Sistemas Multi-Agentes

Teorias de Agentes

Prof. Dr. Jaime Simão Sichman
Universidade de São Paulo
Depto. Eng. Computação e Sistemas Digitais
jaime@pcs.usp.br
http://www.pcs.usp.br/~jaime

Para que teorias?

[Wooldridge e Jennings 95]

- métodos formais têm pouco impacto no desenvolvimento de software
- mas fornecem uma *semântica* (significado) para as arquiteturas e linguagens
- sem um significado, não sabemos nunca claramente *o que* ocorre e *porque* ocorre
- usuários finais nunca lêem ou entendem tal semântica, mas ela é *necessária* para desenvolver ou aperfeiçoar as linguagens

Para que teorias?

[Wooldridge e Jennings 95]

- sistemas baseados em agentes utilizam uma grande gama de conceitos teóricos e ferramentas de abstração
- tais elementos são *facilmente compreendidos* (por analogia ou metáforas) e têm obviamente um grande potencial de utilização
- entretanto necessitamos de uma teoria para obter um conhecimento mais *profundo* destas ferramentas

Sistemas Intencionais

[Wooldridge e Jennings 95]

- Por onde começar?
- Ao explicar as atividades humanas, quase sempre utilizamos frases como:
  - Anna levou o guarda-chuva porque ela *acreditava* que iria chover
  - Jaime trabalhou duro porque ele *desejava* obter o seu Doutorado
  - Túlio *pretende* enviar o paper pro Congresso

Sistemas Intencionais

[Wooldridge e Jennings 95]

- Estas frases utilizam alguns conceitos de "*folk psychology*", pelos quais o comportamento humano é explicado pela atribuição de *atitudes*, como crer, querer, esperar, temer, desejar etc.
- As atitudes empregadas nestas descrições são denominadas *noções intencionais*

Sistemas Intencionais

[Wooldridge e Jennings 95]

- O filósofo Daniel Dennett utilizou o termo *sistema intencional* para descrever "entidades cujo comportamento pode ser predito através da atribuição de crenças, desejos e astúcia racional."
- Identificou também *graus* de intencionalidade:
  - um sistema intencional de 1a. ordem tem crenças, desejos, mas não crenças e desejos *sobre* crenças e desejos

Sistemas Intencionais

[Wooldridge e Jennings 95]

- um sistema intencional de 2a. ordem têm crenças, desejos (e outros estados intencionais) sobre crenças, desejos (e outros estados intencionais), tanto em relação a si próprio como em relação aos outros.

- Mas será que é legítimo ou útil atribuir-se crenças, desejos a sistemas computacionais?
- McCarthy sugeriu que em certas ocasiões, utlizar uma "*postura intencional*" é apropriada:

Sistemas Intencionais

[Wooldridge e Jennings 95]

- Atribuir *crenças, vontade própria, intenções, consciência, habilidades ou desejos* a uma máquina pode ser / é:
  - *legítimo*: expressa o sentido usual
  - *útil*: ajuda a entender, reparar, melhorar
  - *logicamente necessário:* quase nunca
  - *mais fácil:* máquinas de estrutura conhecida
  - *mais útil:* entidades de estrutura desconhecida

Sistemas Intencionais

[Wooldridge e Jennings 95]

- Que objetos podem ser descritos sob uma postura intencional?
- Todos ou quase todos
  - Ex: interruptor de luz [Shoham 91]
  - Entretanto, às vezes é inutil, pois uma versão *mecanicista* é mais simples e informativa
- À medida que sistemas se tornam mais complexos, precisamos de abstrações e metáforas mais poderosas para entendê-los/explicá-los.

Sistemas Intencionais

[Wooldridge e Jennings 95]

- Sistemas intencionais são portanto *ferramentas abstratas*, que nos fornecem um modo conveniente de descrever, explicar e predizer o comportamento de sistemas complexos
- Tal fato não é novo em computação: abstração procedural, tipos abstratos de dados, objetos
- Agentes como sistemas intencionais representam um passo além nesta hierarquia de abstrações
- Teorias de agentes são teorias que usam posturas intencionais para descrever seu comportamento

Atitudes Mentais

[Wooldridge e Jennings 95]

- Queremos portanto criar sistemas computacionais utilizando noções mentais
- Que atitudes devemos utilizar?
  - atitudes informativas
    - » *crença, conhecimento*
  - pró-atitudes
    - » *desejo, intenção, obrigação, engajamento ...*

Atitudes Mentais

[Wooldridge e Jennings 95]

- Atitudes informativas são relacionadas à informação que um agente têm sobre o mundo
- Pró-atitudes são aquelas que de alguma forma "guiam" a ação dos agentes
- Qual combinação deve ser adotada?
  - existem várias combinações possíveis

Atitudes Mentais

[Wooldridge e Jennings 95]

- Cohen e Levesque
  - crenças e objetivos
- Rao e Georgeff
  - crenças, desejos e intenções
- Outros formalismos
  - Singh, Shoham, Kraus, Wooldridge e Jennings, Castelfranchi ...

