Caracterización y Evaluación de Explicaciones Causales en Aplicaciones Computacionales

Fabrizio Meschini

Septiembre 2024

Índice

1. Introducción

El presente trabajo describirá un formalismo que apuntará a representar de un diálogo explicativo causal asimétrico entre dos participantes, un inquiridor y un explicador, en el que el segundo intenta transmitir los motivos detrás de un evento o suceso, abstraído a través una proposición propuesta por el primero. De acuerdo con lo planteado por Douglas Walton, durante el transcurso de este intercambio, "el explicador tomará esta proposición que el inquiridor no comprende e intentará clarificarla, relacionándola a otras que le resulten familiares y sea capaz de entender"[3]. A tales efectos, se hará uso de varios de los conceptos introducidos en *Programación en Lógica Rebatible: Lenguaje, Semántica Operacional y Paralelismo*[1] como mecanismos de representación de información. Algunos de ellos, como las reglas rebatibles y estructuras de argumento, serán resignificados o adaptados con el propósito de integrarlos de manera elegante a este nuevo contexto.

A diferencia de lo que ocurre en instancias de argumentación, los agentes involucrados en el diálogo explicativo no tienen como objetivo persuadir a su contraparte o hacer prevalecer alguna postura en particular. Sin embargo, y a pesar del espíritu colaborativo que lo caracteriza, la presencia en él de información contradictoria, sea ésta preexistente a su inicio o como consecuencia del mismo, es una posibilidad por demás tangible, especialmente al conjugar el conocimiento de ambos participantes. Dado que dicha información puede conducir a conclusiones contrapuestas, parte del objetivo de este trabajo será, en consecuencia, establecer lineamientos para lidiar con esta problemática en aquellas ocasiones donde corresponda.

Finalmente, se especificarán criterios de evaluación para la información recibida durante el proceso en función de los méritos de las estructuras a partir de éste construidas. De manera concordante, se otorgará al inquiridor la facultad de comunicar de manera adecuada el resultado de este análisis, así como también la de integrar a su base de conocimiento la porción del diálogo asociada.

2. Preliminares

2.1. Diálogo explicativo

Como se mencionó en la introducción, se entenderá el concepto de *explicación* como el resultado de un diálogo colaborativo entre dos partes, de aquí en adelante identificadas como *inquiridor* y *explicador*, en el que se buscará clarificar una proposición en particular, el *tópico* de la explicación. Esta conversación estará, por su parte, constituida de una secuencia finita de interacciones, y cada una de éstas, en cambio, por dos mensajes en sucesión: uno en el que el inquiridor comunicará alguna necesidad específica, y otro en el que el explicador transmitirá la información que considere pertinente a modo de respuesta.

Definición 2.1 (Mensaje de Solicitud)**.** Un *mensaje de solicitud* es un término $op(l)$, donde l es un literal y $op \in \{explain, relate\}$.

La intencionalidad detrás del uso de cada uno de los distintos mensajes de solicitud se resumen de la siguiente forma:

explain(l) El inquiridor solicita le sean transmitidas reglas a efectos de tornar *fundamentable* el literal l.

relate(l) El inquiridor solicita le sean transmitidas reglas a efectos de tornar *relevante* el literal l.

Los conceptos de *relevancia* y *fundamentabilidad* son presentados en la sección 3.

Definición 2.2 (Mensaje de respuesta)**.** Un *mensaje de respuesta* R es un conjunto finito de reglas.

Por caso, tanto $\{q \prec r, a \;,\; a \prec b, \sim c \;,\; b \prec \;\}$ como $\{\,\}$ califican como mensajes de respuesta válidos, mientras que uno como $\{a \;, \sim c \prec e \;, f \;, \sim b\}$ no.

Observación 2.1. Al describir conjuntos de reglas, se indicará la presencia simultánea de $(h \prec b_1, \ldots, b_k \;, \ldots b_n\;)$ como de $(b_k \prec\;)$ entre sus elementos mediante la notación $h \prec b_1, \ldots, \underline{b_k} \;, \ldots b_n$. Este marcado se aplicará a todas las reglas que involucren a $(b_k \prec\;)$; como consecuencia directa de esto último, todos los literales que no se hallen destacados de esta forma en las reglas incluidas en el conjunto no contarán con su presuposición asociada presente en el mismo. Por ejemplo, supóngase que se escribe la respuesta i-ésima como $R_i = \{(a \prec \underline{b}, \sim c)\,,\,(d \prec \underline{b})\}$. De ser así, $R_i = \{(a \prec b, \sim c)\,,\,(d \prec b)\,,\,(b \prec\ )\}$.

Definición 2.3 (Interacción)**.** Una *interacción* es un par $i : (Q, R)$, donde Q es un mensaje de solicitud y R es un mensaje de respuesta.

Los pares $(explain(q), \{q \prec r, a\,,\, r \prec b\})$ y $(relate(r), \{\})$ son, entonces, interacciones que se ajustan a las definiciones anteriores.

Definición 2.4 (Diálogo explicativo)**.** Un *diálogo explicativo* es una secuencia de interacciones $D :< i_1, i_2, \ldots, i_n >$, donde $Q_1 = explain(q)$ para algún literal q que se dirá el *tópico de la explicación*.

Ejemplo 2.1. *La siguiente secuencia ilustra un posible diálogo explicativo:*

$$i_1 : (explain(q), \{r \prec \underline{t}\})$$

$$i_2 : (relate(r), \{t \prec r, \underline{s}\})$$

$$i_3 : (relate(t), \{q \prec t\})$$

Lejos de oficiar como comandos que restrinjan el conjunto de posibles respuestas del explicador, los mensajes de solicitud oficiarán meramente como guías o sugerencias en relación a lo que el inquiridor espera de la respuesta venidera. Análogamente, el mismo tampoco se hallará obligado a denotar coherencia entre sus peticiones y la necesidades dialécticas expuestas en la solicitud inicial; por caso, nada le impedirá pedir que se expliquen literales, en apariencia, irrelevantes.

La elección de una modalidad más irrestricta de comunicación se basa en el intento de abstraer, más allá de las evidentes limitaciones del formalismo, lo más fielmente posible un intercambio libre entre individuos. En el contexto del presente trabajo, las condiciones que se imponen atañen únicamente a la sintaxis de los mensajes enviados por cada participante, quedando exentos tanto la intención como el significado de los mismos. Sin embargo, de considerarse apropiado o necesario, sería posible incorporar imposiciones que fuercen al explicador a únicamente incluir en su contestación reglas concernientes a los literales presentados en los mensajes de solicitud, o que limiten el conjunto de aquellos por los que el inquiridor está habilitado a consultar, por caso.

A pesar de lo manifiesto, en la totalidad de los ejemplos presentados en este trabajo se asumirá la presencia de una actitud colaborativa entre los quienes conversan. De este modo, por ejemplo, se entenderá que el inquiridor enviará el mensaje $relate(r)$ sólo cuando tenga una necesidad real de relacionar el literal r con el tópico abordado, y resultará coherente esperar que, a modo de respuesta, se obtendrá la información que el explicador posea y considere de utilidad para satisfacer dicha necesidad.

2.2. Caracterización del inquiridor

En el apartado anterior se introdujeron los dos participantes del diálogo explicativo, junto con los lineamientos que sus mensajes deben respetar conforme su rol asignado en el intercambio. Sin perjuicio de ello, dado que el foco del presente estudio se encuentra posado enteramente sobre la experiencia del inquiridor durante el proceso, se optará por únicamente caracterizar formalmente a este último.

En primer lugar, resulta imperativo proveer al agente interesado en la explicación de un mecanismo de acceso a su conocimiento preexistente, así como también a aquel del cual es provisto tras cada sucesiva interacción. Si bien la capacidad de distinguir entre ambas categorías no hace a su capacidad razonar sobre él, muchos de los criterios a presentar más adelante dependerán de que el inquiridor sea capaz de distinguir el origen de cada una de las reglas en su poder.

Definición 2.5 (Base de conocimiento del inquiridor)**.** Durante un diálogo explicativo, la *base de conocimiento del agente inquiridor luego de la i-ésima interacción* quedará caracterizada por el par (K, L_i), donde tanto K como L_i son conjuntos de reglas y L_i se define de la siguiente manera:

1. Si $i = 0$, $L_0 = \emptyset$.
2. Si $i > 0$, $L_i = L_{i-1} \cup S_i$, donde $S_i \subseteq R_i$ es tal que $S_i \cap \mathcal{K} = \emptyset$ y, de existir $S_i{}' \subseteq R_i$ tal que $S_i{}' \cap \mathcal{K} = \emptyset$, entonces $S_i{}' \subseteq S_i$.

