficiency. *Artificial Intelligence* 149, 2 (2003), 221–250.
- [DG14] DUPIN DE SAINT-CYR BANNAY, F., AND GUILLAUME, R. Towards a transparent deliberation protocol inspired from supply chain collaborative planning. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems* (2014), A. Laurent, O. Strauss, B. Bouchon-Meunier, and R. R. Yager, Eds., vol. 443, Springer International Publishing, pp. 335–344.
- [DHKP11] DOMSHLAK, C., HÜLLERMEIER, E., KACI, S., AND PRADE, H. Preferences in AI: an overview. *Artificial Intelligence* 175, 7-8 (2011), 1037–1052.

- [DKT06] DUNG, P. M., KOWALSKI, R., AND TONI, F. Dialectic proof procedures for assumption-based, admissible argumentation. *Artificial Intelligence* 170, 2 (2006), 114–159.
- [DKT09] DUNG, P. M., KOWALSKI, R., AND TONI, F. *Assumption-Based Argumentation*. Springer US, Boston, MA, 2009, pp. 199–218.
- [dSCLS94] DE SAINT-CYR, F. D., LANG, J., AND SCHIEX, T. Penalty logic and its link with dempster-shafer theory. In *Proceedings of the 10th International Conference on Uncertainty in Artificial Intelligence (1994)*, R. L. de Mantaras and D. Poole, Eds., Morgan Kaufmann, pp. 204–211.
- [Dun95] DUNG, P. M. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence* 77, 2 (1995), 321–357.
- [EDP⁺22] ENGELMANN, D., DAMASIO, J., PANISSON, A. R., MASCARDI, V., AND BORDINI, R. H. Argumentation as a method for explainable ai : A systematic literature review. In *17th Iberian Conference on Information Systems and Technologies (CISTI) (2022)*, pp. 1–6.
- [Eur16] EUROPEAN UNION PARLIAMENT. General data protection regulation. <https://gdpr-info.eu> (2016), last access: 14/06/2022.
- [FCS⁺13] FAN, X., CRAVEN, R., SINGER, R., TONI, F., AND WILLIAMS, M. Assumption-based argumentation for decision-making with preferences: A medical case study. In *Computational Logic in Multi-Agent Systems (2013)*.
- [FEGS08] FERRETTI, E., ERRECALDE, M. L., GARCÍA, A. J., AND SIMARI, G. R. Decision rules and arguments in defeasible decision making. In *Proceedings of Computational Models of Argument (COMMA'08), Toulouse, France (2008)*, vol. 172 of *Frontiers in Artificial Intelligence and Applications*, IOS Press, pp. 171–182.
- [FEGS14] FERRETTI, E., ERRECALDE, M. L., GARCÍA, A. J., AND SIMARI, G. A possibilistic defeasible logic programming approach to argumentation-based decision-making. *Journal of Experimental & Theoretical Artificial Intelligence* 26, 4 (2014), 519–550.

- [Fis70] FISHBURN, P. C. *Utility Theory for Decision Making*. John Wiley and Sons, New York, 1970.
- [FT11] FAN, X., AND TONI, F. Assumption-based argumentation dialogues. In *Proceedings of the IJCAI International Joint Conference on Artificial Intelligence* (2011), pp. 198–203.
- [FT12] FAN, X., AND TONI, F. Agent strategies for aba-based information-seeking and inquiry dialogues. In *ECAI 2012 - 20th European Conference on Artificial Intelligence. Including Prestigious Applications of Artificial Intelligence (PAIS-2012) System Demonstrations Track, Montpellier, France* (2012), vol. 242 of *Frontiers in Artificial Intelligence and Applications*, IOS Press, pp. 324–329.
- [FT14] FAN, X., AND TONI, F. Decision making with assumption-based argumentation. In *Theory and Applications of Formal Argumentation* (2014), Springer Berlin Heidelberg, pp. 127–142.
- [FTG⁺17] FERRETTI, E., TAMARGO, L. H., GARCÍA, A. J., ERRECALDE, M. L., AND SIMARI, G. An approach to decision making based on dynamic argumentation systems. *Artificial Intelligence 242* (2017), 107–131.
- [FTMW14] FAN, X., TONI, F., MOCANU, A., AND WILLIAMS, M. Dialogical two-agent decision making with assumption-based argumentation. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems* (2014), AAMAS '14, International Foundation for Autonomous Agents and Multiagent Systems, pp. 533–540.
- [GCRS13] GARCÍA, A. J., CHESÑEVAR, C. I., ROTSTEIN, N. D., AND SIMARI, G. R. Formalizing dialectical explanation support for argument-based reasoning in knowledge-based systems. *Expert Systems with Applications 40*, 8 (2013), 3233–3247.
- [GDS09] GARCÍA, A. J., DIX, J., AND SIMARI, G. R. *Argument-based Logic Programming*. Springer US, 2009, pp. 153–171.

