Relación de Accesibilidad u Operador Modal: un Dilema para Formalizar la Evolución de las Creencias en el Cálculo de Situaciones

Robert Demolombe y Pilar Pozos Parra ⋆

ONERA Toulouse, France

Resumen. La formalización de interacciones entre agentes se enfrenta al problema de la modelación de la evolución de las creencias de los agentes. Apoyándonos en la solución propuesta por el grupo de investigadores "Cognitive Robotic Group de Toronto" para el caso de un agente, la cual esta basada en el cálculo de situaciones, mostramos una breve comparación de dos formalismos que proponemos para el caso de varios agentes.

1 Introdución

Uno de los problemas más importantes y difícil de resolver para formalizar los aspectos dinámicos es el "frame problem". Este consiste en expresar la idea siguiente (que en principio parece ser muy simple): después de haber realizado una acción, ninguna proposición cambia su valor de verdad, excepto las proposiciones que tienen una relación lógica con los efectos de la acción realizada. Muchas soluciones han sido estudiadas desde el inicio de la Inteligencia Artificial [Bro87], casi todas se caracterizan por salir del marco de la lógica clásica de primer orden.

La solución propuesta por R. Reiter [Rei91] ha sido extendida por algunos investigadores de Cognitive Robotic Group de Toronto [SL93,LL98,SPLL00]. En [DP00] hemos propuesto una solución alternativa, la cual presenta fuertes restricciones pero que, sin embargo, ha podido ser implementada parcialmente. La solución de Reiter es expresada en el cálculo de situaciones, el cual es una lógica clásica de primer orden con igualdad, donde se puede distinguir, por cada predicado que puede cambiar dinamicamente de valor de verdad, un argumento (el último) con un tipo especial, tipo "situación".

Por ejemplo, el hecho que un vaso x se rompa en la situación s es representado por: $roto(x, s)$. Las situaciones son denotadas ya sea por constantes (como la situación inicial S_0), o por variables (generalmente denotadas por s, s', s'', ...), o por términos de la forma $do(a, s)$ cuyo primer argumento es un término de tipo acción y el segundo es un término de tipo situación. Si $tirar(x)$ y $reparar(x)$ son de tipo acción, entonces los términos siguientes: $do(tirar(x), S_0)$, $do(reparar(x), do(tirar(x), S))$, $do(a, s)$, ... son de tipo situación.

⋆ ONERA/DTIM, 2 Avenue E. Belin B.P. 4025, 31055 Toulouse Cedex, France. e-mail: {demolomb,pozos}@cert.fr. Esta investigación ha sido subsidiada por CONACyT.

El significado intuitivo de la fórmula atómica $roto(X1, do(tirar(X1, S_0)))$ es que el vaso está roto en la situación que resulta después de realizar la acción de $tirar(X1)$ a partir de la situación S_0. Las variables de tipo acción y situación pueden ser cuantificadas. Por ejemplo, las fórmulas siguientes forman parte del cálculo de situationes: $\exists a \exists x\ \neg roto(x, do(a, do(tirar(x), S_0)))$, $\forall s\ roto(X1, do(tirar(X1), s))$. Veremos después que la posibilidad de cuantificar sobre las acciones y situaciones juega un rol muy importante en la solución propuesta para el frame problem.

Presentaremos primero la problemática apoyándonos en un ejemplo. Supongamos que en toda situación, si se realiza la acción de $tirar(x)$, el resultado es que el vaso x está roto y si se realiza la acción de $reparar(x)$, el resultado es que x no está roto. Expresado formalmente por:
$(S1)$ $\forall s \forall a \forall x(a = tirar(x) \rightarrow roto(x, do(a, s)))$
$(S2)$ $\forall s \forall a \forall x(a = reparar(x) \rightarrow \neg roto(x, do(a, s)))$.

La solución al frame problem tiene como hipótesis que no existen otras acciones o condiciones, sólo las que se expresan en el antecedente de $(S1)$ (respectivamente $(S2)$) y que tienen por efecto que en la situación $do(a, s)$ se tiene $roto(x)$ verdadero (respectivamente falso). Aún más simple, se supone que estas condiciones dan una definición completa de las situaciones que tienen por efecto que $roto(x)$ sea verdadero (respectivamente falso). Estas hipótesis se expresan formalmente como:
$(S3)$ $\forall s \forall a \forall x(\neg roto(x, s) \wedge roto(x, do(a, s)) \rightarrow a = tirar(x))$
$(S4)$ $\forall s \forall a \forall x(roto(x, s) \wedge \neg roto(x, do(a, s)) \rightarrow a = reparar(x))$,
se puede demostrar que el conjunto $(S1)$, $(S2)$, $(S3)$ y $(S4)$ es equivalente a (S),
(S) $\forall s \forall a \forall x(roto(x, do(a, s)) \leftrightarrow a = tirar(x) \vee roto(x, s) \wedge \neg(a = reparar(x)))$.