El conjunto $\mathcal{K}$ representa el conocimiento propio con el que el agente cuenta previo al inicio del intercambio, y permanece inmutable a través del mismo.

L_i, por su parte, es indicativo de la totalidad de las reglas con las que se ha hecho luego de la i-ésima respuesta; dado que no existe una 0-ésima, L_0 denota el estado del aprendizaje antes de que la primera interacción tenga lugar. Como en este punto aún no se ha transmitido información alguna, $\mathcal{L}_0$ es necesariamente el conjunto vacío.

La restricción $S_i \cap \mathcal{K} = \emptyset$ podría omitirse si, en lugar de en lugar de L_i, se estableciera $(K \cup L)_i$ como $(K \cup L_{i-1}) \cup S_i$. Sin embargo, al menos tres decisiones de diseño diferentes sustentan la elección original. Por un lado, el cambio en L_i es más afín al proceso incremental al que es sometido el conjunto de aprendizaje, ajeno a $\mathcal{K}$. Por otra parte, desde una perspectiva semántica, esto representa más fielmente la reacción de un inquiridor hipotético ante la introducción de información que ya conoce de antemano. Finalmente, y quizás más concerniente a las definiciones que seguirán, la clara distinción entre las reglas preexistentes y aquellas adquiridas resulta fundamental al momento de clasificar posibles conflictos.

En lo que sigue, y a excepción de que sea necesario hacer la distinción, se usará simplemente $\mathcal{L}$ para hacer referencia al estado más reciente del conjunto de aprendizaje; esto es, si el diálogo explicativo se compone hasta el momento de $< i_1, i_2, \ldots i_n >$, $\mathcal{L} = \mathcal{L}_n$. Asimismo, se empleará $\mathcal{B}$ para simbolizar el conjunto resultante de la unión de $\mathcal{K}$ y $\mathcal{L}$, es decir, $\mathcal{B} = \mathcal{K} \cup \mathcal{L}$. De manera acorde, $\mathcal{B}_i = \mathcal{K} \cup \mathcal{L}_i$.

Previo a su comienzo, y especialmente durante el transcurso del diálogo, resulta por demás probable que el inquiridor sea capaz de construir estructuras de explicación para un conjunto variado de proposiciones. En particular, se espera que, a su finalización, este último cuente con el conocimiento necesario para concebir una explicación concerniente al tópico que haya decidido abordar. Como se mostrará más adelante, la facultad de atribuir un "responsable" a cada una de estas ocupará un rol fundamental no sólo al momento de decidir si dar por terminado el intercambio, sino también la manera de proceder ante posibles problemáticas surgidas del arribo a conclusiones en directa contraposición en base a la misma base de conocimiento.

Definición 2.6 (Criterio de Asociación)**.** Un *criterio de asociación* es una función Γ que toma como entrada una estructura de explicación $E : \langle A, q \rangle$ y retorna un elemento del conjunto $\{$*explainee*, *explainer*$\}$, y a la vez verifica

1. $\Gamma(E) =$ *explainer*, si toda regla r en A pertenece a $\mathcal{L}$.
2. $\Gamma(E) =$ *explainee*, si toda regla r en A pertenece a $\mathcal{K}$.

A excepción de aquellos casos en donde se encuentre constituida por reglas provenientes de un mismo conjunto, la definición anterior no establece de qué manera vincular una estructura de explicación dada con alguna de las dos salidas posibles, quedando esto enteramente librado a la discreción del inquiridor. La variada índole de los posibles aspectos a ponderar tornan la elección de un criterio en una tarea no trivial. *¿Basta con contar cuántas reglas provienen de $\mathcal{K}$ y cuántas de $\mathcal{L}$? ¿Qué ocurre si están equiparadas? ¿Debe atribuirse mayor importancia a aquellas más "cercanas" a la conclusión, o a las menos distantes a los hechos o presuposiciones?*. La posibilidad de poder ponderar uno por sobre otro permite la caracterización de agentes inquiridores de naturaleza diversa.

Definición 2.7 (Regla dominante)**.** Sea $E : \langle A, q \rangle$ una explicación. Se llamará *regla dominante* a aquella regla $r \in A$ tal que la cabeza de r es q.

Observación 2.2. La regla dominante necesariamente existe y es única: es imperativo que exista alguna regla con cabeza q en el conjunto A (de no ser así, q no podría ser derivado), y, si existiera más de una que cumpliera con este requerimiento, A no sería mínimo en cardinalidad como exige la definición de estructura de explicación.

Si bien no es el objetivo de este trabajo imponer un criterio de asociación particular, se presentará a continuación aquel a emplearse en aquellos ejemplos que demanden la utilización de alguno. El mismo, además de resultar de cómputo sencillo, hace uso de la noción previamente introducida de regla dominante.

Definición 2.8 (Criterio de asociación basado en regla dominante)**.** Sea E una explicación, r su regla dominante. El *criterio de asociación basado en regla dominante* Γ_D está definido de la siguiente manera:

- $\Gamma_D(E) =$ *explainer*, si $r \in \mathcal{L}$.
- $\Gamma_D(E) =$ *explainee*, si $r \in \mathcal{K}$.

Ejemplo 2.2. *Supóngase que, luego de la i-ésima interacción, el estado de $\mathcal{K}$ y $\mathcal{L}$ es*

$$\mathcal{K} = \{(c \prec\,), (b \prec d, {\sim}e\,), ({\sim}x \prec\,), (y \prec z\,)\}$$

$$\mathcal{L} = \{(q \prec \sim a, b\,)\,,\,(\sim w \prec \sim x\,)\,,\,(d \prec \underline{f}\,)\,,\,(\sim e \prec \underline{g}, c\,)\}$$

y considérese la estructura de explicación para el literal q

$$E : \langle\{(q \prec \sim a, b\,)\,,\,(\sim a \prec \underline{c}\,)\,,\,(b \prec d, \sim e\,)\,,\,(d \prec \underline{f}\,)\,,\,(\sim e \prec \underline{g}, \underline{c}\,)\},\ q\,\rangle$$

Dado que la regla dominante de E es $(q \prec \sim a, b\,)$*, y la misma es un miembro de* $\mathcal{L}$*, entonces*

$$\Gamma_D(E) = explainer$$

El último elemento constitutivo de un inquiridor será un *criterio de comparación entre explicaciones*, $\prec$, afín al concepto análogo para argumentos propio de DeLP. Dado que se lo empleará tal y como se encuentra allí definido (a excepción, claramente, del cambio en su nomenclatura para denotar el nuevo del dominio de aplicación), se optará por no hacerlo nuevamente aquí. De este modo, se procede a conjugar los conceptos introducidos en esta sección en la definición siguiente.

Definición 2.9 (Estado interno del inquiridor)**.** Durante un diálogo explicativo, se caracterizará el *estado interno del inquiridor* como la tupla $(\prec\,,\,\Gamma\,,\,\mathcal{B})$, donde $\prec$ es un criterio de comparación entre explicaciones, Γ es un criterio de asociación, y $\mathcal{B}$ es una base de conocimiento.

En el contexto donde se deba hacer referencia a la tupla anterior y se considere que no exista ambigüedad al respecto, se hablará sencillamente del inquiridor $(\prec\,,\,\Gamma\,,\,\mathcal{B})$.

3. Relevancia y Fundamentabilidad

En esta sección se explorarán brevemente dos conceptos que apuntan a retratar la manera en la que el inquiridor percibe cada respuesta en función de la información ya adquirida. A su vez, estas mismas nociones se generalizarán para llevar a cabo una evaluación análoga en relación al estado del intercambio, ya sea de manera parcial o total.

3.1. Relevancia

El inquiridor asociará la calidad de *relevante* con aquellos literales que considere se encuentran relacionados con el tópico abordado. Las definiciones próximas intentarán caracterizar esta apreciación.

Definición 3.1 (Relevancia directa entre literales)**.** Sean l_h y l_b dos literales, y $r : h \prec b_1, b_2 \ldots b_n$ una regla. Se dice que l_b es *relevante de forma directa* (o *directamente relevante*) *a* l_h *con respecto a* r si $l_b = b_i$ para algún i, con $1 \leq i \leq n$, y $h \in \{l_h, \overline{l_h}\}$.