- [GF17] GOODMAN, B., AND FLAXMAN, S. European union regulations on algorithmic decision-making and a “right to explanation”. *AI Magazine* 38, 3 (2017), 50.
- [GPS20] GARCÍA, A. J., PRAKKEN, H., AND SIMARI, G. R. A comparative study of some central notions of ASPIC+ and DeLP. *Theory and Practice of Logic Programming* 20, 3 (2020), 358–390.
- [GS04] GARCÍA, A. J., AND SIMARI, G. R. Defeasible logic programming: An argumentative approach. *Theory Pract. Log. Program.* 4, 2 (jan 2004), 95–138.
- [GS14] GARCÍA, A. J., AND SIMARI, G. R. Defeasible logic programming: Delp-servers, contextual queries, and explanations for answers. *Argument & Computation* 5 (2014), 63–88.
- [GTWX16] GAO, Y., TONI, F., WANG, H., AND XU, F. Argumentation-based multi-agent decision making with privacy preserved. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems* (2016), AAMAS ’16, pp. 1153–1161.
- [HW12] HUNTER, A., AND WILLIAMS, M. Aggregating evidence about the positive and negative effects of treatments. *Artificial Intelligence in Medicine* 56, 3 (2012), 173–190.
- [Kac11] KACI, S. Preference representation languages. In *Working with Preferences Less Is More Less Is More* (2011), Springer Berlin Heidelberg, pp. 19–85.
- [KM03] KAKAS, A., AND MORAITIS, P. Argumentation based decision making for autonomous agents. In *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems* (2003), AAMAS ’03, Association for Computing Machinery, p. 883?890.
- [KR93] KEENEY, R. L., AND RAIFFA, H. *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*. Cambridge University Press, 1993.
- [Lif96] LIFSCHITZ, V. Foundations of logic programming. In *Principles of Knowledge Representation*, G. Brewka, Ed. CSLI Publications, 1996, pp. 69–128.

- [LM11] LEITE, J., AND MARTINS, J. Social abstract argumentation. In *IJCAI International Joint Conference on Artificial Intelligence (2011)*, pp. 2287–2292.
- [Mil19] MILLER, T. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence 267* (2019), 1–38.
- [MP14] MODGIL, S., AND PRAKKEN, H. The ASPIC+ framework for structured argumentation: a tutorial. *Argument & Computation 5*, 1 (2014), 31–62.
- [MRF⁺10] MOGUILLANSKY, M. O., ROTSTEIN, N. D., FALAPPA, M. A., GARCÍA, A. J., AND SIMARI, G. R. Argument theory change through defeater activation. In *Proceedings of the 2010 Conference on Computational Models of Argument: Proceedings of COMMA 2010* (NLD, 2010), IOS Press, p. 359?366.
- [MTB⁺13] MODGIL, S., TONI, F., BEX, F., BRATKO, I., CHESÑEVAR, C. I., DVOŘÁK, W., FALAPPA, M. A., FAN, X., GAGGL, S. A., GARCÍA, A. J., GONZÁLEZ, M. P., GORDON, T. F., LEITE, J., MOŽINA, M., REED, C., SIMARI, G. R., SZEIDER, S., TORRONI, P., AND WOLTRAN, S. *The Added Value of Argumentation*. Springer Netherlands, Dordrecht, 2013, pp. 357–403.
- [MTV09] MATT, P.-A., TONI, F., AND VACCARI, J. Dominant decisions by argumentation agents. In *6th International Workshop on Argumentation in Multi-Agent Systems (ArgMAS'09)* (2009), pp. 42–59.
- [NJ17] NUNES, I., AND JANNACH, D. A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction 27*, 3-5 (oct 2017), 393–444.
- [ODLM11] OUERDANE, W., DIMOPOULOS, Y., LIAPIS, K., AND MORAITIS, P. Towards automating decision aiding through argumentation. *Journal of Multi-Criteria Decision Analysis 18*, 5-6 (2011), 289–309.
- [Pra10] PRAKKEN, H. An abstract framework for argumentation with structured arguments. *Argument & Computation 1*, 2 (2010), 93–124.
- [PTV16] PIGOZZI, G., TSOUKIÀS, A., AND VIAPPANI, P. Preferences in artificial intelligence. *Annals of Mathematics and Artificial Intelligence 77* (08 2016), 361–401.