(S) es llamado un axioma de estado sucesor ("successor state axiom" en [Rei99]). Este axioma es suficiente para determinar el valor de verdad de $roto(x)$ en $do(a, s)$ para cualquier acción realizada.

Por ejemplo, si tenemos $\neg roto(X1, S_0)$, podemos deducir $\neg roto(X1, do(tirar(X2), S_0))$ y $roto(X2, do(tirar(X2), S_0))$. Lo que significa que después de tirar el vaso $X2$, éste se rompe pero el vaso $X1$ queda incambiable.

En general, por cada predicado p que puede cambiar su valor de verdad de una situación a otra, se tiene un axioma de estado sucesor de la forma (S_p). Para simplificar las notaciones hemos omitido todas las variables excepto a y s,
(S_p) $\forall s \forall a(p(do(a, s)) \leftrightarrow \Gamma_p^+(a, s) \vee p(s) \wedge \neg \Gamma_p^-(a, s))$.

Las condiciones Γ_p^+ y Γ_p^- conciernen a la situación s y la acción a pero no a la situación $do(a, s)$. Además suponemos que las dos no pueden ser verdaderas en la misma situación, es decir $\neg \exists s \exists a(\Gamma_p^+(a, s) \wedge \Gamma_p^-(a, s))$.

El conjunto de axiomas de la forma (S_p) determina el valor de verdad de cada fórmula atómica después de realizar cualquier acción. Lo cual resuelve el frame problem.

En la siguiente sección presentamos una extensión a este formalismo que se aplica a las creencias de un agente. Posteriormente mostramos dos extensiones que proponemos para el caso de las creencias de varios agentes, seguidas de una comparación entre ambas extensiones.

2 Extensión a la evolución de creencias de un agente: relación de accesibilidad

El cálculo de situaciones ha sido extendido a las creencias de un agente por Scherl y Levesque en [SL93]. Desde un punto de vista formal esta extensión se basa en la idea de expresar en lógica clásica de primer orden la definición de modalidad de creencia. Para esto, el lenguaje es agrandado con un predicado binario $B(s', s)$[1] donde los argumentos son de tipo situación. Este predicado juega el mismo rol que una relación de accesibilidad en la semántica de mundos posibles. La diferencia es que éste es ligado a las situaciones y no a los mundos posibles, además el predicado es introducido en la axiomática y no en la semántica.

El hecho que en una situación s un agente crea que ϕ es verdadera es denotado por $Bel(\phi, s)$, donde ϕ es una fórmula obtenida a partir de una fórmula del cálculo de situaciones en la cual todos los argumentos de tipo situación estan suprimidos. Se da por hecho que todos conciernen a la situación s.

Por ejemplo, a partir de la fórmula $roto(X1, s) \wedge \neg roto(X2, s)$ obtenemos $\phi_1 = roto(X1) \wedge \neg roto(X2)$ y $Bel(roto(X1) \wedge \neg roto(X2), s)$.

El significado atribuido a $Bel(roto(X1) \wedge \neg roto(X2), s)$ es que en la situación s el agente cree que en todas las situaciones s' compatibles con sus creencias (es decir, tales que $B(s', s)$), se tiene $roto(X1, s') \wedge \neg roto(X2, s')$ que es verdadero. Es por esta razón que las creencias se restringen a sólo incluir fórmulas ϕ donde todos los argumentos de tipo situación se refieran a la situación s que aparece en $Bel(\phi, s)$.

Si tenemos $Bel(\phi, s)$, denotamos $\phi[s']$ a la fórmula obtenida a partir de ϕ agregando el argumento s' a cada predicado que puede cambiar de valor de verdad. Tenemos así la definición de Bel.

