

TOMLG

Lucas Simões Matheus Ferraz

UTFPR

November 30, 2018

Sumário

- 1 Introdução
- 2 ToM
- 3 Lorini
- 4 Reinforcement Learning
- 5 Instan. of the model
- 6 Experimentos
- 7 Resultados
- 8 Conclusões

Pergunta

Como encarar um problema cognitivo de teoria da mente como um problema computacional?

Theory of mind

- O que é Theory of Mind?
- Experimento Sally e Anne.

Theory of mind

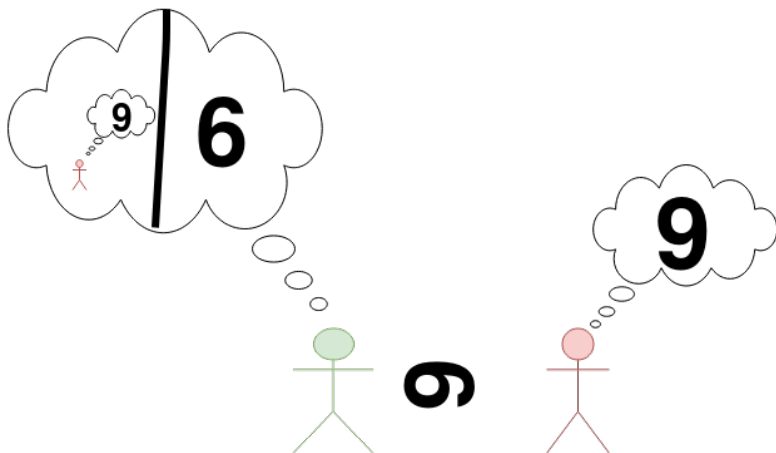


Figure 1: Theory of mind.

Experimento Sally e Anne

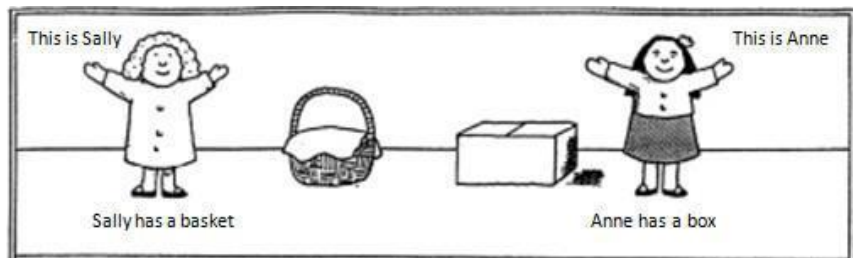


Figure 2: Experimento Sally e Anne 1.[1]

Experimento Sally e Anne

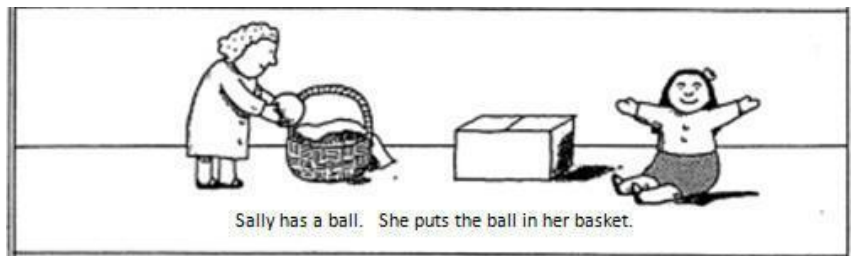


Figure 3: Experimento Sally e Anne 2.[1]

Experimento Sally e Anne

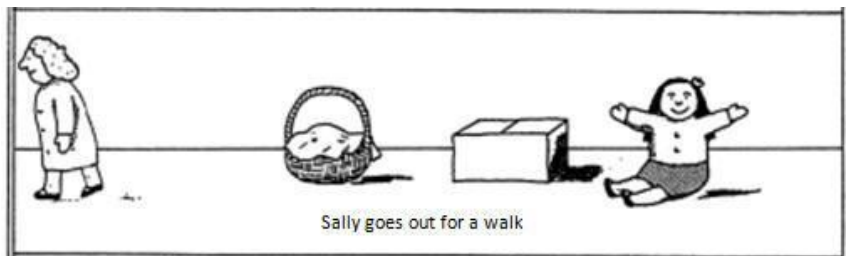


Figure 4: Experimento Sally e Anne 3.[1]