- [RCB⁺21] RAGO, A., COCARASCU, O., BECHLIVANIDIS, C., LAGNADO, D., AND TONI, F. Argumentative explanations for interactive recommendations. *Artificial Intelligence 296* (2021), 103506.
- [RCBT20] RAGO, A., COCARASCU, O., BECHLIVANIDIS, C., AND TONI, F. Argumentation as a framework for interactive explanations for recommendations. In *Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning* (9 2020), pp. 805–815.
- [RCT18] RAGO, A., COCARASCU, O., AND TONI, F. Argumentation-based recommendations: Fantastic explanations and how to find them. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18* (2018), International Joint Conferences on Artificial Intelligence Organization, pp. 1949–1955.
- [RMGS10] ROTSTEIN, N. D., MOGUILLANSKY, M. O., GARCÍA, A. J., AND SIMARI, G. R. A dynamic argumentation framework. In *Proceedings of the 2010 Conference on Computational Models of Argument: Proceedings of COMMA 2010* (2010), pp. 427–438.
- [TGGS15] TEZE, J. C., GOTTIFREDI, S., GARCÍA, A. J., AND SIMARI, G. R. Improving argumentation-based recommender systems through context-adaptable selection criteria. *Expert Systems with Applications 42*, 21 (2015), 8243–8258.
- [TGGS20] TEZE, J. C., GOTTIFREDI, S., GARCÍA, A. J., AND SIMARI, G. R. An approach to generalizing the handling of preferences in argumentation-based decision-making systems. *Knowledge-Based Systems 189* (2020), 105112.
- [TM07] TINTAREV, N., AND MASTHOFF, J. A survey of explanations in recommender systems. In *2007 IEEE 23rd International Conference on Data Engineering Workshop* (2007), pp. 801–810.
- [TM15] TINTAREV, N., AND MASTHOFF, J. *Explaining Recommendations: Design and Evaluation*. Springer US, Boston, MA, 2015, pp. 353–382.
- [Ton13] TONI, F. A generalised framework for dispute derivations in assumption-based argumentation. *Artificial Intelligence 195* (2013), 1?43.

- [Ton14] TONI, F. A tutorial on assumption-based argumentation. *Argument & Computation* 5, 1 (2014), 89–117.
- [VAHJ12] VISSER, W., AYDOGAN, R., HINDRIKS, K. V., AND JONKER, C. M. A framework for qualitative multi-criteria preferences. In *Proceedings of the 4th International Conference on Agents and Artificial Intelligence* (2012), ICAART '12, pp. 243–248.
- [VHJ12] VISSER, W., HINDRIKS, K., AND JONKER, C. Argumentation-based qualitative preference modelling with incomplete and uncertain information. *Group Decision and Negotiation* 21 (01 2012), 99–127.
- [VHJ13] VISSER, W., HINDRIKS, K. V., AND JONKER, C. M. Reasoning about interest-based preferences. In *Agents and Artificial Intelligence* (2013), Springer Berlin Heidelberg, pp. 115–130.
- [Wak17] WAKAKI, T. Assumption-based argumentation equipped with preferences and its application to decision making, practical reasoning, and epistemic reasoning. *Computational Intelligence* 33 (2017), 706–736.
- [Woo00] WOOLDRIDGE, M. Reasoning about rational agents. In *Intelligent robots and autonomous agents* (2000).
- [YH95] YOON, K. P., AND HWANG, C.-L. *Multiple attribute decision making. Quantitative Applications in the Social Sciences*. SAGE Publications, January 1995.
- [YH00] YOKOO, M., AND HIRAYAMA, K. Algorithms for distributed constraint satisfaction: A review. *Autonomous Agents and Multi-Agent Systems* 3 (2000), 185–207.
- [ZFLT14] ZHONG, Q., FAN, X., LUO, X., AND TONI, F. Explaining best decisions via argumentation. In *European Conference on Social Intelligence (ECSI-2014)* (2014), vol. 1283, pp. 224–237.
- [ZFLT19] ZHONG, Q., FAN, X., LUO, X., AND TONI, F. An explainable multi-attribute decision model based on argumentation. *Expert Systems with Applications* 117 (2019), 42–61.

- [ZFM⁺18] ZENG, Z., FAN, X., MIAO, C., LEUNG, C., JIH, C. J., AND SOON, O. Y. Context-based and explainable decision making with argumentation. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems* (2018), AAMAS '18, International Foundation for Autonomous Agents and Multiagent Systems, pp. 1114–1122.
- [ZST⁺20] ZENG, Z., SHEN, Z., TAN, B. T. H., CHIN, J. J., LEUNG, C., WANG, Y., CHI, Y., AND MIAO, C. Explainable and argumentation-based decision making with qualitative preferences for diagnostics and prognostics of alzheimer's disease. In *Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning* (2020), pp. 816–826.