$Bel(\phi, s) \stackrel{\text{def}}{=} \forall s'(B(s', s) \rightarrow \phi[s'])$,

por ejemplo:

$Bel(roto(X1) \wedge \neg roto(X2), s) \stackrel{\text{def}}{=} \forall s'(B(s', s) \rightarrow roto(X1, s') \wedge \neg roto(X2, s'))$.

Las fórmulas ϕ pueden contener modalidades. Por ejemplo, podemos tener $Bel(\neg Bel(roto(X2)), S_0)$ que expresa que en S_0 el agente cree que el no cree que $X2$ está roto.

Para definir la manera en que las creencias evolucionan cuando las acciones son realizadas, se distinguen dos aspectos:

El primero concierne a la evolución de la relación de accesibilidad $B(s', s)$. Se debe definir las situaciones s'' que son accesibles después de haber realizado la acción a a partir de s, es decir, tales que $B(s'', do(a, s))$. Para ello distinguimos dos tipos de acciones (ver figura 1):

– De percepción. Las que permiten saber si una información sobre el mundo representada por una fórmula ϕ es verdadera o falsa. Se da por hecho que éstas pueden modificar las creencias pero no modifican el mundo real.

[1] En [Che88] dentro de relación de accesibilidad $R(w, w')$, el mundo w' representa una alternativa al mundo w con respecto a la creencia. En este trabajo hemos respetado la notación utilizada por [Rei99] cuyo orden de los argumentos es inverso.

– Sobre el mundo. Las acciones que modifican el mundo. Por cada situación posible, éstas proyectan sus efectos sobre la situación siguiente y eventualmente, por consecuencia, modifican las creencias.

El segundo aspecto concierne a los valores de verdad de los átomos en las situaciones s'' accesibles desde $do(a, s)$ por B. Para ello se emplean los axiomas de estado sucesor (S_p).

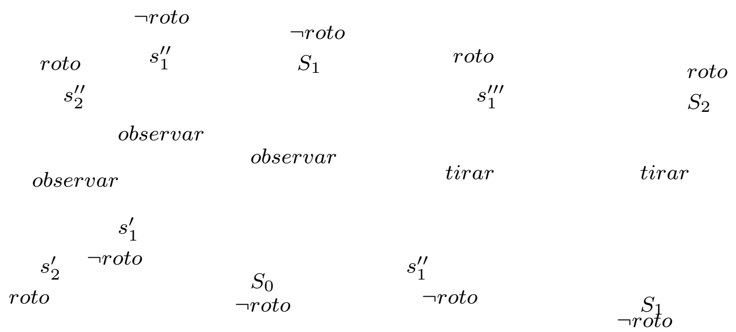

Fig. 1. Evolución de las creencias de un agente. Acción de percepción y acción sobre el mundo.

La forma general del axioma que define la evolución de la relación de accesibilidad B es el siguiente:

$$
\begin{aligned}
(S_B)\quad & \forall s \forall s'' \forall a (B(s'', do(a, s)) \leftrightarrow \exists s'(B(s', s) \wedge s'' = do(a, s') \wedge (\\
& (\neg(a = \alpha_1) \wedge \ldots \wedge \neg(a = \alpha_n)) \\
& \vee\ a = \alpha_1 \wedge (\phi_1(s) \leftrightarrow \phi_1(s')) \\
& \vdots \\
& \vee\ a = \alpha_n \wedge (\phi_n(s) \leftrightarrow \phi_n(s')))),
\end{aligned}
$$

donde cada α_i denota una acción de percepción que permite conocer el valor de verdad de la fórmula ϕ_i. En esta formalización de la evolución de las creencias se tienen implícitas dos hipótesis:

La primera es que cada vez que una acción ha sido realizada, el agente sabe que ha sido realizada. De hecho, las creencias de un agente evolucionan cada vez que una acción ha sido realizada. En caso de considerar sólo un agente y de no haber acciones "exógenas" (realizadas por el medio ambiente o por otro agente), la hipótesis es aceptable, pero no lo es cuando se consideran varios agentes.

La segunda hipótesis es que el agente sabe cuáles son los efectos de las acciones. De hecho, la evolución de sus creencias es determinada, entre otras cosas, por los axiomas de estado sucesor, los cuales son los mismos en s y en s' si $B(s', s)$. En el caso de varios agentes, puede ser que ciertos agentes crean que las cosas evalucionan de una manera y que otros crean que evolucionan de una manera diferente. Por consiguiente esta hipótesis tampoco es aceptable.

