Lógica Epistémica

Andrés Laurito

Primer cuatrimestre 2016

andy.laurito@hotmail.com

July 3, 2016

Lo que vamos a ver

1. Introducción al problema
   - Los papers
   - Lógica epistémica
2. El problema - Parte 1
   - Definiendo la semántica
   - Modelando la noción de poder de razonamiento
   - Estudiando la complejidad computacional
3. El problema - Parte 2
   - Redefiniendo la semántica
   - Bisimulación en la semántica de Neighbourhood
   - Estudiando la complejidad computacional

Qué papers elegí para presentar?

1. On the Complexity of Epistemic Reasoning
2. Which Semantics for Neighbourhood Semantics?

Idea de hacía donde vamos

Tratan el problema de la lógica epistémica. En el primero nos vamos a enfocar en la complejidad computacional de SAT para la lógica epistémica, mientras que en el otro vamos a usar esté estudio, para definir la complejidad computacional de la bisimulación en 2 modelos semánticos de Neighbourhood.

Qué es la lógica epistémica?

Wikipedia

La lógica epistémica es un campo de la lógica modal que se ocupa del razonamiento sobre el conocimiento. Está tiene aplicaciones en numerosos campos, tales como filosofía, ciencia computacional teórica, inteligencia artificial, economía y linguística.

Stanford Enciclopedy of Phylosophy

Epistemic logic is the logic of knowledge and belief. It provides insight into the properties of individual knowers, has provided a means to model complicated scenarios involving groups of knowers and has improved our understanding of the dynamics of inquiry.

Qué es la lógica epistémica?

Mi definición

Es la lógica de la representación del conocimiento y las creencias de un individuo (en I.A. un agente). A partir de la extensión de la lógica proposicional con operadores modales, podemos modelar de manera formal el poseer conocimientos y adquirirlos (en donde adquirir puede ser visto como una forma de razonamiento del individuo).

Super relacionada con Belief Revision (me parece bastante volátil el fin de una y el comienzo de otra).

Introduciendo a los nuevos operadores

En las lógicas epistémicas existen dos operadores modales. Si a representa un agente, escribimos:

- a conoce una fórmula ϕ como $K_a\phi$
- a cree una fórmula ϕ como $B_a\phi$

Para el problema que vamos a atacar, es indistinto el operador. En todo lo que sigue de la presentación , voy a usar el primer operador y lo voy a notar con la notación utilizada en toda la materia, pero debe etenderser que es indistinto usar cualquiera de los dos.

Queremos modelar a la epistemología, tenemos la sintaxis,
los operadores ... Qué modelo semántico usamos?

Usando el modelo de Kripke

Nos encontramos con un problema conocido como el "logical omniscience problem". Sabemos que:

Lema

Si ϕ y $(\phi \implies \psi)$ son válidas en un modelo, también lo es ψ.

Y recordando que K tiene la regla de necesitación que dice:

Necesitación

Si ϕ es válida en un modelo, entonces también lo es $\Box\phi$.

Llegamos a que la lógica modal K es muy fuerte para modelar conocimiento y creencia!.

Neighbourhood al rescate!

Vamos a definir a una estructura epistémica, (los famosos modelos de vecindad con otro sabor), como una tripla $M = (W, N, I)$ en donde, si A es un conjunto de agentes, y a $\in$ A entonces:

- W es un conjunto de mundos
- $N : AxW \mapsto 2^{2^W}$ es la función que asigna a cada agente en un mundo, el correspondiente conjunto de proposiciones que conoce (es decir, un conjunto epistémico).
- $I : P \mapsto 2^W$ es la función que asigna a cada proposición atómica, el conjunto de mundos en donde dicha proposición es satisfecha.

La definición de satisfacibilidad de una fórmula será la misma que vimos en la materia.