Experimento Sally e Anne



Figure 5: Experimento Sally e Anne 4.[1]

Experimento Sally e Anne

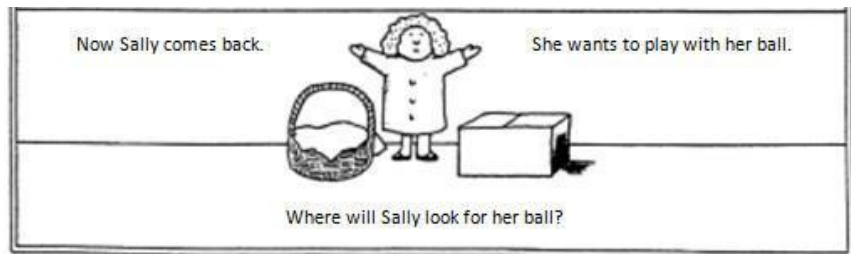


Figure 6: Experimento Sally e Anne 5.[1]

Objetivos

- Propor um modelo de teoria da mente, segundo o modelo cognitivo lógico desenvolvido por Lorini;
- Extrair a linha de raciocínio gerada pelo agente na lógica formalizada;
- Demonstrar como o agente pode, através de teoria da mente, inferir as ações futuras de outro agente.

Lorini

Define uma lógica multi-modal chamada \mathcal{LIA} . Permitindo:

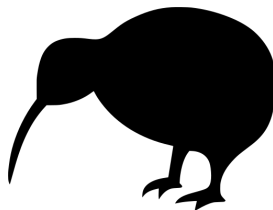
- Lidar com noção de **tentativa**, dentro de um *framework* que permite raciocinar sobre crenças, objetivos, intenções e ações básicas.
- Lidar com a geração de intenções instrumentais e entender sobre que condições um agente decide buscar um certo plano como uma forma de atingir um objetivo superior.
- Lidar com a persistência de intenções, permitindo entender sobre que condições uma intenção de executar uma certa ação pode ser abandonada.

Conceitos

- Objetivo
- Crença
- Ações
- Tentativa
- Surpresa

Crença

- Expressa ideias em que o agente acredita
- Nem sempre são fatos
 - $Belief_{Lucas}(TitiVoa)$
 - $Belief_{Matheus}(TitiNaoVoa)$



Ação

- Atuação do agente no ambiente
- Ação simples: Qualquer ação que conste no repertório do agente, e que ele consiga executar imediatamente.

Objetivo

- É um estado mental motivacional. É o que incentiva o agente a buscar algo.
- Intenção: objetivo com conteúdo acionário
 - $Goal_i(Passar_{TCC})$
- Achievement goal:
 - $AGoal_i(SeFormar)$
- Future directed Intention:
 - $FDI_i(Apresentar_{tcc}, SeFormar)$

Intenção

- Uma intenção é definida como um objetivo com conteúdo acionário. As intenções são divididas em proximais e distais.
- As intenções proximais são responsáveis por exercer controle do agente no presente.
- As intenções distais são responsáveis por guiar o comportamento e é utilizado no raciocínio do agente para geração de planos.

Tentativa

A expressão o "agente i intenciona fazer α pode ter dois significados":

- O agente i intenciona fazer α com sucesso;
- O agente i intenciona tentar fazer α .

Uma tentativa de fazer α é sempre iniciado por uma intenção proximal de tentar fazer α com sucesso ou uma intenção proximal de tentar fazer α .

- $Goal_I \ll I : pickupPassenger \gg T$

Surpresa

- Ocorre quando um evento não esperado pelo agente ocorre.
- Precursor de revisão de crenças

Reinforcement Learning

Uma técnica de Inteligência Artificial em que o aprendizado de agente acontece por reforço. Ou seja, pelas recompensas de suas ações.

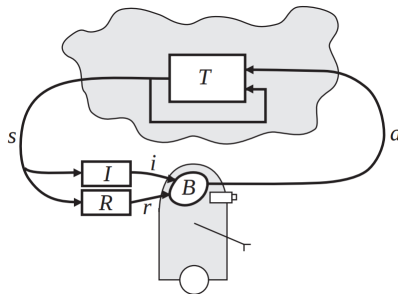


Figure 7: Modelo padrão de RL, Kaelbling et al., 1996.

