

# Um Mecanismo Construtivista para a Aprendizagem de Estrutura de MDPs Fatorados e Parcialmente Observáveis

Filipo Studzinski Perotto<sup>1</sup>, Jean-Christophe Buisson<sup>2</sup>, Luis Otavio Alvares<sup>1</sup>

<sup>1</sup>Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)  
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brasil

<sup>2</sup>Institut de Recherche en Informatique de Toulouse (IRIT) – Institut National  
Polytechnique de Toulouse (INPT) – França

{fsperotto,alvares}@inf.ufrgs.br, jean-christophe.buisson@enseeiht.fr

**Resumo.** Este artigo apresenta o CALM (Constructivist Anticipatory Learning Mechanism), um mecanismo de aprendizagem para agentes baseado na abordagem construtivista da IA. Ele é desenvolvido para enfrentar de forma dinâmica e interativa ambientes que são ao mesmo tempo parcialmente determinísticos e parcialmente observáveis. O mecanismo é descrito em detalhe, explicando o método de aprendizagem empregado, e a estrutura de representação do mundo, baseada em MDPs fatorados. São analisados os tipos de regularidade que o CALM é capaz de descobrir, constituindo um caminho para a construção de conceitos abstratos e formas de representação de mais alto-nível.

**Abstract.** This paper presents CALM (Constructivist Anticipatory Learning Mechanism), an agent learning mechanism based on a constructivist approach of AI. It is designed to deal dynamically and interactively with environments which are at the same time partially deterministic and partially observable. We describe in detail the mechanism, explaining how the learning methods operate, and how it represents a world model using a factored MDP. We analyze the kinds of environmental regularities that CALM is able to discover, trying to show that our proposition follows the way towards the construction of more abstract or high-level representational concepts.

## 1. Introdução

O mundo real é um ambiente de alta complexidade, e a transição da *inteligência sensório-motora* à *inteligência simbólica* constitui um aspecto crucial para explicar como os seres humanos conseguem lidar tão bem com ele (Piaget, 1957). O problema é o mesmo para um agente artificial situado, que precisa aprender de forma incremental as regularidades de seu ambiente, apenas através de sua observação e interação com o mundo.

Em ambientes complexos, propriedades especiais ‘macroscópicas’ emergem das interações entre elementos ‘microscópicos’, e geralmente estas características emergentes não estão diretamente presentes em nenhum dos componentes que as geram. Os fenômenos importantes neste tipo de ambiente tendem portanto a estar relacionados

com processos e objetos de um nível mais alto de abstração (Thornton, 2003), e nesse caso é inadequado representar o mundo somente através de termos sensoriomotores primitivos (Drescher, 1991). Um agente inteligente (humano ou artificial) vivendo sob essas condições precisa ter a capacidade de superar os limites da percepção sensorial direta, organizando o universo através de conceitos mais abstratos. Em outras palavras, o agente precisa ser capaz de detectar regularidades de alto-nível na dinâmica do ambiente, mas isso não é possível se ele está preso a um “vocabulário representacional” pré-estabelecido e rígido.

Este artigo apresenta o CALM (*Constructivist Anticipatory Learning Mechanism*) um mecanismo de aprendizagem para agentes autônomos, inspirado na concepção construtivista da inteligência (Piaget, 1957), capaz de criar um modelo descritivo de seu universo, usando elementos abstratos, definidos por ele mesmo, para representar propriedades não-observáveis do ambiente. Trata-se do problema da indução de estrutura de MDPs fatorados e parcialmente observáveis a partir da experiência.

Na sequência do artigo, a seção 2 descreve os conceitos de agente e ambiente utilizados, e a seção 3 define o problema de aprendizagem no *framework* dos MDPs. A seção 4 apresenta o CALM, detalhando a forma de representação do conhecimento e os algoritmos de aprendizagem. A seção 5 apresenta o método utilizado para tratar propriedades não-observáveis, mostrando como sua existência pode ser descoberta e como seus valores podem ser previstos através do uso de “elementos sintéticos”. A seção 6 apresenta um problema e sua solução seguindo o método apresentado. A seção 7 compara trabalhos relacionados, e a seção 8 finaliza o artigo, argumentando que este é um importante passo na direção de formas mais abstratas de representar o mundo.

## 2. Agente e Ambiente

Neste trabalho adotamos a concepção de que agente e ambiente são as duas partes de um sistema integrado, relacionando-se por acoplamento dinâmico (Beer, 1995). Cada um deles interfere na trajetória do outro, mas sem exercer controle absoluto. O agente é uma entidade situada em seu ambiente, e é dotado de sensores e atuadores limitados, seja porque estão restritos ao alcance e ponto de vista momentâneos do agente, seja porque captam apenas uma parcela das características do ambiente. O ambiente é portanto apenas parcialmente observável pela percepção sensorial do agente, e apenas parcialmente passível de ser transformado por suas ações (Suchman, 1987). Devido a isso, o agente pode frequentemente se encontrar incapaz de distinguir perceptivamente entre diferentes estados do mundo que aparecem para ele como idênticos. Da mesma forma, o controle que o agente tem sobre as transformações do ambiente é apenas parcial e localizado.

Esta ambiguidade na percepção dos estados do mundo (*perceptual aliasing*), pode tornar muito mais difícil a tarefa realizada por um algoritmo de aprendizagem de construir um modelo consistente e políticas estáveis (Crook; Hayes, 2003). É comum neste caso separar o problema de aprendizagem em duas partes (Degris et al. 2006): primeiro (a) construir um modelo antecipatório da dinâmica de interações entre o agente e o mundo e, baseado nele, e em seguida (b) encontrar uma política que maximize seu desempenho em uma janela de longo-prazo.