Para no restringirse al uso de estas hipótesis, hemos propuesto dos variantes del formalismo, que son descritas en las siguientes secciones. La primera es una extensión a este formalismo, mientras que la segunda es una extensión del cálculo de situaciones, mencionado en la introducción.

3 Evolución de creencias de varios agentes: relación de accesibilidad

Para representar formalmente el hecho que las creencias de cada agente pueden evolucionar de forma diferente, suponemos que existe una relación de accesibilidad B_i por cada agente i. Esto permite definir una modalidad $Bel_i(\phi, s)$ que representa las creencias del agente i. Así tenemos la definición siguiente:

$Bel_i(\phi, s) \stackrel{\text{def}}{=} \forall s'(B_i(s', s) \rightarrow \phi[s'])$.

Para no estar restringido a las hipótesis implícitas que mencionamos en la sección precedente, consideramos que:

- el conjunto de acciones cuya realización es conocida por un agente es dado explícitamente y puede ser diferente de un agente a otro,
- los axiomas de estado sucesor dependen de los tipos de situación a los que son aplicados. Ya sea que se trate de una situación real o una situación imaginada por un cierto agente, podemos tener axiomas de estado sucesor diferentes.

Primero presentamos un ejemplo de cómo se pueden expresar formalmente estos dos principios. Consideremos dos agentes: una madre y su hijo. Las acciones que puede realizar la madre son: $tirar_m$, $reparar_m$ y $observar_m$ y las que puede realizar el hijo son: $tirar_h$ y $observar_h$. Tenemos las relaciones de accesibilidad B_m y B_h que representan respectivamente las creencias de la madre y del hijo.

Supongamos que las acciones cuya realización es conocida por la madre son: $tirar_m$, $reparar_m$, $observar_m$ y $tirar_h$ y las conocidas por el hijo son: $tirar_h$, $observar_h$ y $reparar_m$.

En la solución que presentamos, consideramos que un agente razona como si las acciones que no conoce, jamas se hubiesen realizado. Así él ignora estas acciones siempre que no haya observado sus efectos. Esta actitud es de confianza o credulidad. En [LLR99], al contrario, la actitud del agente es demasiado prudente o escéptica.

El hecho que la realización de la acción $tirar_m$ no sea conocida por el hijo es expresado formalmente por el hecho que las situaciones de accesibilidad a partir de $do(tirar_m, s)$ por la relación B_h son las mismas que aquellas accesibles desde s por la relación B_h.

En caso de las acciones cuya realización es conocida por el hijo, la evolución de las creencias es definida de la manera siguiente:

$$\begin{aligned}
&\forall s \forall s'' \forall a(\neg(a = tirar_m) \wedge \neg(a = observar_m) \\
&\quad \rightarrow (B_h(s'', do(a, s)) \leftrightarrow \exists s'(B_h(s', s) \wedge s'' = do(a, s') \wedge (a = tirar_h \vee \\
&\quad a = reparar_m \vee\ a = observar_h \wedge (roto(s) \leftrightarrow roto(s')))))),
\end{aligned}$$

y para cualquier tipo de acción, la evolución de las creencias es definida como:

$$\begin{aligned}
&\forall s \forall s'' \forall a(B_h(s'', do(a, s)) \leftrightarrow \\
&\quad (B_h(s'', s) \wedge \neg(a = tirar_h) \wedge \neg(a = observar_h) \wedge \neg(a = reparar_m)) \vee \\
&\quad \exists s'(B_h(s', s) \wedge s'' = do(a, s') \wedge (\\
&\quad a = tirar_h \vee a = reparar_m \vee \\
&\quad a = observar_h \wedge (roto(s) \leftrightarrow roto(s')))).
\end{aligned}$$

Para representar la evolución de las creencias de la madre, se tiene un axioma similar. La forma general del axioma que define la evolución de una relación B_i es:

(S_{B_i}) $\forall s \forall s'' \forall a(B_i(s'', do(a, s)) \leftrightarrow$
$(B_i(s'', s) \wedge \neg(a = \alpha_1) \wedge \ldots \wedge \neg(a = \alpha_n) \wedge \neg(a = \beta_1) \wedge \ldots \wedge \neg(a = \beta_m))\vee$
$\exists s'(B_i(s', s) \wedge s'' = do(a, s')\wedge$
$(a = \beta_1 \vee \ldots \vee a = \beta_m$
$\vee\ a = \alpha_1 \wedge (\phi_1(s) \leftrightarrow \phi_1(s'))$
$\vdots$
$\vee\ a = \alpha_n \wedge (\phi_n(s) \leftrightarrow \phi_n(s')))),$

donde las alphas son las acciones de percepción realizadas por el agente i y las betas son el resto de las acciones cuya realización es conocida por el agente i. La forma de este axioma podría ser generalizada definiendo el conjunto de acciones conocidas por el agente i en la situación s utilizando una fórmula $\psi_i(a, s)$, en lugar de dar explícitamente estas acciones. La fórmula $\psi_i(a, s)$ podría expresar, por ejemplo, que a es conocida por i en s si la acción a es realizada por un agente que está en la misma habitación que i o en una vecindad próxima.

Para determinar la evolución de las creencias se deben proporcionar también los axiomas de estado sucesor que se aplican a las situaciones accesibles por B_m o B_h.

Supongamos, por ejemplo, que en S_0 el vaso no esté roto y que la madre y el hijo sepan que el vaso no está roto, Es decir $\neg roto(S_0) \wedge Bel_m(\neg roto, S_0) \wedge Bel_h(\neg roto, S_0)$. En la situación s'_1 y s'_2 respectivamente accesibles por B_h y B_m tenemos $roto(s'_1)$ y $\neg roto(s'_2)$.

Si el hijo cree que el vaso no se rompe si se tira y si en S_0 tira el vaso (acción $tirar_h$), entonces después de tirar el vaso (es decir, en la situación $do(tirar_h, S_0)$) cree que el vaso no está roto, por lo tanto, en las situaciones $do(tirar_h, s'_1)$ que son accesibles desde $do(tirar_h, S_0)$ por B_h, se debe tener $\neg roto(do(tirar_h, s'_1))$. Las creencias de la madre toman otra dirección, si ella cree que el vaso se rompe después de tirarlo, entonces en $do(tirar_h, S_0)$ cree que el vaso está roto, en consecuencia en las situaciones $do(tirar_h, s'_2)$ accesibles desde $do(tirar_h, S_0)$ por B_m se debe tener $roto(do(tirar_h, s'_2))$.

Se debe distinguir el axioma de estado sucesor que se aplica en s'_1 y el que se aplica en s'_2. Para ello, introducimos los predicados $cree_m(s)$ y $cree_h(s)$ que expresan intuitivamente que en s los axiomas de estado sucesor que se aplican son aquellos que corresponden a las creencias de la madre y del hijo respectivamente. Estos axiomas pueden ser, por ejemplo:

$(S_{roto,m})$ $\forall s(cree_m(s) \rightarrow (roto(do(a, s)) \leftrightarrow a = tirar_h \vee a = tirar_m \vee roto(s) \wedge \neg(a = reparar_m)))$

$(S_{roto,e})$ $\forall s(cree_h(s) \rightarrow (roto(do(a, s)) \leftrightarrow roto(s) \wedge \neg(a = reparar_m)))$.

Para definir las situaciones donde se aplican los axiomas de estado sucesor que corresponden a la realidad introducimos el predicado $real(s)$, que significa intuitivamente que s representa una situación real. Asi tenemos por ejemplo:

(S_{real}) $\forall s(real(s) \rightarrow (roto(do(a, s)) \leftrightarrow a = tirar_h \vee a = tirar_m \vee roto(s) \wedge \neg(a = reparar_m)))$.

En general, por cada predicado p y por cada agente i o por las situaciones reales, se debe tener un axioma de estado sucesor de la forma:
$(S_{p,i})$ $\forall s \forall a(cree_i(s) \rightarrow (p(do(a,s)) \leftrightarrow \Gamma^+_{p,i}(a,s) \vee p(s) \wedge \neg\Gamma^-_{p,i}(a,s)))$
(S_{real}) $\forall s \forall a(real(s) \rightarrow (p(do(a,s)) \leftrightarrow \Gamma^+_{p,0}(a,s) \vee p(s) \wedge \neg\Gamma^-_{p,0}(a,s)))$.

Supongamos que en S_0 el hijo cree que la madre cree que el vaso no está roto y que la madre cree que el hijo cree también que el vaso no está roto, formalmente $Bel_h(Bel_m(\neg roto), S_0)$ y $Bel_m(Bel_h(\neg roto), S_0)$. En consecuencia tenemos $Bel_m(\neg roto, s'_1)$ y $Bel_h(\neg roto, s'_2)$.