KWIK

Know what it knows (KWIK) é um tipo de algoritmo que realiza apenas predições exatas, retornando "eu não sei" (\perp) quando não consegue prever o resultado de um evento. O número de vezes em que o algoritmo retorna \perp deve ser limitado.

Rmax

O algoritmo R_{max} é um KWIK que opera atribuindo a máxima recompensa possível para cada estado desconhecido. Ou seja, o algoritmo tem um viés otimista sob incerteza.

DOORMAX - Deterministic Object-Oriented Rmax

DOORMAX [2] é um algoritmo baseado na estrutura da família R_{MAX} , projetado para Processo de decisão de Markov orientado à objetos (OOMDP - do inglês *Object-Oriented Markov Decision Process*) proposicional determinístico.

- Algoritmo gera hipóteses através do que ele percebe.
- Aprendizado de Efeitos
- Aprendizado de Condições dos Efeitos

- Efeitos → mudanças que ocorrem no mundo quando um agente executa uma ação cujas precondições estão satisfeitas.
- Cada ação pode resultar em múltiplos efeitos.
- Precondição → São as condições necessárias para um evento acontecer quando uma determinada ação é realizada.

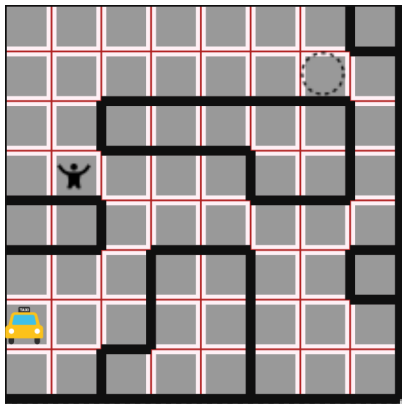
Operadores Perceptuais - retornam percepções sobre o ambiente onde o agente está inserido

- *touch* → informações referentes a casas adjacentes;
- *on* → informações referentes a casa atual do agente;
- *see* → informações referentes a demais casas do tabuleiro;

Ambiente de Testes

O ambiente escolhido para realizar os experimentos é o domínio *Taxi*. Este domínio consiste em um mundo de grid 2D, onde um agente (táxi) tem que levar um passageiro para um destino. Tanto o táxi quanto o passageiro e o destino ocupam uma posição na grid.

Ambiente de Testes



Ciclo de raciocínio - Geral

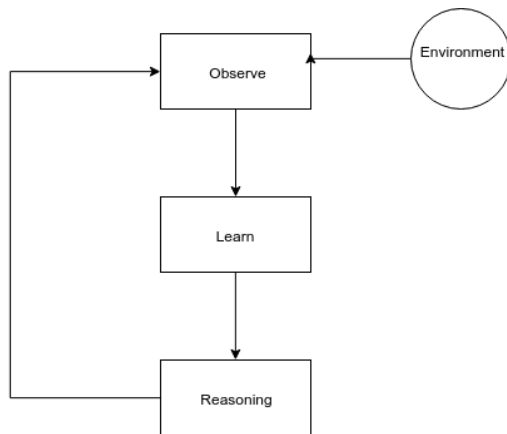


Figure 8: Arquitetura geral de um agente

Ciclo de raciocínio - Agente agindo no mundo

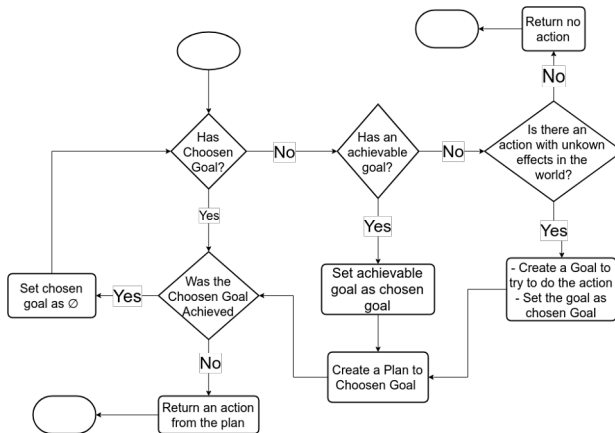


Figure 9: Ciclo de raciocínio de um agente agindo no mundo.