Por ejemplo, p es relevante a q con respecto a $(q \prec p, r)$ dado que aparece en el cuerpo de esta última regla. Adicionalmente, y por el mismo motivo, también lo es a $\sim q$.

Definición 3.2 (Relevancia transitiva entre literales)**.** Sean S un conjunto de reglas, y h y l literales. Se dice que l es *relevante de forma transitiva a* h *con respecto a* S (o simplemente *relevante a* h *con respecto a* S) si se cumple alguna de las siguientes condiciones:

- l es directamente relevante a h con respecto a alguna regla $r \in S$.
- Existe un literal l' relevante a h con respecto a S, tal que l es directamente relevante a l' con respecto a alguna regla $r' \in S$.

Ejemplo 3.1. *Considérense los conjuntos* S_1 *y* S_2 *como se describe a continuación:*

$$S_1 = \{q \prec r\,, r \prec s\,\} \quad S_2 = \{\sim q \prec r\,\}$$

El literal s *es relevante a* q *con respecto a* S_1*, pero no lo es con respecto a*

S_2. Por su parte, r es relevante a q (y también a $\sim q$) con respecto tanto a S_1 como a S_2.

Definición 3.3 (Relevancia al tópico de la explicación)**.** Sea q el tópico de la explicación. Se dice que el literal l es *relevante al tópico de la explicación* (o simplemente *relevante*) si l es relevante a q con respecto a $\mathcal{B}$.

En lo que sigue, y dado que se trabajará con foco en el conjunto $\mathcal{B}$, si el literal l resulta relevante de forma general al literal h con respecto a este último, se dirá simplemente que l *es relevante a* h.

Una vez que el inquiridor determina que un literal es relevante, esta apreciación permanece inmutable durante la totalidad del intercambio. Sin embargo, dependiendo de la naturaleza de las respuestas vertidas por el explicador, la calidad de relevante de un literal puede no evidenciarse de manera inmediata.

Ejemplo 3.2. *Considérese el siguiente estado para $\mathcal{B}$ luego de la j-ésima interacción en el contexto de una explicación para a:*

$$B = \{a \prec b\ ,\ f \prec\ \}$$

Supóngase, además, que las respuestas $(j+1)$ y $(j+2)$-ésimas son los conjuntos $R_A = \{d \prec f\ \}$ y $R_B = \{b \prec d\ \}$, correspondientemente, y asúmase, además, que ambos conjuntos son incorporados a $\mathcal{B}$ (esto es, ninguna regla de las presentadas es rechazada). En el instante posterior a recibir R_A, pero previo a hacer lo propio con R_B, el inquiridor carece de los elementos suficientes como para considerar a los literales d y f como relevantes. Posteriormente, y luego de ser presentado con R_B, b, d, f y -trivialmente- a gozan de esta clasificación.

Si, por el contrario, R_A hubiera precedido a R_B, todos los literales mencionados habrían sido considerados relevantes de forma instantánea.

El ejemplo anterior ilustra cómo la forma en la que se estructuran las respuestas puede impactar en la percepción del inquiridor en relación a la relevancia de los literales involucrados en las mismas. Elegir transmitir una regla antes que otra puede conducir a que este último suponga que un literal no guarda relación con la proposición siendo clarificada, al menos hasta que se haga con las reglas que lo conecten a ella.

El concepto de relevancia puede generalizarse para alcanzar no sólo a literales, sino también a conjuntos de reglas. La *relevancia de una respuesta* evalúa cada respuesta y la califica en función de cuánta relación guarda con la proposición siendo explicada.

Definición 3.4 (Relevancia de una respuesta)**.** Sea R_i la i-ésima respuesta. Se define al número $\Psi(R_i)$, la *relevancia de la respuesta* R_i (o la *relevancia local a* R_i), como

$$\Psi(R_i) = \frac{|Rel_{R_i}|}{|R_i|}$$

donde Rel_{R_i} es el conjunto de los reglas de R_i cuyas cabezas son relevantes en relación al conjunto $\mathcal{B}_i$.

Existen dos razones por las cuales la relevancia de una respuesta R debe ser calculada en función al estado de conjunto de conocimiento en el instante inmediatamente posterior a incorporarla, y no en alguno previo o posterior. En primer lugar, puede existir interdependencia entre las reglas de R en lo que a relevancia respecta. Por ejemplo, supóngase que se está tratando de explicar q, $\mathcal{B} = \{x\}$, y la última respuesta recibida es $R = \{a \prec b\ ,\ b \prec c\}$. Resulta lógico considerar al literal c como relevante dado que "conduce" a a a través de la regla de la que es cuerpo, $(b \prec c)$, vía $(a \prec b)$, que se halla también en R, lo que justifica su incorporación a $\mathcal{B}$ previamente a llevar a cabo el cómputo. En este caso, $\Psi(R) = \frac{2}{2} = 1$.

Por otra parte, y como ya se había mencionado, el inquiridor juzga la relevancia de un literal en función del conocimiento con el que cuenta al momento de recibir una regla que lo involucra, y es incapaz de involucrar en este análisis aquel que aún no le ha sido transmitido. Para ilustrar esta situación, considérese una leve modificación al ejemplo presentado en el párrafo anterior, resultando en un contexto similar al del ejemplo 3.2: en lugar de recibir R como única respuesta, supóngase que se perciben de forma consecutiva $R_A : \{b \prec \underline{c}\}$ y $R_B : \{a \prec b\ ,\ g \prec\ \}$. Si esto fuera así, $\Psi(R_A) = \frac{0}{1} = 0$ y $\Psi(R_B) = \frac{1}{2} = 0{,}5$. Por otra parte, si, en cambio, R_B precediera a R_A, la relevancia local de R_B se mantendría inmutable (b es directamente relevante a a, y g no guarda relación con ningún literal en $\mathcal{B}$), pero ahora $\Psi(R_A) = \frac{1}{1} = 1$.

Extender el análisis de relevancia al conjunto de aprendizaje se presenta como una alternativa interesante si el inquiridor desea evaluar la proporción de la información hasta ahora adquirida que le está resultando pertinente al tópico del diálogo. En particular, al finalizar el mismo, esta métrica puede emplearse para cuantificar la eficiencia del intercambio contrastando la cantidad de reglas que atañeron a este último con la correspondiente a aquellas que sí lo hicieron.

Definición 3.5 (Relevancia general)**.** Se define el número Ψ_i, la *relevancia general* del diálogo luego de la i-ésima respuesta, como

$$\Psi_i = \frac{|Rel_{\mathcal{L}_i}|}{|\mathcal{L}_i|}$$

donde $Rel_{\mathcal{L}_i}$ es el conjunto de las reglas de $\mathcal{L}_i$ cuyas cabezas son relevantes.

En particular, si el diálogo explicativo está compuesto por las interacciones $< i_1, i_2, \ldots, i_n >$, el número Ψ_n se escribirá directamente como Ψ.

Ejemplo 3.3. *Considérese el siguiente estado inicial para $\mathcal{K}$, y asúmase q como tópico de la explicación:*

$$\mathcal{K} = \{(\sim r \prec \sim t, \sim u\,)\,,\,(s \prec\,)\}$$

A partir del mismo, se incorpora a $\mathcal{L}$ el contenido de las siguientes respuestas:

R_1*:* $\{(p \prec s\,)\,,\;(\sim c \prec \underline{\sim d}\,)\}$

R_2*:* $\{q \prec p, \sim r\,\}$

R_3*:* $\{\sim k \prec l\,\}$

R_4*:* $\{(\sim t \prec\,)\,,\;(\sim u \prec \underline{v}\,)\,,\;(\sim c \prec e\,)\,,(b \prec \underline{a}, c\,)\}$

Luego, se tiene:

i	$\Psi(R_i)$	Ψ_i
1	$\frac{0}{3} = 0$	$\frac{0}{3} = 0$
2	$\frac{1}{1} = 1$	$\frac{2}{4} = 0{,}5$
3	$\frac{0}{1} = 0$	$\frac{2}{5} = 0{,}4$
4	$\frac{3}{6} = 0{,}5$	$\frac{5}{11} \approx 0{,}45$

3.2. Fundamentabilidad

La noción de *fundamentabilidad* aplica a aquellos literales para los cuales es posible construir una estructura de explicación que los sustente.

Definición 3.6 (Literal fundamentable)**.** Una literal l se considerará *fundamentable* si existe una regla $r \in \mathcal{B}$ con cabeza l que verifique alguna de las siguientes condiciones:

- r es una presuposición.
- Para cada literal b en el cuerpo de r, b es fundamentable.