Formalização

[Gasser, Rosenschein e Ephrati 95]

- Lógicas de Conhecimento e Crença
  - Para que agentes possam raciocinar sobre seu próprio conhecimento ou sobre o conhecimento de outros agentes, precisamos utilizar formalismos para descrever o que agentes *conhecem* e *crêem.*

Formalização

[Gasser, Rosenschein e Ephrati 95]

- Para representar que um agente A *crê* numa proposição ϕ, usamos o operador modal B:
  - B (ϕ, Father(John,Bill))
- Para representar que um agente A *conhece* uma proposição ϕ, usamos o operador modal K:
  - K(ϕ, Father(John,Bill))
- Proposições conhecidas são assumidas verdadeiras

Semântica Sentencial

[Gasser, Rosenschein e Ephrati 95]

- Todas as fórmulas ordinárias da lógica de 1a. ordem são fórmulas da lógica
- Se ϕ é uma fórmula lógica ordinária, e se α é um termo fechado, então B(α,ϕ) é uma fórmula legal (chamada de *átomo de crença*)
- Pode-se usar os conectivos lógicos usuais como *and*, *or*, *not*, etc.

Semântica Sentencial

[Gasser, Rosenschein e Ephrati 95]

Qual a semântica dos átomos de crença?

- Não desejamos que o valor verdade de uma crença seja o mesmo quando termos equivalentes são substituídos:

  B(α,Hungry(John)) não implica que

  B(α,Hungry(Pai-de(Bill))), mesmo que John seja o pai de Bill
- *referencialmente opaco* / referenc.transparente

Semântica Sentencial

[Gasser, Rosenschein e Ephrati 95]

Para definir a semântica de B, estendemos o nosso domínio do seguinte modo:

- um conjunto de agentes
- cada agente α tem um conjunto de crenças básicas Δ_α , composto por um conjunto de fórmulas ordinárias, e um conjunto de regras de inferência ρ_α
- Uma teoria de agente é composta por todas as fórmulas possíveis de serem inferidas, utilizando-se estas regras de inferência

Semântica Sentencial

[Gasser, Rosenschein e Ephrati 95]

- Deste modo, tudo o que um agente *conhece* é determinado pelas suas regras de inferência
- Estas regras de inferências podem limitar o número de deduções que um agente realiza, ou os tipos de deduções feitas
- Semântica é referencialmente opaca

Semântica Sentencial

[Gasser, Rosenschein e Ephrati 95]

- B(α, ϕ) tem o valor verdade verdadeiro sse ϕ pertence à teoria associada com o agente denotado por α.
- Método de prova:
  - partindo de B(α,ϕ) quero concluir que B(α,ψ)
  - assumimos que temos um modelo de prova do método de dedução do agente
- Konolidge, "Deduction Model of Belief"

Semântica de Mundos Possíveis

[Gasser, Rosenschein e Ephrati 95]

- Uma conceitualização alternativa inclui objetos w_0, w_1, w_2, …, w_i, … denominados *mundos possíveis*
- A sintaxe da linguagem é basicamente a mesma apresentada para o caso de lógica sentencial, ou seja, fórmulas ordinárias do cálculo de predicados de 1a. ordem mais um operador modal K

Semântica de Mundos Possíveis

[Gasser, Rosenschein e Ephrati 95]

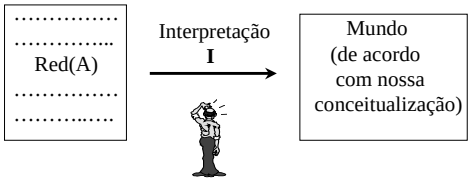
- A semântica é definida do seguinte modo:
  - Uma determinada fórmula bem formada (wff) não é *verdadeira* ou *falsa*, mas sim verdadeira ou falsa num *determinado mundo possível*
  - Ao invés de uma única interpretação (com objetos, funções, relações), temos agora uma interpretação para cada *mundo possível*

Semântica de Mundos Possíveis

[Gasser, Rosenschein e Ephrati 95]

Numa semântica descritiva:

Símbolos e Sentenças

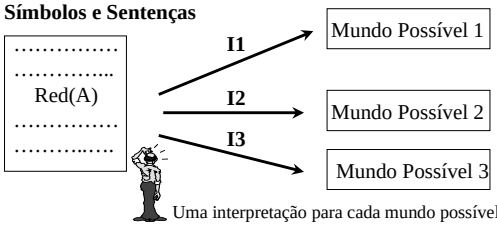


Semântica de Mundos Possíveis

[Gasser, Rosenschein e Ephrati 95]

Numa semântica de mundos possíveis:

Símbolos e Sentenças



Uma interpretação para cada mundo possível

Semântica de Mundos Possíveis

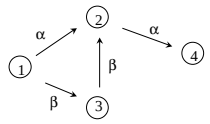
[Gasser, Rosenschein e Ephrati 95]

- Na semântica de mundos possíveis, uma fórmula ordinária bem formada (wff) tem um valor verdade em relação ao mundo w_i quando este valor é calculado usando a função de interpretação associada a w_i
- Portanto, uma fórmula ϕ pode ser verdadeira num mundo w_0 e ser falsa num mundo w_1

Semântica de Mundos Possíveis

[Gasser, Rosenschein e Ephrati 95]

- Como tais mundos são relacionados?
  - existe uma *relação de acessibilidade* $K(\alpha,w_i,w_j)$
  - quando esta relação é satisfeita, dizemos que w_j é acessível a partir de w_i para o agente α



Semântica de Mundos Possíveis

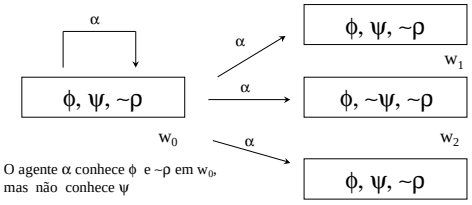
[Gasser, Rosenschein e Ephrati 95]

- Um átomo de conhecimento $B(\alpha, \phi)$ tem o valor verdade verdadeiro em relação ao mundo w_i sse ϕ tem o valor verdade verdadeiro em *todos* os mundos possíveis acessíveis a partir de w_i para o agente α.
- Outras propriedades (crença) podem ser definidas usando outras definições para o valor verdade

Semântica de Mundos Possíveis

[Gasser, Rosenschein e Ephrati 95]

Exemplo de relação de acessibilidade:

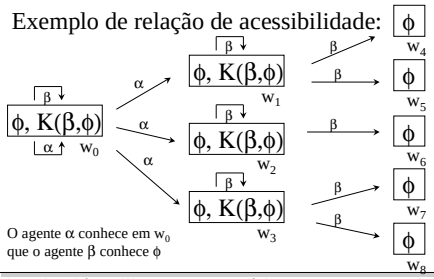


O agente α conhece φ e ~ρ em w_0, mas não conhece ψ

Semântica de Mundos Possíveis

[Gasser, Rosenschein e Ephrati 95]

Exemplo de relação de acessibilidade:



O agente α conhece em w_0 que o agente β conhece φ

Semântica de Mundos Possíveis

[Gasser, Rosenschein e Ephrati 95]

- a relação de acessibilidade divide os mundos possíveis em aqueles que são acessíveis e os que não são acessíveis, a partir de uma dado mundo
- porém esta relação nos dá mais do que uma simples partição de conjuntos acessíveis ou não
- *propriedades de conhecimento:* várias propriedades intuitivas sobre conhecimento de um agente podem ser capturadas usando propriedades destas relações, como reflexividade e transitividade

Semântica de Mundos Possíveis

[Gasser, Rosenschein e Ephrati 95]

Omnisciência Lógica

- De $\phi \rightarrow \psi$ e de $K(\alpha,\phi)$
  infiro $K(\alpha,\psi)$
- Isso significa que um agente modelizado pela lógica de mundos possíveis conhece todas as consequências lógicas de seu conhecimento (hipótese não realista!)

Problemas Abertos

[Wooldridge e Jennings 95]

- Maior parte dos esforços dedicados à *interrelação* entre as atitudes e como estas levam um agente a agir
  - ausência de teoria definitiva, unificada
  - na maioria, usa-se semântica dos mundos possíveis, que apesar de elegante é não realista
  - ausência de relação entre a teoria e o sistema computacional que desejamos especificar usando a teoria, estas tornam-se abstratas

Bibliografia

Drogoul, A. and Ferber, J. (1994). Multi-agent simulalation as a tool for studying emergent processes in societies. In: Gilbert, N. and Doran, J. eds., *Simulating Societies*. London, UCL Press.

Werner, E. and Demazeau, Y. (1992). *Decentralized AI III*, North Holland, 1992.