Sean s'_3 y s'_4 las situaciones accesibles desde s'_1 por B_m y desde s'_2 por B_h respectivamente, la pregunta es ¿se debe considerar s'_3 como una situación que representa las creencias de la madre o como una situación que representa las creencias del hijo con respecto a las creencias de la madre?

Si consideramos que s'_3 representa una creencia de la madre, es decir $cree_m(s'_3)$, entonces el axioma de estado sucesor que se aplica en s'_3 es $(S_{roto,m})$. Para ello, se debe suponer que el hijo cree que su madre cree que si se tira el vaso, éste se rompe, aunque el crea lo contrario. Este podría ser el caso, pero no es muy realista. Eventualmente se debe poder representar que éste no es el caso, es decir que en s'_3 el axioma $(S_{roto,e})$ es el que se aplica y que lo que se tiene es una creencia del hijo con respecto a las creencias de la madre, $cree_h(s'_3)$.

Si ahora consideramos que s'_4 representa las creencias de la madre con respecto a las creencias del hijo, podemos aceptar que la madre sepa que el hijo cree que el vaso no se rompe si se tira, en este caso aceptaremos aplicar en s'_4 el axioma $(S_{roto,e})$.

Desde un punto de vista práctico, no podemos expresar la naturaleza de las situaciones por cada situación s, proporcionando $cree_i(s)$, puesto que existe una infinidad de situaciones. Parece natural considerar que los agentes creen que los efectos de las acciones son siempre los mismos, es decir, que las situaciones siguientes a una situación s son de la misma naturaleza que esta situación. Expresado formalmente por el axioma:
(CR_1) $\forall s(cree_i(s) \leftrightarrow cree_i(do(a,s)))$.

Pero incluso limitándose a las situaciones accesibles desde la situación inicial, aún queda una infinidad de situaciones. Proponemos atribuir a los agentes dos actitudes posibles que pueden ser descritas por un número limitado de hechos de la forma $cree_i(s)$.

La primera supone que cada agente i sabe cómo evoluciónan las creencias de un agente j. Por ejemplo, éste podría ser el caso de la madre con respecto al hijo. Esta actitud se expresa formalmente por el hecho que la naturaleza de una situación es determinada unicamente por la relación de accesibilidad que accede a esta situación. Por ejemplo, tenemos $cree_m(s'_2)$ porque s'_2 es accesible por B_m y $cree_h(s'_4)$ porque s'_4 es accesible por B_h. En general tenemos el esquema de axioma siguiente:
(CR_2) $\forall s \forall s'(B_i(s', s) \rightarrow cree_i(s'))$.

La segunda supone que cada agente i cree que las creencias de los otros agentes evolucionan como sus propias creencias, como podría ser el caso del hijo que cree que su madre cree, como él, que un vaso no se rompe si se tira. En este

caso, la naturaleza de una situación será determinada por el "agente primero" que imagina las creencias de los otros agentes. Por agente primero designamos al agente que está en una situación real y no en una situación imaginada por un agente. En este caso tenemos, por ejemplo, $cree_h(s'_1)$ porque s'_1 es accesible por B_h desde S_0 y además $real(S_0)$. También tenemos $cree_h(s'_3)$, aunque la relación que accede a s'_3 sea B_m, porque s'_3 es accesible desde s'_1 y tenemos que $cree_h(s'_1)$. En general tenemos los esquemas de axioma siguientes:
(CR_3) $\forall s \forall s'(real(s) \wedge B_i(s', s) \rightarrow cree_i(s'))$
(CR_4) $\forall s \forall s'(cree_i(s) \wedge B_j(s', s) \rightarrow cree_i(s'))$.

No importando cuál actitud se elija, debe existir un solo axioma de estado sucesor por predicado que se aplique a una situación dada. Por lo tanto, se deben tener los equemas de axiomas siguientes para i diferente de j.
(CR_5) $\neg\exists s(cree_i(s) \wedge cree_j(s))$
(CR_6) $\neg\exists s(real(s) \wedge cree_j(s))$.

4 Evolución de creencias de varios agentes: operador modal

En [DP00] proponemos otra extensión a la evolución de creencias de varios agentes, la cual está basada en el uso de un operador modal.