De ser así, se dirá que r *fundamenta* a l.

Ejemplo 3.4. *Considérese el siguiente estado para $\mathcal{B}$:*

$$B = \{ \ f \prec g \ , \quad a \prec b \ , \quad b \prec \quad \}$$

Los literales a y b son fundamentables, mientras que tanto f como g no lo son. La regla $a \prec b$ fundamenta a a, mientras que la presuposición $b \prec$ hace lo propio con b.

El hecho de que una regla fundamente a un literal cobrará especial importancia al extender el concepto a conjuntos de estas últimas. Para entender por qué, considérese a modo de ejemplo uno en que se incluya una regla $h \prec b_1, b_2, \ldots, b_n$, para la cual todo literal b_i es fundamentable. En consecuencia, h necesariamente deberá serlo también. Con el propósito de que esta métrica apunte a ilustrar la proporción de reglas en condiciones de usarse en una estructura de explicación, sería incorrecto extender automáticamente esta clasificación a todas las demás que tengan también a h como cabeza. Por caso, supóngase que se recibe la respuesta

$$R = \{(h \prec a, b\,), (\sim x \prec y\,), (h \prec \sim c\,)\}$$

y asúmase que tanto a como b son fundamentables, pero que ni $\sim c$ ni y lo son. De este modo, y a pesar de la condición de h, se tiene $\Theta(R) = \frac{1}{3} \approx 0{,}33$.

Definición 3.7 (Fundamentabilidad de una respuesta)**.** Sea R_i la i-ésima respuesta. Se define al número $\Theta(R_i)$, la *fundamentabilidad de la respuesta R_i* (o la *fundamentabilidad local a R_i*), como

$$\Theta(R_i) = \frac{|Fun_{R_i}|}{|R_i|}$$

donde Fun_{R_i} es el conjunto de las reglas de R_i para las cuales existe un literal l por ellas fundamentado.

Es visible la similitud del concepto anterior a su análogo en el contexto de relevancia. El motivo detrás de este parecido es lógico: se espera que aquello que el explicador escoja transmitir esté relacionado con el tópico abordado y, simultáneamente, sea tangible para el inquiridor.

Definición 3.8 (Fundamentabilidad general)**.** Se define el número Θ_i, la *fundamentabilidad general* del diálogo luego de la i-ésima respuesta, como

$$\Theta_i = \frac{|Fun_{\mathcal{L}_i}|}{|\mathcal{L}_i|}$$

donde $Fun_{\mathcal{L}_i}$ es el conjunto de las reglas de $\mathcal{L}_i$ para las cuales existe un literal l por ellas fundamentado.

En particular, si el diálogo explicativo está compuesto por las interacciones $< i_1, i_2, \ldots, i_n >$, el número Θ_n se escribirá directamente como Θ.

Ejemplo 3.5. *Supóngase el mismo estado inicial para $\mathcal{K}$ y sucesión de respuestas presentado en el ejemplo 3.3. Los valores correspondientes a la fundamentabilidad local a cada respuesta y general a la base de conocimiento son los presentados a continuación:*

i	$\Theta(R_i)$	Θ_i
1	$\frac{3}{3} = 1$	$\frac{3}{3} = 1$
2	$\frac{0}{1} = 0$	$\frac{3}{4} = 0{,}75$
3	$\frac{0}{1} = 0$	$\frac{3}{5} = 0{,}6$
4	$\frac{3}{6} \approx 0{,}83$	$\frac{9}{11} \approx 0{,}81$

3.3. Aplicación

Los criterios de éxito a presentarse en la sección 5 no harán uso de las nociones introducidas en esta para determinar si una estructura de explicación dada ha logrado satisfacer las necesidades dialécticas del inquiridor. Sin embargo, con el propósito de representar a uno particularmente demandante, el formalismo podría extenderse de manera sencilla para exigir, además de los requisitos a introducir como necesarios, condiciones que atañan a cualquiera de las dos métricas, como, por ejemplo, imponer la necesidad de que la relevancia local a cada respuesta siempre sea mayor a un número α, o que la fundamentabilidad general no caiga por debajo de β. Adicionalmente, ambas podrían combinarse para dar lugar a nuevas formas de juzgar el desempeño del explicador; por caso, podría observarse la fundamentabilidad de aquellos literales que pueden considerarse relevantes y viceversa.

La sección 2.1 introdujo el concepto de mensaje de solicitud, y expuso que sus variantes cuentan con sendos aspectos de los aquí explorados vinculados a ellas: *explain* atañe a la fundamentabilidad, mientras que *relate* hace lo propio con la relevancia. A pesar de que la capacidad de los agentes de enviar los mensajes que consideren apropiados es un aspecto intencional del formalismo (más allá de las obvias restricciones en relación a su sintaxis), las dos nociones pueden emplearse también para guiar el proceso de decisión detrás de cada uno de ellos.

Ejemplo 3.6. *El siguiente es un posible ejemplo de este tipo de aplicación:*

1. *explain*(l)*, para algún literal* l *relevante no fundamentable.*

1. *relate*(l)*, para algún literal* l *fundamentable no relevante para el cual exista una regla que lo tenga como cabeza en* $\mathcal{L}$*.*

2. *relate*(l)*, para algún literal* l *ni fundamentable ni relevante para el cual exista una regla que lo tenga como cabeza en* $\mathcal{L}$*.*

3. *Cualquier otro mensaje.*

El acto de valorar las dos primeras reglas por igual permite al inquiridor construir estructuras de explicación tanto de manera top-down (comenzando de la proposición a explicar y descendiendo hacia las presuposiciones) como bottom-up (comenzando de la información que se considera cierta y ascendiendo hacia el tópico), priorizando la información introducida al diálogo por el explicador.

4. Conflictividad

A pesar de no ser la finalidad del diálogo explicativo que la información en él vertida *prevalezca* por sobre alguna otra postura, nada impide que la misma acabe conflictuando con el conocimiento previo del inquiridor. Del mismo nodo, no existen restricciones impuestas que fuercen al explicador a transmitir información no contradictoria. Si bien a primera vista distinguir estas situaciones puede parecer intuitivo, origen posiblemente diverso de las reglas que componen a las estructuras de explicación complejiza considerablemente la tarea. Las definiciones que siguen pretenden capturar ciertas situaciones problemáticas que se consideran de particular importancia.

En el contexto de DeLP, se dice que los literales l_1 y l_2 se hallan *en desacuerdo* si el conjunto $\Pi \cup \{l_1, l_2\}$ es contradictorio, siendo Πaquel que aglomera tanto hechos como reglas estrictas. Dado que tanto $\mathcal{K}$ como $\mathcal{L}$ hacen uso únicamente de reglas rebatibles y presuposiciones, la definición acaba simplificándose para el contexto actual de la siguiente forma:

> **Definición 4.1** (Literales en desacuerdo)**.** Se dirá que los literales l_1 y l_2 se hallan *en desacuerdo* si son complementarios, es decir, $l_1 = \overline{l_2}$ (o, equivalentemente, $\overline{l1} = l2$).

Ejemplo 4.1. *Los literales a y $\sim a$ están en desacuerdo. Ninguno de ellos lo está con f o con $\sim d$.*

De forma similar, resulta necesario adaptar a continuación la definición de *contra-argumento o ataque*. Continuando con la práctica de llamar explicaciones a las estructuras de argumento, y teniendo presente la definición anterior, considérese lo siguiente:

> **Definición 4.2** (Explicaciones en conflicto)**.** Sean $E_1 : \langle A_1, q_1 \rangle$ y $E_2 : \langle A_2, q_2 \rangle$ dos estructuras de explicación obtenidas a partir de $\mathcal{B}$. Se dirá que E_2 se encuentra *en conflicto* con E_1 si y sólo si existe una sub-estructura de explicación $E_{sub} : \langle A_s, q_s \rangle$ de E_1 tal que q_2 y q_s se encuentran en desacuerdo. La explicación E_{sub} y el literal q_s se dirán la *sub-explicación de conflicto* y el *punto de conflicto*, respectivamente. De existir explicaciones E_1 y E_2 que satisfagan esta definición, se dirá -por consiguiente- que *existe* o que se ha *producido* el conflicto $\mathfrak{C} : E_2 \rightarrow E_{1_{E_{sub}}}$.