Problema con Neighbourhood

Si bien con el modelo definido recién, solucionamos el problema de logical omniscience, nos surge un nuevo problema. Citando la diapo 6 de la tecera clase:

Teorema

Si $\phi \iff \psi$ es válida en una clase de modelos de vecindad, entonces en dicha clase también lo es $\Box\phi \iff \Box\psi$.

Nos vamos a permitir vivir con este problema

El razonamiento como fórmulas

1. $\neg\Box \mathit{false}$
2. $\Box \mathit{true}$
3. $\Box(p \wedge q) \implies \Box q$
4. $(\Box p \wedge \Box q) \implies \Box(p \wedge q)$
5. $\Box p \implies \Box\Box p$
6. $\neg\Box p \implies \Box\neg\Box p$
7. $\Box p \implies p$

El razonamiento modelado sobre conjuntos epistémicos!

1. $\emptyset \notin N(a, w)$
2. $W \in N(a, w)$
3. Si $U \in N(a, w)$ y $U \subseteq V, \implies V \in N(a, w)$
4. Si $U \in N(a, w)$ y $V \in N(a, w) \implies (U \cap V) \in N(a, w)$
5. Si $U \in N(a, w) \implies \{u : U \in N(a, u)\} \in N(a, w)$
6. Si $U \notin N(a, w) \implies \{u : U \notin N(a, u)\} \in N(a, w)$
7. Si $U \in N(a, w) \implies w \in U$

A partir de está definición, si llamamos ϵ a la clase de todos los modelos epistémicos, entonces podemos decir:

Definición

Sea S un subconjunto de $\{1, ..., 7\}$, entonces ϵ_S será la clase de los modelos epistémicos que satisfaga $C_j \; \forall j \in S$

Complejidad computacional de cada modelo ?

Sabemos que una fórmula ϕ es ϵ_S-satisfacible si $\exists w \in W$ tal que $\epsilon_S \models_w \phi$. Cúal será la relación, si es que la hay, entre S y la complejidad computacional de satisfacer una fórmula?

EL problema a resolver en este paper

EL Teorema

Si S es un subconjunto de 1, ..., 7 entonces resolver SAT está en PSPACE.
Ahora, si S es un subconjunto de 1, ..., 7 y 4 $\notin S$, entonces SAT está en NP

El paper de "On the Complexity of Epistemic Reasoning" demuestra este teorema (todo el paper está enfocado en esto).

La idea para resolver el problema

El objetivo para demostrar el teorema anterior viene por este lado

La idea

Sea M uno de los modelos epistémicos ϵ_i con $i \in S$, ϕ una fórmula tal que $\epsilon_i \models \phi$. La idea será probar que siempre que $i \neq 4$ existe una valuación v para ϕ tal que las subformulás con operadores modales en ϕ son satisfacibles, si ψ es satisfacible, con ψ una fórmula perteneciente al calcúlo proposicional.

Con esto podemos aplicar el algoritmo de tableau de una manera más eficiente. Además, al dar propiedades de está pinta, nos podemos construir un algoritmo no determinístico de tiempo polinomial.
Veamos la propiedad enunciada más importante del paper:

SAT de una fórmula

Proposición 1

Una fórmula ϕ es ϵ-satisfacible *iff* $\exists$ v valuación para ϕ tal que si $\Box\psi_1$ y $\Box\psi_2 \in sub(\phi)$, $v(\Box\psi_1) = 1$, y $v(\Box\psi_2) = 0 \implies (\psi_1 \wedge \psi_2) \vee (\neg\psi_1 \wedge \psi_2)$ es ϵ-satisfacible

A partir de está propiedad, podemos inferir el siguiente algoritmo:

Algoritmo SAT para ϵ

- De manera no deterministísca adivinar una valuación v para ϕ
- $\forall \Box\psi_1 \wedge \Box\psi_2 \in sub(\phi)$ / $v(\Box\psi_1) = 1 \wedge v(\Box\psi_2) = 0$ no determinísticamente y de manera recursiva, verificar que $(\psi_1 \wedge \neg\psi_2)$ ó $(\neg\psi_1 \wedge \psi_2)$ es satisfacible.