Ciclo de raciocínio - ToM

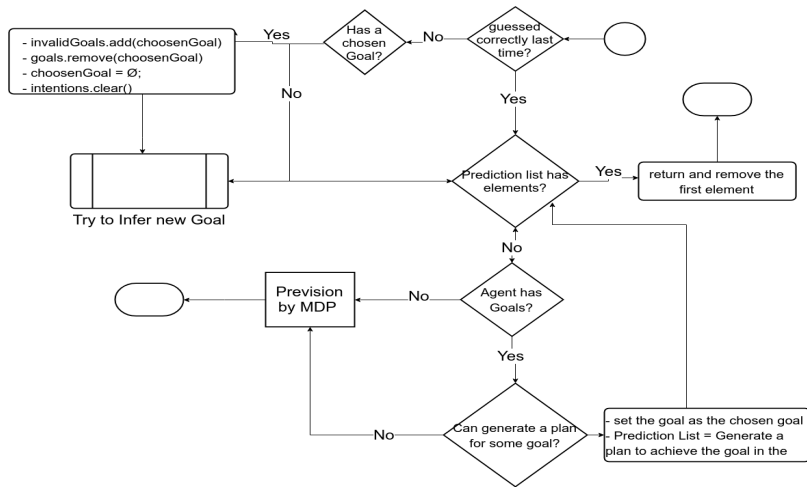


Figure 10: Ciclo de raciocínio ToM

Experimentos

- O agente é capaz de agir e aprender sobre o mundo ao seu redor?
- O agente é capaz de aprender sobre outros agentes, e prever suas ações?
- O agente é capaz de revisar suas próprias crenças, quando observando outros agentes?

Experimentos

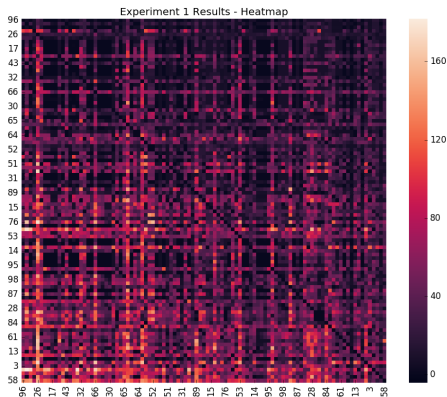
- 1. O agente é inserido em um ambiente para aprender as relações de ação/efeito e intenção. Ao aprendê-las, o mesmo agente, já com o conhecimento obtido no passo anterior é inserido em n novos ambientes e seu desempenho nos mesmos é analisado.
- 2. O agente observa outro agente, aprendendo suas relações de ação/efeito com o ambiente. O agente tenta prever a intenção e ações do agente após observá-lo várias vezes. A acurácia das previsões é analisada.
- 3. O agente observa outro agente, aprendendo suas relações de ação/efeito com o ambiente. Após algumas iterações, o mesmo agente é inserido em outro ambiente e tenta replicar o objetivo

Experimento 1

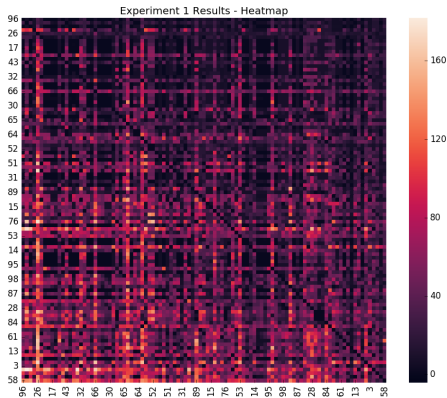
Foram gerados 100 cenário de teste. Em cada cenário, foi inserido um agente sem conhecimento prévio para realizar o aprendizado nele, gerando 100 mentes diferentes. Cada mente for colocado para agir em cada um dos cenários, para verificar o desempenho dela em relação ao que seria o número de passos ideal.

Experimento 1

Agentes no eixo x, cenário no eixo y. Os cenários foram ordenados pela complexidade do seu número de passos ideal.