Observación 4.1. Análogamente a lo que ocurre en DeLP, si ocurre el conflicto $\mathfrak{C}_1 : A \to B_C$, entonces también se registra el conflicto $\mathfrak{C}_2 : C \to A_A$. En ocasiones como esta, en donde la estructura bajo ataque coincida con la sub-explicación de conflicto, el mismo se notará directamente como $C \to A$.

Ejemplo 4.2. *Supóngase el siguiente estado para la base de conocimiento:*

$\mathcal{K}$	$\mathcal{L}$
$(\sim b \prec f\,),\ (f \prec\ ),$	$(c \prec e\,),\ (d \prec\ ),\ (e \prec\ ),$
$(b \prec \sim c, d\,),\ (\sim c \prec\ )$	$(\sim a \prec c\,),\ (a \prec b\,)$

y considérense algunas de las estructuras de explicación que es posible construir a partir de ella

$$E_a : \langle\{(a \prec b\,), (b \prec \underline{\sim c}, \underline{d}\,)\},\ a\,\rangle$$

$$E_b : \langle\{(b \prec \underline{\sim c}, \underline{d}\,)\},\ b\,\rangle$$

$$E_{(\sim c)} : \langle\{\sim c \prec\ \},\ \sim c\,\rangle$$

$$E_{(\sim a)} : \langle\{(\sim a \prec c\,), (c \prec \underline{e}\,)\},\ \sim a\,\rangle$$

$$E_c : \langle\{c \prec \underline{e}\,\},\ c\,\rangle$$

$$E_{(\sim b)} : \langle\{(\sim b \prec \underline{f}\,)\},\ \sim b\,\rangle$$

A partir de lo anterior, es posible identificar los siguientes conflictos:

- $\mathfrak{C}_1 : E_a \to E_{(\sim a)}$ *y, correspondientemente,* $\mathfrak{C}_1{}' : E_{(\sim a)} \to E_a$.
- $\mathfrak{C}_2 : E_{(\sim c)} \to E_{(\sim a)}{}_{E_c}$ *y* $\mathfrak{C}_3 : E_c \to E_a{}_{E_{(\sim c)}}$. *Está claro que* E_c *y* $E_{(\sim c)}$ *están también mutuamente en conflicto.*
- $\mathfrak{C}_4 : E_b \to E_{(\sim b)}$. *Nuevamente,* $\mathfrak{C}_4{}' : E_{(\sim b)} \to E_b$.

Tal como ocurre con la estructura de las explicaciones involucradas en un conflicto, el agente asociado a cada una de aquellas en directa contraposición es un aspecto que resultará de utilidad al evaluarlo. En función al resultado de esta observación, y con el propósito de discernir más ágilmente su naturaleza, cada conflicto será clasificado en función al siguiente criterio:

Definición 4.3 (Clasificación de conflictos). Dadas las explicaciones E y E_{con} de manera tal que la segunda se encuentra en conflicto con la primera, con E_{sub} sub-explicación de conflicto. El conflicto anteriormente mencionado se clasificará en función al siguiente criterio:

- Si $\Gamma\left(E_{sub}\right) = \Gamma\left(E_{con}\right) = explainee$, el conflicto se dirá *dicotomía*.
- Si $\Gamma\left(E_{sub}\right) \neq \Gamma\left(E_{con}\right)$, el conflicto se dirá *discrepancia*.
- Si $\Gamma\left(E_{sub}\right) = \Gamma\left(E_{con}\right) = explainer$, el conflicto se denominará *inconsistencia*.

Ejemplo 4.3. *Considérense el resultado de aplicar la clasificación anterior a los conflictos presentados en el ejemplo 4.2 en función del criterio de asociación basado en regla dominante:*

- *Debido a que tanto la regla* $(\sim a \prec c)$ *como* $(a \prec b)$ *residen en* $\mathcal{L}$, $\mathfrak{C}_1$ *y* $\mathfrak{C}_1{}'$ *son* <u>*inconsistencias*</u>*.*
- $(\sim c \prec\ )$ *es un elemento de* $\mathcal{K}$, *mientras que* $(c \prec e)$ *proviene de* $\mathcal{L}$*. Los conflictos* $\mathfrak{C}_2$ *y* $\mathfrak{C}_3$ *son* <u>*discrepancias*</u>*.*
- $(\sim b \prec f)$ *y* $(b \prec \sim c, d)$ *son miembros de* $\mathcal{K}$*. En consecuencia,* $\mathfrak{C}_4$ *y* $\mathfrak{C}_4{}'$ *califican como* <u>*dicotomías*</u>*.*

5. Suficiencia, Aceptación y Éxito

Idealmente, luego de efectuar una consulta en relación a un tópico en particular, y a través de una serie de interacciones, el inquiridor debería ser capaz de formular una explicación para el mismo. Lo que sigue intenta establecer los lineamientos para determinar cuando esto efectivamente es así, y cuándo el resultado alcanzado puede considerarse exitoso.

5.1. Suficiencia

La noción de suficiencia establece los requerimientos mínimos que una explicación debe cumplir para siquiera considerarse evaluable.

Definición 5.1 (Explicación suficiente)**.** Sea q el tópico de la explicación, Γ un criterio de asociación y $E : \langle A, q \rangle$ una explicación construida para q a partir de $\mathcal{B}$ cuya regla dominante es r. La explicación E se dirá *suficiente* si se verifican de forma simultánea

1. $r \in \mathcal{L}$
2. $r \neq (q \mathrel{-\!\!\prec})$
3. $\Gamma\,(E) = \mathit{explainer}$

Observación 5.1. Como consecuencia de la naturaleza mínima del conjunto A, la condición 2 se verifica si y sólo si $A \neq \{q \mathrel{-\!\!\prec} \}$.

Observación 5.2. De adoptar el criterio de asociación basado en regla dominante, $\Gamma_D(E) = \mathit{explainer} \iff r \in \mathcal{L}$. En otras palabras, o se verificarían simultáneamente 1 y 3, o ninguna de ellas.

Observación 5.3. Como consecuencia directa de la definición de estructura de explicación para q, todos los literales que aparezcan en las reglas usadas en la construcción de E necesariamente resultan tanto relevantes como fundamentables.

La restricción impuesta al origen de la regla r en la definición anterior cumple un propósito doble. Por un lado, fuerza a emplear al menos una parte de la información adquirida; al fin y al cabo, sería cuestionable considerar que una explicación resultó de un diálogo si todo lo en él vertido por el explicador fue desestimado. Por otra parte, supóngase que existiera, previo a su inicio, en $\mathcal{K}$ una regla $q \mathrel{-\!\!\prec} b$, que esta fuera la usada en la explicación E, y que, en este instante, b no fuera fundamentable. Si esto ocurriera, parece razonable asumir que el interés real del inquiridor no pasaría tanto por comprender los motivos detrás de q, sino que el foco real del intercambio se centraría en hacer lo propio en relación a b.

La segunda condición de la definición apunta a evitar respuestas del estilo ”porque sí”: si el inquiridor da comienzo a un diálogo con el objeto de conocer los motivos detrás de la ocurrencia de la proposición q, es apropiado considerar que no estará dispuesto a considerar explicaciones que no denoten causante alguno.

5.2. Aceptación

Ostentar la calidad de suficiente no resguarda a una explicación de la posible presencia de información que contraste con la usada en su concepción. Lo que sigue tiene como propósito establecer un marco deductivo para identificar, de forma determinística, aquellas que, además de suscribir a la definición anterior, sean también capaces de soportar el escrutinio al que son sometidas por otras que sustenten conclusiones antagónicas.

Retomando el foco sobre una hipotética estructura de explicación E, en primer lugar, se impondrá una responsabilidad significativa sobre el explicador exigiendo que la información por él vertida no resulte contradictoria bajo la óptica del inquiridor. A tales efectos, se demandará que al momento de llevar a cabo el análisis no se registren inconsistencias a partir de $\mathcal{B}$, independientemente de si atañen a la explicación en cuestión o no.

A continuación, supóngase que la explicación E_{con} se encuentra en conflicto con E en E_{sub}, y asúmase $\Gamma(E_{con}) = explainer$:

- Si $\Gamma(E_{sub}) = explainer$, entonces dos explicaciones atribuidas al explicador se hallarían en conflicto directo, resultando en una inconsistencia. Como consecuencia de lo mencionado anteriormente, esto no puede ocurrir.
- Si $\Gamma(E_{sub}) = explainee$, entonces el inquiridor estaría “enmendando” parte del conocimiento transmitido por el explicador con información propia, a pesar de que este último se habría manifestado (al menos en función del criterio de asociación en uso) de manera opuesta.