Estamos recortando un monton de ramas del algoritmo de Tableau !!. (Otra propiedad muy importante que enuncia el paper)

Propiedas sobre los modelos de razonamiento

A partir de la proposición 1, podemos dar propiedades sobre cada ϵ_i, y obtener el algoritmo de manera similar a como lo hicimos recién. Veamos algunos ejemplos:

Proposición 2

Una fórmula ϕ es ϵ_2-satisfacible *iff* $\exists$ v valuación para ϕ tal que:

- Si $\Box\psi \in sub(\phi)$ y $v(\Box\psi) = 0 \implies \neg\psi$ es ϵ_2−satisfacible, y
- $\Box\psi_1$ y $\Box\psi_2 \in sub(\phi), v(\Box\psi_1) = 1$, y $v(\Box\psi_2) = 0 \implies (\psi_1 \wedge \psi_2) \vee (\neg\psi_1 \wedge \psi_2)$ es ϵ_2-satisfacible

Proposición 3

Una fórmula ϕ es ϵ_3-satisfacible *iff* $\exists$ v valuación para ϕ tal que:

- $\Box\psi_1$ y $\Box\psi_2 \in sub(\phi), v(\Box\psi_1) = 1$, y $v(\Box\psi_2) = 0 \implies (\psi_1 \wedge \neg\psi_2)$ es ϵ_3-satisfacible

Un momento...

Antes había mencionado que si el agente podía razonar: Si $U \in N(a, w)$ y $V \in N(a, w) \implies (U \cap V) \in N(a, w)$ (poder inferir conocimiento), el problema de SAT se volvía PSPACE. Por qué pasa esto?

Lema

Sea $M = (W, N, I) \in \epsilon_4$, sea ϕ una fórmula y $w \in W$. Entonces $M \not\models_w \Box\phi \iff \cap_{j=1}^{k} U_j \neq I(\phi)$ para todo subconjunto no vacío $U_1, \ldots, U_k \subseteq N(a, w)$

Y entonces la propiedad que caracteriza a ϵ_4 es la siguiente:

Proposición 4

Una fórmula ϕ es ϵ_4-satisfacible *iff* $\exists$ v valuación para ϕ tal que:

- $\Box\psi_1, \ldots, \Box\psi_k \in sub(\phi), v(\Box\psi_j) = 1 \; \forall 1 \leq j \leq k \wedge v(\Box\psi_k) = 0$ $\implies \wedge_{j=1}^{k-1} \psi_j \wedge \neg\psi_k$ es ϵ_4-satisfacible ó $\neg\psi_j \wedge \psi_k$ es ϵ_4-satisfacible para algún j, $1 \leq j \leq k$

Repasando lo que tenemos

En toda está primera parte, al igual que en la materia, decimos que:

Definición

Sea M un modelo epistémico, $w \in W$,
$M \models_w \Box\phi \iff \{v/v \models \phi\} \in N(a, w)$.

Vamos a decir lo mismo de otra forma:

Reescribiendo la definición

Sea M un modelo epistémico, $w \in W$,
$M \models_w \Box\phi \iff \{v/v \models \phi\} \in N(a, w) = \{X_1, ..., X_n\} \iff \exists Y \in N(a, w)/Y = \{v/v \models \phi\} \iff \exists Y \in N(a, w)/Y = ||\phi||_M$

Esto puede leerse como : Un agente a cree una fórmula ϕ, desde un estado o mundo w $\iff$ el conjunto de TODOS los mundos en donde ϕ es cierta pertenece a N(a,w)... No será mucho pedir todos?

Nuevo enfoque

Cambio la notación por una cuestión de practicidad. Vamos a representar al operador modal que conocemos, el $\Box$ como [], y vamos a definir una nueva semántica.