En general, un agente puede tener cuatro actitudes correspondientes a sus creencias con respecto a una proposición, por ejemplo las cuatro actitudes de la madre con respecto al vaso roto son representadas formalmente por: $B_m(roto(x, s))$, $B_m(\neg roto(x, s))$, $\neg B_m(roto(x, s))$ y $\neg B_m(\neg roto(x, s))$.

Si consideramos que el hecho de observar y ver que el vaso está roto o de tirar el vaso implica que la madre crea que el vaso está roto o bien que no crea que el vaso no está roto, representado formalmente por:
$a = observar_m \wedge roto(x, s) \vee a = tirar_m(x) \vee a = tirar_h(x) \rightarrow B_m(roto(x, do(a, s)))$,
$a = observar_m \wedge roto(x, s) \vee a = tirar_m(x) \vee a = tirar_h(x) \rightarrow \neg B_m(\neg roto(x, do(a, s)))$.
Por otro lado, si el hecho de observar y ver que el vaso no está roto o de reparar el vaso implica que la madre no crea que el vaso está roto o bien que crea que el vaso no está roto, representado por:
$a = observar_m \wedge \neg roto(x, s) \vee a = reparar_m(x) \rightarrow \neg B_m(roto(x, do(a, s)))$,
$a = observar_m \wedge roto(x, s) \vee a = reparar_m(x) \rightarrow B_m(\neg roto(x, do(a, s)))$.

Se han definido las cuatro actitudes de la madre con respecto a la proposición *roto*. Si consideramos que estos hechos representan a "todas" las acciones que la madre conoce y que tienen como consecuencia que la madre crea que el vaso está roto, y que no existen hechos que la dejen en la ignorancia, entonces aplicando una metodología similar a la de Reiter, se pueden deducir los axiomas de estado sucesor de las creencias de la madre con respecto a $roto(x, s)$ y $\neg roto(x, s)$, representados por:
-$B_m(roto(x, s)) \leftrightarrow a = observar_m \wedge roto(x, s) \vee a = tirar_m(x) \vee a = tirar_h(x) \vee$
$B_m(roto(x, s)) \wedge \neg(a = observar_m \wedge \neg roto(x, s) \vee a = reparar_m(x))$
-$B_m(\neg roto(x, s)) \leftrightarrow a = observar_m \wedge \neg roto(x, s) \vee a = reparar_m(x) \vee B_m(\neg roto(x, s))$
$\wedge \neg(a = observar_m \wedge roto(x, s) \vee a = tirar_m(x) \vee a = tirar_h(x))$.

Se tiene como hipótesis que estos axiomas describen “todas” las condiciones que cambian el valor de verdad de las creencias de la madre con respecto al vaso roto. Para no deducir contradiciones, se deben imponer ciertas restricciones, ver [DP00]. Se puede mostrar, por inducción sobre las situaciones, que la consistencia en una situación dada sólo depende de la consistencia en la situación inicial.

5 Comparación entre los formalismos

El formalismo que utiliza la relación de accesibilidad posee un problema en la revisión de las creencias. Si no existe situación accesible por la relación de accesibilidad a partir de una situación que contiene como última acción, una acción de percepción, la consecuencia es que el agente tiene creencias contradictorias, evidentemente esto no es aceptable.

Consideremos el ejemplo de la seccion 2. Si en S_0 el agente cree que el vaso no está roto $Bel(\neg roto, S_0)$, tenemos que $\forall s'(B(s', s) \rightarrow \neg roto(s'))$ y como $roto(S_0)$, por lo tanto tenemos $\neg\exists s'(B(s', s) \wedge (roto(S_0) \leftrightarrow roto(s'))$, utilizando (S_B) no existe ninguna situación accesible desde $do(observar, S_0)$ por B. La consecuencia es que para cualquier fórmula ϕ tenemos $Bel(\phi, do(observar, S_0))$ y $Bel(\neg\phi, do(observar, S_0))$.

En [SPLL00] y [DP01], los autores han propuesto una modificación a la formalización, en la que no se tiene este problema. La idea es de asignar a cada situación s, un valor de posibilidad definida por la función $pl(s)$ y agregan condiciones suplementarias (ver el Teorema 3 en [SPLL00]).

En lo concerniente a la formalización que emplea el operador modal, estos problemas no se presentan, pues se debe describir explícitamente cómo evolucionan cada una de las cuatro actitudes, las leyes de creencia son independientes unas de otras. Por lo que la consistencia de las creencias, si existe en la situación inicial, es persistente.