Cualquiera sea el caso, admitir que el conflicto en cuestión provenga del explicador supone faltar al espíritu del diálogo explicativo: o éste estaría obrando contrariamente a la expectativa de que su discurso resulte consistente, o este último estaría siendo manipulado por el inquiridor con el objetivo de arribar a una conclusión a la que, posiblemente, de otra manera no habría sido capaz de hacerlo.

Pese a que impedir que los ataques provengan del emisor necesariamente fuerza a que, de producirse alguno, éste provenga de quien ocupa el rol de receptor, existen razones adicionales -ajenas a esta restricción- que apoyan esta exigencia. Análogamente al caso anterior, supóngase ahora $\Gamma\left(E_{con}\right) =$ *explainee*:

- Si $\Gamma\left(E_{sub}\right) =$ *explainee*, el inquiridor contaría con al menos dos explicaciones a él asociadas en directa contraposición. Esta no es necesariamente situación que se desee evitar: a menudo, la carencia de información en relación a un tópico en particular tiene como producto la coexistencia de explicaciones dialécticamente correctas para proposiciones que no pueden verificarse simultáneamente. Esta circunstancia, justamente, suele oficiar de motivadora para la búsqueda de otras que sean capaces de dirimir el conflicto en favor de una postura u otra.
- Si $\Gamma\left(E_{sub}\right) =$ *explainer*, la explicación propuesta por el explicador conflictúa con alguna noción preconcebida del inquiridor. Esta ocurrencia es por demás habitual en las interacciones humanas, y, por consiguiente, resulta adecuado contemplarla como válida.

En base a lo anterior, se desprende que la existencia de ciertos conflictos con origen en el inquiridor no deberían impedir calificar a una explicación como aceptable. Distinguir cuáles, precisamente, es el último aspecto que resta por dilucidar.

Resulta razonable demandar, si se pretende tanto poder considerar a una explicación como meritoria como también tolerar la presencia de otras que sustenten posturas antagónicas, que la primera no sucumba dialécticamente ante las segundas. Con el propósito de verificar si esto ocurre es que aparece en escena el criterio de comparación de explicaciones ($\prec$) propio de la configuración del agente inquiridor.

De este modo, las consideraciones precedentes quedan conjugadas en la definición presentada debajo:

Definición 5.2 (Explicación Aceptable)**.** Sea E una explicación suficiente. Se dirá que E es *aceptable* si para todo conflicto $\mathfrak{C}$ existente a partir de $\mathcal{B}$ se cumple simultáneamente:

1. $\mathfrak{C}$ no es una inconsistencia.
2. Si $\mathfrak{C}$ es de la forma $E_{con} \rightarrow E_{E_{sub}}$ (esto es, $\mathfrak{C}$ tiene como sub-explicación de conflicto a una sub-estructura de explicación de E, E_{sub}), entonces también se verifican las siguientes dos condiciones:
   a) $\Gamma(E_i) = $ *explainee*, para todo $E_i \in \Lambda_I$, donde Λ_I es el conjunto de explicaciones de interferencia de E. En particular, Λ_I está compuesto por las estructuras de explicación $E_1, E_2, \ldots, E_n$, donde $E_1 = E_{con}$.
   b) El árbol de dialéctica para E_{con} no se encuentra marcado como U.

5.3. Éxito

Si la noción de aceptación pretende capturar la circunstancia en la que el inquiridor, pese a sus reservas, carece de elementos para desacreditar el contenido de una explicación, la de *satisfacción* intentará hacer lo propio con aquellos casos donde, además, la misma es interpretada como superadora en función de la información disponible. Coloquialmente hablando, una explicación satisfactoria es aquella que más allá de no perder, gana.

Definición 5.3 (Explicación Satisfactoria)**.** Sea E una explicación aceptable. E se dirá *satisfactoria* o *exitosa* si su árbol de dialéctica se encuentra marcado como U.

El ejemplo que se presentará a continuación ilustrará una situación en la que un mismo diálogo explicativo da a lugar a diferentes estructuras de explicación, y donde cada una de las tres nociones introducidas en esta sección quedará representada por al menos una de ellas.

Ejemplo 5.1. *Supóngase que, luego de la n-ésima interacción de un diálogo explicativo cuyo tópico es q, el estado actual de $\mathcal{B}$, descompuesto en $\mathcal{K}$ y $\mathcal{L}$, es el siguiente:*

$\mathcal{K}$	$\mathcal{L}$
$(c \prec)$, $(b \prec f)$, $(f \prec g, \sim h)$, $(a \prec c)$, $(q \prec r)$, $(x \prec)$, $(\sim y \prec z)$	$(q \prec \sim a, b)$, $(\sim a \prec c, d)$, $(d \prec \sim e)$, $(\sim e \prec)$, $(g \prec)$, $(\sim h \prec c)$, $(r \prec \sim s)$, $(\sim s \prec)$, $(q \prec x, \sim y)$, $(y \prec z)$, $(z \prec w)$, $(w \prec)$

A partir de esta base de conocimiento, es posible construir las explicaciones $E_1 : \langle A_1, q \rangle$, $E_2 : \langle A_2, q \rangle$ y $E_3 : \langle A_3, q \rangle$ para q, donde

- $A_1 = \{(q \prec \sim a, b), (\sim a \prec \underline{c}, d), (b \prec f), (f \prec \underline{g}, \sim h), (d \prec \underline{\sim e}), (\sim h \prec \underline{c})\}$
- $A_2 = \{q \prec r, r \prec \underline{\sim s}\}$
- $A_3 = \{(q \prec \underline{x}, \sim y), (\sim y \prec z), (z \prec \underline{w})\}$

Usando especificidad generalizada como criterio de comparación, y usando el de regla dominante para hacer lo propio en relación al criterio de asociación, se puede concluir lo siguiente:

- *La regla dominante de A_2, $q \prec r$, tiene origen en $\mathcal{K}$. Por este motivo, E_2 **no es suficiente**.*
- *E_3 satisface todas las condiciones de suficiencia. Sin embargo, es posible construir la explicación*

  $$B_3 : \langle \{(y \prec z), (z \prec \underline{w})\}, y \rangle$$

  en conflicto con E_3 con sub-explicación de conflicto

  $$E_3' : \langle \{(\sim y \prec z), (z \prec \underline{w})\}, \sim y \rangle$$

  *Dado que $\Gamma_D(B_3) =$ explainee, A_3 **no es aceptable**: de permitir A_3, el inquiridor estaría usando una regla propia para enmendar parte del discurso explicador con el propósito de alcanzar la conclusión deseada.*
- *E_1 también es suficiente, y se encuentra únicamente bajo ataque de*

  $$B_1 : \langle \{a \prec \underline{c}\}, a \rangle$$

  en la sub-explicación de conflicto

  $$E_1' : \langle \{(\sim a \prec \underline{c}, d), (d \prec \underline{\sim e})\}, \sim a \rangle$$

  *Dado que el conjunto de explicaciones de interferencia de E_1 está únicamente compuesto por B_1, se verficia $B_1 \prec E_1'$ y, adicionalmente, $\Gamma_D(B_1) =$ explainee, E_1 no sólo es aceptable, sino también **satisfactoria**.*

6. Finalización y Asimilación de Conocimiento

Para concluir, se introducirá un mecanismo mediante el cual el inquiridor será capaz de poner fin al diálogo explicativo, así como también un criterio de decisión que orientará el proceso de traspaso de la información vertida en este último al conjunto de conocimiento permanente del primero.

6.1. Mensajes de Finalización

En cualquier momento posterior a una respuesta, el inquiridor podrá dar por terminado el diálogo explicativo haciendo uso de ciertos mensajes destinados únicamente a este propósito.

Definición 6.1 (Mensaje de finalización)**.** Un *mensaje de finalización* es un término que respeta alguno de los siguientes formatos:

- $success(E_{suc})$
- $accepted(E_{acc_1}, E_{acc_2})$
- $rejected(E_{rej_1}, E_{rej_2})$
- $inconsistency(E_{inc_1}, E_{inc_2})$
- $unknown$

Donde cada E_i es una estructura de explicación.

Si se pretende que la función de los mensajes de finalización sea coherente con su clasificación, los mismos deberían aparecer una única vez durante la conversación y no deberían de registrarse interacciones posteriores a su emisión. A tales efectos, la siguiente extensión a la definición de diálogo explicativo garantiza simultáneamente su integración y el cumplimiento de ambas condiciones.