Semántica fuerte

Vamos a decir que $||[=]\phi||_M = \{w | \exists X \in N(a, w) \text{ tal que } X = ||\phi||_M\}$

Semántica debil

Vamos a decir que $||[\subseteq]\phi||_M = \{w | \exists X \in N(a, w) \text{ tal que } X \subseteq ||\phi||_M\}$

Estas semánticas son conocidas hoy en día para la lógica epistémica, e incluso ambas poseen una noción de bisimulación. El objetivo del segundo paper, "Which Semantics for Neighbourhood Semantics?", tratará el problema de definir un nuevo concepto de bisimulación para la primer semántica, y mostrará la complejidad computacional usando fuertemente las propiedades vistas del primer paper.

Definiendo bisimulación

$N_{\subseteq}$-bisimulación

Sean M y M' dos modelos epistémicos. Una $N_{\subseteq}$-bisimulación entre M y M' es una relación no vacía $Z \subseteq WxW'$ tal que si xZx' entonces:

- Prop: $\forall p \in PROP$, $x \in ||p|||_M \iff x' \in ||p||_{M'}$
- Zig: $\forall T \in N_x$, $\exists T' \in N'_x talque \forall w' \in T' \exists w \in T$ tal que wZw'
- Zag: $\forall T' \in N'_x$, $\exists T \in N_x talque \forall w \in T \exists w' \in T'$ tal que wZw'

Es muy parecida a la noción de bisimulación que vimos en la materia. Veamos ahora como definimos la noción de bisimulación para la otra semántica.

Definiendo bisimulación

$N_=$-bisimulación

Sean M y $M^{'}$ dos modelos epistémicos. Una $N_{\subseteq}$-bisimulación entre M y $M^{'}$ es una relación no vacía $Z \subseteq WxW^{'}$ tal que si $xZx^{'}$ entonces:

- Prop: $\forall p \in PROP$, $x \in ||p|||_M \iff x^{'} \in ||p||_{M'}$
- Zig: $\forall T \in N_x$, $\exists T^{'} \in N^{'}_x$ tal que
  - $\forall w^{'} \in T^{'} \exists w \in T$ tal que $wZw^{'}$
  - $\forall w^{'} \in W^{'} \setminus T^{'} \exists w \in W \setminus T$ tal que $wZw^{'}$
- Zig: $\forall T^{'} \in N^{'}_x$, $\exists T \in N_x$ tal que
  - $\forall w \in T\ \exists w^{'} \in T^{'}$ tal que $wZw^{'}$
  - $\forall w \in W \setminus T\ \exists w^{'} \in W^{'} \setminus T^{'}$ tal que $wZw^{'}$

El problema con está definición es que es muy fuerte. La bisimulación preserva satisfacibilidad de las fórmulas de $N_=$, pero límita el poder de expresividad de $N_=$

Redefiniendo $N_=$-bisimulación

Definición de diferenciación

Se define la clase de modelos epistémicos diferenciales D como D $= \{M | \forall w \in W \; \forall T \in N_w \; \exists \phi$ tal que $||\phi^M|| = T\}$

Lo importante es notar que para cada conjunto de mundos en la vecindad, existe una fórmula que caracteriza a dicho conjunto.
A partir de está definición, podemos volver a definir el concepto de $N_=$-bisimulación como sigue

Redefiniendo $N_{=}$-bisimulación

$N_{=}$-bisimulación

Sean M y $M^{'}$ dos modelos epistémicos. Una $N_{\subseteq}$-bisimulación entre M y $M^{'}$ es una relación no vacía $Z \subseteq WxW^{'}$ tal que si $xZx^{'}$ entonces:

- Prop: $\forall p \in PROP,\ x \in ||p|||_M \iff x^{'} \in ||p||_{M'}$
- Zig: $\forall T \in N_x,\ \exists T^{'} \in N^{'}_x$ tal que
  - $\forall w^{'} \in T^{'} \exists w \in T$ tal que $wZw^{'}$
  - $\forall X^{'} \notin N^{'}_{w^{'}}$ diferenciable en $M^{'}$ tal que $T^{'} \subset X^{'} \subseteq W^{'}$, existen $w^{'} \in X^{'} \setminus T^{'}$, $w \in W \setminus T$ tal que $wZw^{'}$
- Zig: $\forall T^{'} \in N^{'}_x,\ \exists T \in N_x$ tal que
  - $\forall w \in T\ \exists w^{'} \in T^{'}$ tal que $wZw^{'}$
  - $\forall X \notin N_w$ diferenciable en M tal que $T \subset X \subseteq W$, existen $w \in X \setminus T$, $w^{'} \in W^{'} \setminus T^{'}$ tal que $wZw^{'}$

Por que pasa que la definición se vuelve tan complicada ?.

Extendiendo a $N_=$

El aporte importante que hace el paper, es resolver a la pregunta anterior. Definimos al operador $\bigtriangledown$ como sigue:

Definición nuevo operador $\bigtriangledown$

$||\phi \bigtriangledown \psi||^M = \{w | \exists X_1, X_2 \in N_w$ tal que
$X_1 \neq X_2 \wedge X_1 = ||\phi||_M \wedge X_2 = ||\phi \vee \psi||_M\}$

Con este operador, podemos probar que que la lógica $N_=(\bigtriangledown)$ (es la lógica que obtenemos al extender $N_=$ con el operador $\bigtriangledown$), matchea el poder expresivo de las $N_=$-bisimulaciones.
El problema con este nuevo operador, es que es alto galerazo. No habrá algo más natural?

El operador modal de existencia

Definimos al operador modal de existencia E como sigue:

Definición operador modal de existencia

$$||E\phi|| = \begin{cases} W & si||\phi||_M \neq \emptyset \\ \emptyset & sino \end{cases}$$

Se puede probar que vale la siguiente igualdad
$||\phi \bigtriangledown \psi||^M = ||[=]\phi \wedge [=](\phi \vee \psi) \wedge E(\psi \wedge \neg\phi)||^M$
Por lo que a partir de esto, podemos inferir que la extensión $N_=(E)$ matchea el poder expresivo de las $N_=$-bisimulaciones.
Surge la siguiente pregunta... Habremos modificado la complejidad computacional, al haber extendido de está manera la bisimulación?

Satisfacibilidad de las semánticas definidas

Proposición

$\forall \phi$ y $\forall$ M modelo epistémico, $||\phi||_{\subseteq}^{M} = ||\phi||_{=}^{M^+}$

Usando esto y lo visto en la primera parte, podemos probar entonces que

Proposición

El problema de satisfacibilidad de $N_{\subseteq}$ es NP

Falta ver que $N_{=}(E)$ sigue estando en NP-completo. Para ver esto, primero demos el algoritmo

Algoritmo satisfaciblidad $N_{=}(E)$

Algorithm: $\mathsf{sat}(\varphi)$

```
guess a universal valuation μ for φ
R ← witness^φ_μ(φ, ⊥, ∅)
for all Eψ ∈ sub(φ) s.t. μ(Eψ) = 1 do
    if there is no ν ∈ R where ν(ψ) = 1 then
        R ← witness^φ_μ(ψ, ⊥, R)
    end if
end for
return R
```

Algorithm: $\mathsf{witness}^{\varphi}_{\mu}(\psi_1, \psi_2, R)$

```
guess a valuation ν for φ compatible with μ s.t.: ν(ψ1) = 1 and
ν(ψ2) = 0, or ν(ψ1) = 0 and ν(ψ2) = 1
if there is not such a valuation then
    print UNSAT and exit
end if
R ← R ∪ {ν}
for all [=]η1, [=]η2 ∈ sub(φ) s.t. ν([=]η1) = 1 and
ν([=]η2) = 0 do
    if there is no ν' ∈ R such that ν'(η1) = 1 and ν'(η2) = 0,
    or ν'(η1) = 0 and ν'(η2) = 1 then
        R ← witness^φ_μ(η1, η2, R)
    end if
```

Preguntas??