Heat Map

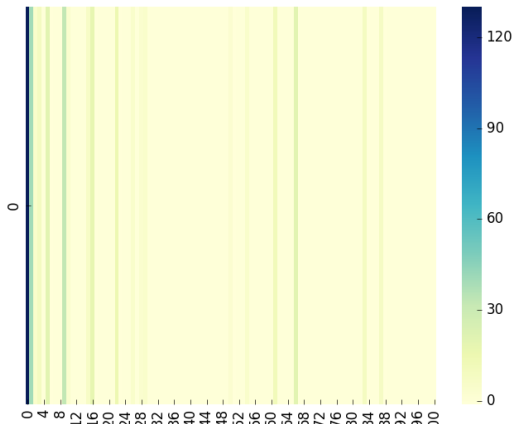


Experimento 1

Agentes de cenários que necessitaram um maior número de passos, e portanto maior exploração, tiveram um desempenho significativamente superior quando jogados em cenário distintos. Cenários que possuíam alguma característica incomum (g.e. Ponto de pickup no canto) necessitaram passos de aprendizado extra de quase todos os agentes.

Experimento 1

Um agente, a princípio sem nenhum conhecimento, aprendendo e agindo nos 100 cenários.



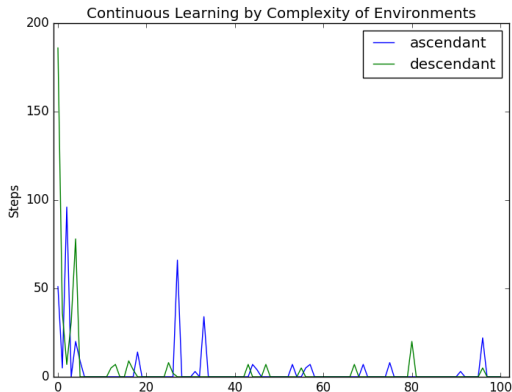
Experimento 1

Estendendo o mesmo experimento anterior para 100 agentes.
Todos começando sem conhecimento, e executando cada um nos
100 cenários dispostos de forma aleatória.



Experimento 1

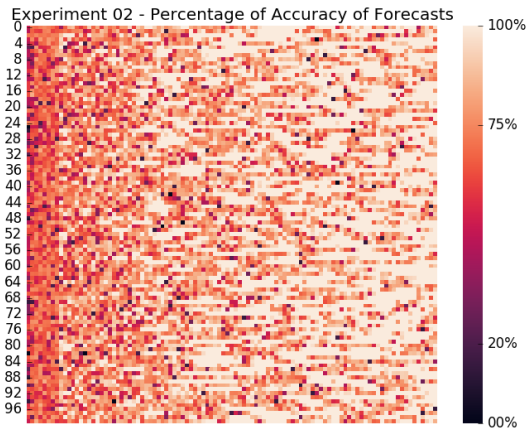
E por fim, dois agentes executando em sequências pré-determinadas: do cenário mais "difícil" ao mais "fácil", e vice-versa.



Experimento 2

Para o experimento 2, os mesmos 100 cenários foram usados. Foram gerados 100 agentes, cada um sem conhecimentos prévio, para observar as ações de um outro agente no mundo, atribuindo-lhe intenções e aprendendo suas relações. Cada um dos 100 agentes observou os cenários em uma ordem aleatória, e a razão de predições corretas foi obtida em cada episódio.

Experimento 2

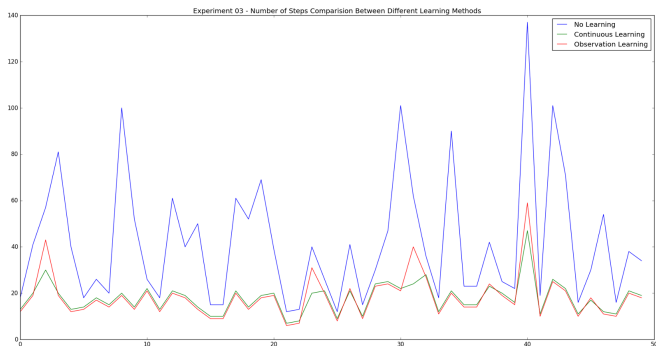


Experimento 3

Os 100 cenários foram divididos em um subconjuntos de 50 cenários de treino e 50 de teste. O desempenho de três agentes são treinados por experimentação em um dos cenários de treino.