Definición 6.2 (Diálogo explicativo finalizado)**.** Se denominará *diálogo explicativo finalizado* a la secuencia $< i_1, i_2, \ldots, i_n, e >$, donde la subsecuencia $< i_1, i_2, \ldots, i_n >$ es un diálogo explicativo y e es un mensaje de finalización.

Las definiciones anteriores no incluyen restricción alguna en lo que a las explicaciones que el inquiridor es capaz de construir en base a $\mathcal{B}$ respecta, en particular aquellas asociadas a su solicitud inicial. Sin embargo, es razonable esperar que exista un correlato entre la elección de un mensaje de finalización en particular y la situación de éstas al momento de su emisión. El criterio detallado debajo apunta a lograr que esto sea, en efecto, así.

Definición 6.3 (Criterio de correctitud para el envío de mensajes de finalización)**.** Sea q el tópico de la explicación y $(\prec, \Gamma, \mathcal{B})$una configuración para el inquiridor. Un mensaje de finalización m se considerará *correcto* si, de no detectarse ningún conflicto $\mathfrak{C}$ a partir de $\mathcal{B}$ tal que $\mathfrak{C}$ sea una inconsistencia, suscribe al siguiente criterio:

- Si existe al menos una estructura de explicación **suficiente** E_α para q:
  - Si existe al menos una estructura de explicación **aceptable** E_β para q:
    - Si existe al menos una estructura de explicación **satisfactoria** E_γ para q:
      - $m = success(E_\gamma)$.
    - Si no:
      - $m = accepted(E_\beta,\ E_c)$, donde E_c es una estructura de explicación en conflicto con E_β tal que $\Gamma(E_c) = explainee$.
  - Si no:
    - $m = rejected(E_\alpha,\ E_c)$, donde E_c es una estructura de explicación en conflicto con E_α tal que $\Gamma(E_c) = explainer$ o su árbol de dialéctica esté marcado como U.
- Si no:
  - $m = unknown$.

Si $\mathfrak{C}$ fuera una inconsistencia, entonces $m = inconsistency(E_{inc_1},\ E_{inc_2})$, donde E_{inc_1} y E_{inc_2} son dos estructuras de explicación tal que el conflicto entre ellas es $\mathfrak{C}$.

Como se mencionó previamente, la noción de diálogo finalizado garantiza que el mensaje de terminación sea único, así como también que no existan solicitudes que lo sucedan en la secuencia. Integrando el criterio de correctitud, es posible asegurar también que el mismo ilustra con fidelidad la percepción del inquiridor con respecto al resultado del intercambio al momento de su envío.

Definición 6.4 (Diálogo explicativo correctamente finalizado)**.** Se llamará *diálogo explicativo correctamente finalizado* a aquel diálogo explicativo finalizado cuyo mensaje de finalización suscribe al criterio de correctitud presentado anteriormente.

Excepto en aquellos casos que se desee meramente ejemplificar lo contrario, se asumirá que el inquiridor apuntará a finalizar el diálogo explicativo de manera correcta.

Ejemplo 6.1. *Tómese como punto de partida la situación expuesta en el ejemplo 5.1. Dado que la explicación E_1 resultó satisfactoria, el mensaje de finalización que corresponde de apuntar a que el mismo sea correcto es*

$$success(E_1)$$

Supóngase ahora una situación idéntica, pero con la única diferencia de que la regla $q \prec \sim a, b$ no fue transmitida. En consecuencia, E_1 no puede ser construida, por lo que el mensaje de finalización que cabe ahora es

$$rejected(E_3,\ B_3)$$

Finalmente, supóngase que tampoco se cuenta con $q \prec x, \sim y$, imposibilitando ahora también la concepción de E_3. Dado que E_2 no cumple las condiciones necesarias para ser considerada siquiera suficiente, corresponde

$$unknown$$

6.2. Integración de conocimiento

Habiendo concluido el diálogo explicativo, el inquiridor deberá incorporar total o parcialmente el conocimiento que fue almacenado en $\mathcal{L}$ como resultado del mismo. Este proceso se verá guiado por la asignación

$$\mathcal{K} \leftarrow \mathcal{K} \cup \mathcal{L}'$$

donde $\mathcal{L}'$ es un subconjunto de aquel de aprendizaje. Al momento de identificar puntualmente qué elementos de $\mathcal{L}$ han de formar parte de $\mathcal{L}'$, se buscará lograr balancear dos intereses en ocasiones contrapuestos: por un lado, se intentará que la porción a integrar guarde coherencia con el mensaje de finalización que dio cierre al diálogo, en esto reconociendo el mérito del explicador si lo por éste expuesto conllevó resultó en una explicación satisfactoria o, cuanto menos, aceptable. Por otra parte, se deseará permitir un grado de libertad semejante al apreciable tanto en las solicitudes del inquiridor como en sus correspondientes respuestas, admitiendo cierta arbitrariedad en la selección.

A tales efectos, se optará por no especificar con rigurosidad absoluta los elementos que han de componer al conjunto $\mathcal{L}'$, ni tampoco se inhibirá al inquiridor de incorporar reglas de manera arbitraria. No obstante, sí se impondrán restricciones con respecto al contenido que no deberá quedar excluido, conforme se verifiquen las condiciones asociadas. Puntualmente, sea q el tópico de la explicación y m el mensaje de finalización que concluyó el diálogo explicativo correcto, se exigirá que se cumpla:

- $A \subseteq \mathcal{L}'$, si $m = \mathit{success}(\langle A, q \rangle)$.
- $(A_a \cup A_c) \subseteq \mathcal{L}'$, si $m = \mathit{accepted}(\langle A_a, q_a \rangle, \langle A_c, q_c \rangle)$.

En otras palabras, se esperará que si el conocimiento expuesto por el explicador fue lo suficientemente adecuado como para permitir al inquiridor construir una explicación cuanto menos aceptable, las reglas que la componen sean incorporadas. Si esto ocurre, pero la explicación en cuestión no logra alcanzar la calificación de satisfactoria, se deberá asimilar también la información propia a al menos una que provoque este impedimento; debe recordarse que, si bien la explicación $\langle A_c, q_c \rangle$ necesariamente tiene al inquiridor como figura asociada, esto no necesariamente implica que la totalidad de los elementos en A_c provengan de $\mathcal{K}$.

Ejemplo 6.2. *Considérense nuevamente lo ocurrido en el ejemplo 5.1. Como E_1 es considerada satisfactoria, todas las reglas que componen a A_1 deberán formar parte de la nueva iteración de $\mathcal{K}$. En particular, destacando únicamente aquellas provenientes de $\mathcal{L}$, se tiene*

$$\{(q \prec {\sim}a, b), ({\sim}a \prec c, d), (d \prec {\sim}e), ({\sim}h \prec c), (g \prec), ({\sim}e \prec)\} \subseteq \mathcal{L}'$$

7. Conclusión

El formalismo descripto logra, como se había propuesto inicialmente, modelar un diálogo explicativo entre dos agentes con bases de conocimiento posiblemente diferentes. Para hacerlo, se recurrió a nociones y conceptos basados fundamentalmente en el material señalado en la referencia de este documento, pero también en el conocimiento adquirido durante el cursado de la carrera de Licenciatura en Ciencias de la Computación.

El proceso no careció de puntos salientes que, ya sea por lo desafiante de su concepción o inherentemente interesante de los temas a los que atañeron, son, en retrospectiva, producto de un inmenso grado de satisfacción. En particular, se destacan los criterios de suficiencia, aceptación y éxito: descifrar qué significa realmente que una explicación logre su objetivo y, una vez decidido, vincularlo con los criterios de conflictividad de la sección previa fue, sin dudas, el reto más arduo de todos los que este trabajo demandó superar.

Son igualmente destacables las falencias y carencias presentes. Para empezar, la realidad dicta que muchas de las nociones expuestas no son más que simples abstracciones de fenómenos propios de la comunicación humana, cuyas complejidades el autor no puede pretender conocer ni capturar en su totalidad. Este entendimiento limitado conlleva inevitablemente que las definiciones introducidas presenten un grado considerable de arbitrariedad en su fundamentación; no cabe duda de que existen otras interpretaciones, igualmente o más meritorias, que justificarían lineamientos diferentes para los mismos conceptos.