La dependencia entre creencias que posee la formalización que utiliza la relación de accesibilidad no es muy realista. Por ejemplo, el hecho que un vaso esté *roto* no tiene porque influir, en principio, en el hecho que esté $frío$, sin embargo bajo esta formalización existe una relación implícita. Los axiomas de estado sucesor de creencias permiten tener independencia entre ambos hechos. Así por ejemplo, los axiomas que definen la creencia que el vaso está *roto* son independientes de las condiciones que definen la creencia que el vaso está $frío$.

En contraparte, la formalización que emplea el operador modal posee una fuerte limitante de expresión. Por ejemplo, supongamos que un agente escucha un ruido de ruptura de un vaso, y sabe que existen dos vasos, $X1$ y $X2$ pero no sabe cual de ellos fue roto, que puede ser descrito formalmente como $B(roto(X1, s) \vee roto(X2, s))$, pero que no puede ser representado bajo este formalismo.

6 Conclusión

El cálculo de situaciones permite formalizar una posible solución del frame problem, la cual se aplica a la evolución de las creencias de un agente en [SL93] y

[LL98]. Este formalismo supone dos hipótesis poco realistas para el caso de varios agentes: el agente conoce todas las acciones realizadas y sabe cómo evolucionan las cosas en la realidad.

Dos extensiones para el caso de varios agentes han sido propuestas, utilizando una relación de accesibilidad y utilizando un operador modal. Estos formalismos no estan restringuidos al uso de las dos hipótesis. Sin embargo la elección de uno u otro promueve la busqueda de un convenio entre el poder de expresión y los problemas propios al dinamismo. Así por ejemplo, el primer formalismo posee un poder de expresión ilimitado pero no puede garantizar la independencia entre creencias. Mientras que el segundo posee un poder de expresión limitado pero evita problemas delicados tales como el frame problem y la revisión de creencias.

Agradecimientos. Queremos agradecer a Yves Lespérance por sus comentarios y sugerencias que contribuyeron a mejorar la calidad de este trabajo.

Referencias

[Bro87] F. M. Brown, editor. *The Frame Problem in Artificial Intelligence*. Morgan Kaufmann Publishers, 1987.

[Che88] B. F. Chellas. *Modal Logic: An introduction*. Cambridge University Press, 1988.

[DP00] R. Demolombe and M. P. Pozos Parra. A simple and tractable extension of situation calculus to epistemic logic. In Z. W. Ras and S. Ohsuga, editors, *Proc. of 12th Internatioonal Symposium ISMIS 2000*. Springer. LNAI 1932, 2000.

[DP01] R. Demolombe and P. Pozos Parra. Formalisation de l'évolucion des croyances dans le Calcul des Situations. In *Proc Premières Journées Francophones Modèles Formels de l'Interaction (MFI01)*. 2001.

[LL98] G. Lakemeyer and H. Levesque. AOL: a logic of acting, sensing, knowing and only knowing. In *Proc. of the 6th Int. Conf. on Principles of Knowledge Representation and Reasoning*, pages 316–327, 1998.

[LLR99] Y. Lespérance, H. Levesque, and R. Reiter. A situation calculus approach to modeling and programming agents. In A. Rao and M. Wooldbridge, editors, *Foundations and Theories of Rational Agents*. Kluwer, 1999.

[Rei91] R. Reiter. The frame problem in the situation calculus: a simple solution (sometimes) and a completeness result for goal regression. In V. Lifschitz, editor, *Artificial Intelligence and Mathematical Theory of Computation: Papers in Honor of John McCarthy*, pages 359–380. Academic Press, 1991.

[Rei99] R. Reiter. Knowledge in Action: Logical Foundations for Describing and Implementing Dynamical Systems. Technical report, University of Toronto, 1999.

[SL93] R. Scherl and H. Levesque. The Frame Problem and Knowledge Producing Actions. In *Proc. of the National Conference of Artificial Intelligence*. AAAI Press, 1993.

[SPLL00] S. Shapiro, M. Pagnuco, Y. Lespérance, and H. Levesque. Iterated belief change in the situation calculus. In *Proc. of the 7th Conference on Principles on Knowledge Representation and Reasoning (KR2000)*. Morgan Kaufman Publishers, 2000.