- O primeiro será treinado somente no primeiro cenário;
- O segundo treina também nos outros 50 cenários por experimentação;
- O terceiro observa outro agente nos outros 49 cenários.

Experimento 3



Exemplos Saída lógica - Operadores Perceptuais

- $Belief_I(\neg Touch(taxi, North))$
- $Belief_I(See(horizontalWall(leftStartOfWall : 0, rightStartOfWall : 4, wallOffset : 0)))$
- $Belief_I(See(Myself(\neg passengerInTaxi, xLocation : 1, yLocation : 3)))$
- $Belief_I(See(passenger(\neg inTaxi, xLocation : 3, yLocation : 0)))$

Exemplos Saída lógica -

- $Belief_I(\gamma^+(taxiMoveNorth, (passenger, yLocation, ArithmeticEffect, -1))) \rightarrow \neg Touch(taxi, North) \wedge \neg Touch(taxi, South) \wedge On(taxi, passenger) \wedge In(taxi, passenger))$
- $Belief_I(\gamma^+(taxiMoveNorth, (taxi, yLocation, ArithmeticEffect, -1))) \rightarrow \neg Touch(taxi, North))$

Saída lógica

- $Goal_I \ll I : pickupPassenger \gg \top$
- $Belief_I(\gamma^+(taxiMoveSouth, (taxi, yLocation, ArithmeticEffect, 1)) \rightarrow \neg Touch(taxi, South))$
- $Belief_I(\gamma^+(pickupPassenger, (taxi, passengerInTaxi, AssignmentEffect, 1)) \rightarrow Touch(taxi, North) \wedge \neg Touch(taxi, South) \wedge \neg Touch(taxi, West) \wedge Touch(taxi, East) \wedge On(taxi, passenger) \wedge \neg In(taxi, passenger) \wedge \neg On(taxi, destination))$

Saída lógica

- $AGoal_I < I$:
 $dropOffPassenger > (\gamma^+(taxi, passengerInTaxi, AssignmentEffect, 0) \wedge \gamma^+(passenger, inTaxi, AssignmentEffect, 1))$
- $AGoal_I < I$:
 $pickupPassenger > (\gamma^+(taxi, passengerInTaxi, AssignmentEffect, 1) \wedge \gamma^+(passenger, inTaxi, AssignmentEffect, 1))$
- $Goal_I \ll I : dropOffPassenger \gg \top$
- $FDI_I < I$:
 $dropOffPassenger > (\gamma^+(taxi, passengerInTaxi, AssignmentEffect, 0) \wedge \gamma^+(passenger, inTaxi, AssignmentEffect, 0))$

Conclusões

In this work, we showed how to implement an agent capable of not only learning through experimentation and observation, but also distinguish between his and other agent's beliefs.

Conclusões

An important aspect of the approach presented here is how generic the learning process is. The only implicit knowledge the agent has are the set of action and propositions. The learning is done using the feedback of its perceptual operators, and could easily be adapted for a great range of different environments.

Conclusões

Our experiments showed that, with the appropriate planner, the performance of the algorithm rapidly converges to a near optimal policy for most scenarios. The experiments also demonstrated that the mental framework of the agent is decoupled from the ToM representation of other agents minds, although one can be used to update the beliefs from the other.

Conclusões

This work emphasized on the process of learning, both through experimentation and observation, therefore many other steps of the reasoning cycle were simplified, and could be expanded in future works. The process of intention generation and planning are two examples of important aspects that must be done correctly to attain a true cognitive agent.

Agradecimentos

Agradecemos a orientação do Prof. Dr. Gustavo Lugo, e a todos os que fizeram críticas construtivas no decorrer do projeto.

Referências



“The original Sally-Anne illustration by Uta Frith,” 1985.



C. Diuk, A. Cohen, and M. L. Littman, “An object-oriented representation for efficient reinforcement learning,” in *Proceedings of the 25th international conference on Machine learning*, pp. 240–247, ACM, 2008.

Modelo - Definições do Ambiente

O modelo de objetos é definido nos conceitos de classes, atributos, domínio, objetos, instâncias, estados, relações entre classes e efeitos.

Predicados não necessariamente precisam ser associados com uma posição no espaço, e podem representar outros tipos de conhecimento sobre objetos no mundo, como a quantia de energia de um determinado agente. Eles são sempre relações binárias.