Finalmente, no son necesariamente pocas las formas posibles de extender o refinar el modelo actual. Por caso, podrían incorporarse sistemas de reputación o confianza que asocien cada regla con un agente de origen, afectando la manera en que cada regla (y, por consiguiente, la explicación de la que forman parte) es valorada. Del mismo modo, podrían considerarse mecanismos afines a la fundamentabilidad y la comprensibilidad, enfocados en la brevedad esperada del diálogo por parte del inquiridor, constituyendo así otro modo de condicionar el accionar del explicador. Estas alternativas son sólo algunas que se espera (y se desea) puedan ser abordadas en trabajos futuros.

A. DeLP

Uno de los principales desafíos a superar al momento de idear el formalismo presentado en este trabajo consistió en la elección de un mecanismo de representación de la información que los agentes involucrados en el diálogo explicativo habrían de volcar en él. Luego de evaluar diferentes alternativas, y ponderando fundamentalmente aquellos contenidos estudiados a lo largo del cursado de la carrera de Licenciatura en Ciencias de la Computación, la opción que se impuso por sobre las demás en función de la cercanía temática entre lo aquí tratado y el estudio de la argumentación fue la *Programación en Lógica Rebatible*[1] (*DeLP*, por *Defeasible Logic Programming*).

Cercanía, no obstante, no constituye absoluta coincidencia: si bien la mayoría de las nociones propias DeLP pudieron ser integradas sin alteraciones, otras debieron recontextualizarse. La intención detrás de esta sección no es replicar el contenido de la obra original de forma indiscriminada, sino destacar ciertas definiciones que se consideran de particular importancia, sea por el rol preponderante que ocuparon en las secciones anteriores, o porque su nombre, su sintaxis o su semántica original han sido manipulados para preservar la coherencia temática de este último.

En primer lugar, los literales son aplicados de manera idéntica al modo en el que fueron definidos, aplicados e interpretados en DeLP, constituyendo el mecanismo principal de representación de proposiciones.

Definición A.1 (Literal)**.** Un literal l es un átomo a o un átomo negado $\sim a$, donde "$\sim$" representa la negación fuerte. Un literal a se dirá negativo si es un átomo negado, y positivo en caso contrario. Un literal fijo es un átomo fijo o un átomo fijo negado.

Si bien no se hizo mención a este hecho durante el desarrollo de este trabajo, todo uso de literales a lo largo del mismo ha asumido su calidad de fijos, es decir, el átomo que los representa carece de variables.

Ligeramente distinto es lo que ocurre con las reglas rebatibles y las presuposiciones. Nuevamente, su sintaxis adhiere a lo establecido por la obra original.

Definición A.2 (Regla Rebatible)**.** Una *regla rebatible* es un par ordenado, denotado $(Cabeza, Cuerpo)$, donde el primer elemento, $Cabeza$, es un literal fijo y el segundo, $Cuerpo$, es un conjunto finito de literales fijos. Una regla rebatible con cabeza l_0 y cuerpo $\{l_1, \ldots, l_n\}$ $(n \geq 0)$ se escribirá también como: $l_0 \prec l_1, \ldots, l_n$. Una regla rebatible con cuerpo vacío se llama *presuposición*.

Sin embargo, su interpretación semántica merece ser revista. Si en DeLP la regla $a \prec b$ debe interpretarse como "*creer en el la proposición b es motivo para creer en la proposición a*", en el contexto del diálogo explicativo esto debe leerse como "*existen motivos para creer que a ocurre como consecuencia de b*". En particular, la presencia de la presunción $a \prec$ en una base de conocimiento indica que el agente asociado considera que lo propuesto por a es un hecho que ha tenido lugar, y que o bien los motivos detrás de su ocurrencia carecen de relevancia relevantes, sea en términos generales o en relación específica al tópico siendo abordado.

Considérese, a continuación, la definición de *Programa Lógico Rebatible*:

Definición A.3 (Programa Lógico Rebatible)**.** Se denomina *programa lógico rebatible* a un conjunto $\mathcal{P}$, posiblemente infinito, de hechos, reglas estrictas y reglas rebatibles. En un programa $\mathcal{P}$ identificaremos con Θ al conjunto de hechos, con Ω al conjunto de reglas estrictas, y con Δ al conjunto de reglas rebatibles. Por conveniencia, también se denota con Π al conjunto $\Theta \cup \Omega$. Por lo tanto, en algunas ocasiones se refiere a un programa lógico rebatible $\mathcal{P}$ con el par $\mathcal{P}$=(Π,Δ), y, en los casos que se desee identificar los hechos del programa explícitamente, se emplea la terna $\mathcal{P}$=(Θ,Ω,Δ).

El análogo al programa lógico rebatible (PLR de aquí en adelante) en el contexto actual es, claramente, la base de conocimiento $\mathcal{B}$. Nuevamente, es fácilmente apreciable que, a pesar de cumplir funciones similares, existen también marcadas diferencias entre ambos. Mientras que la distinción entre los conjuntos en los que un PLR se descompone está guiada fundamentalmente por el tipo de elemento que cada uno de ellos alberga, $\mathcal{B}$ resulta de la unión de $\mathcal{K}$ y $\mathcal{L}$, ambos constituidos tanto de reglas rebatibles como de presuposiciones. Con respecto a esto último, no se ha hecho uso ni de hechos ni de reglas estrictas. Más allá de cualquier justificación basada en la simplificación de la tarea a mano, el motivo principal por el cual estos fueron excluidos es fundamentalmente conceptual: en la vida cotidiana, rara vez podemos estar completamente seguros de un hecho, y siempre existirá quien considere una mera suposición aquello que otro entiende como irrefutable.

Como último punto saliente, ciertos conceptos debieron ser renombrados por cuestiones de coherencia temática. A pesar de esta alteración, su sintaxis se ha mantenido inalterada en la transición. Quizás simultáneamente más evidente y más relevante es el caso de las *estructuras de argumento*.

Definición A.4 (Estructura de Argumento)**.** Sea h un literal y $\mathcal{P}$=(Π,Δ) un programa lógico rebatible, una *estructura de argumento* para h es un par $\langle\mathcal{A}, h\rangle$, donde $\mathcal{A}$ es conjunto de reglas rebatibles de Δ, tal que:

1. existe una derivación rebatible para h a partir de $\Pi \cup \mathcal{A}$.
2. $\Pi \cup \mathcal{A}$ es no contradictorio, y
3. $\mathcal{A}$ es minimal, es decir, no existe un subconjunto propio $\mathcal{A}'$ de $\mathcal{A}$ tal que $\mathcal{A}'$ satisface las condiciones (1) y (2).

Las tratadas aquí como *estructuras de explicación* o, en casos donde no existe ambigüedad resultante de la simplificación, meramente *explicaciones*, suscriben a esta misma definición, pero observando ciertos aspectos puntuales. Por un lado, el rol del conjunto Δ debe asumido por $\mathcal{B}$. Adicionalmente, puede salvarse la ausencia del conjunto Π asumiendo que, en lugar de faltar, se trata de un conjunto vacío inmutable. De este modo, se arriba a la definición siguiente:

Definición A.5 (Estructura de Explicación)**.** Sea h un literal y $\mathcal{B}$ una base de conocimiento del inquiridor. Una *estructura de explicación* (o simplemente *explicación*) es un par $\langle\mathcal{A}, h\rangle$, donde $\mathcal{A}$ es un subconjunto de $\mathcal{B}$ tal que:

1. Existe una derivación rebatible para h a partir de $\mathcal{A}$.
2. $\mathcal{A}$ no es contradictorio.
3. $\mathcal{A}$ es minimal, es deir, no existe un subconjunto propio $\mathcal{A}'$ de $\mathcal{A}$ tal que $\mathcal{A}'$ satisface las condiciones (1) y (2).

Parecido es lo que ocurre con otros conceptos abordados previamente, como lo es el criterio de comparación entre explicaciones (argumentos), o el conjunto de explicaciones (argumentos) de interferencia. Todos ellos se emplean prácticamente del mismo modo que en el texto original, pero, nuevamente, el cambio de denominación refleja el nuevo contexto en el que son utilizados y el propósito para el cual se emplean.

Referencias

[1] Alejandro Javier García. *Programación en Lógica Rebatible: Lenguaje, Semántica Operacional y Paralelismo*. PhD thesis, Universidad Nacional del Sur, 2000.

[2] J. Anthony Blair Ralph H. Johnson. *Logical Self-Defense*, chapter 1, pages 18 – 19. Key Titles in Rhetoric, Argumentation, and Debate. International Debate Education Association, 2006.

[3] Douglas Walton. *Fundamentals of Critical Argumentation*, chapter 2, pages 75–80. Critial Reasoning and Argumentation. Cambridge University Press, 2005.