Para o problema da definição de políticas, algoritmos eficientes estão bem definidos em (Guestrin et al., 2003) e (Boutilier et al. 2000), entretanto estes métodos

requerem, de antemão, a estrutura e os parâmetros do FMDP para o qual a política será produzida. Uma vez que esta estrutura é desconhecida do agente, como é o caso na maioria dos problemas, então é preciso resolver também o problema da aprendizagem de modelos de mundo. O mecanismo CALM responde por esta tarefa, construindo um modelo compatível com a representação utilizada pelos métodos referidos.

Muitos dos algoritmos propostos para estes problemas de aprendizagem utilizam representações baseadas na enumeração de estados, frequentemente definidos como um *Processo de Decisão de Markov* (MDP), que pode ser também estendido para o caso parcialmente observável (POMDP). O problema é que nestas representações “rasas” o tamanho do problema cresce exponencialmente com o número de atributos considerados (*maldição da dimensionalidade*), tornando tais soluções inviáveis para problemas grandes (Degris et al. 2006).

Em vista disso, a comunidade de IA tem, cada vez mais, utilizado representações fatoradas, onde não se enumeram os estados, mas tratam-se diretamente as propriedades que os definem. O uso de um MDP fatorado (FMDP) possibilita explorar a estrutura do ambiente para representá-lo de forma compacta, mesmo no caso em que o MDP correspondente é exponencialmente grande (Guestrin et al. 2003). Do mesmo modo, um POMDP também pode ser representado de forma fatorada, constituindo então um FPOMDP (Shani et al. 2008).

Um FPOMDP é formalmente definido como uma 4-upla  $\{X, A, R, T\}$  onde  $X = \{X_1, X_2, \dots, X_n\}$  é um conjunto de  $n$  propriedades, entre as quais  $X_{(p)}$ , é o subconjunto das propriedades observáveis, acessíveis através da percepção, e  $X_{(h)}$ , seu complementar, o das propriedades ocultas ou não-observáveis,  $A = \{A_1, A_2, \dots, A_m\}$  é um conjunto de  $m$  variáveis de atuação do agente,  $R = \{R_1, R_2, \dots, R_n\}$  é o conjunto de funções de recompensa, na forma  $R_i : X_i \rightarrow \mathfrak{R}$ , e  $T = \{T_1, T_2, \dots, T_n\}$  é o conjunto de funções de transformação, na forma  $T_i : X \times A \rightarrow X_i$ , definindo a dinâmica de evolução do sistema. Cada função de transformação pode ser representada por uma *Rede Bayesiana Dinâmica* (DBN), na forma de um grafo de duas camadas, acíclico e orientado (Boutilier et al. 2000). Os nós da primeira camada representam o estado do ambiente no tempo  $t$ , e os nós da segunda camada representam o estado do ambiente no instante seguinte  $t+1$ .

A descrição de um ambiente baseado em propriedades (estados fatorados) permite uma relação mais natural com o modelo de agente situado pois assume-se que cada sensor do agente constitui uma propriedade observável de  $X$ , e cada atuador, uma propriedade em  $A$ . Fazendo referência às propriedades é possível representar os estados de um ambiente de forma generalizada. Em ambientes bem estruturados, onde as causas dos eventos são razoavelmente bem definidas, então uma representação fatorada permite a construção de uma descrição muito mais compacta da dinâmica, frente à correspondente representação “rasa” através de estados enumerados (Shani et al. 2008).

### 3. O Problema de Aprendizagem de Modelos de Mundo

O mecanismo CALM é desenvolvido para aprender a estrutura de um ambiente desconhecido através de observação e experimentação, induzindo um *modelo antecipatório* que é representado através de um FPOMDP parcialmente determinístico. O problema de aprendizagem é colocado para o CALM como um caso de *aprendizagem ativa e incremental*, onde além de observar o ambiente, o agente também tem de escolher entre ações alternativas a executar, aprendendo gradativamente a partir da

experiência, ao mesmo tempo em que desempenha sua atividade, não havendo um período prévio e separado de treinamento. A aprendizagem se dá num *horizonte infinito*, onde o agente realiza uma única e ininterrupta “caminhada” sobre o ambiente.

Este trabalho supõe que o ambiente é *parcialmente determinístico*, ou seja, entre a totalidade das transformações em  $T$ , uma parte delas se dá de forma estocástica, e outra parte se dá de forma determinística. CALM é designado para construir uma descrição das regularidades determinísticas em um ambiente parcialmente determinístico. Por um lado, tal situação é mais confortável do que o caso não-determinístico porque o mecanismo não precisa induzir as distribuições de probabilidades das transições não-determinísticas, contentando-se simplesmente em detectá-las. Por outro lado, trata-se de um caso mais difícil do que o caso em que o ambiente é sabidamente determinístico, pois nessa condição uma única observação de uma transição seria suficiente para defini-la no modelo. No caso parcialmente determinístico uma transformação aparentemente estável está sempre sujeita a se mostrar não-determinística. Assim, a rigor, restringindo-se ao caso parcialmente determinístico, pode-se dizer que o CALM constrói a estrutura de um PD-FPOMDP a partir da experiência, mas aí sim a sigla começa a ficar grande.

Entretanto, o CALM lida com situações de observabilidade parcial, ou seja, casos em que parte das propriedades do ambiente relevantes para a descrição das transformações não é perceptivamente acessível pelos sensores do agente. Em ambientes parcialmente observáveis, não é possível identificar os estados apenas com base na observação, e isso traz sérias implicações ao modelo de transformações que o agente está tentando aprender. Ambientes parcialmente observáveis podem apresentar uma dinâmica altamente complexa e não-determinística na sua face observável, quando de fato podem ser determinísticos e portanto facilmente predizíveis em relação a seus elementos subjacentes não-observáveis (Holmes; Isbell, 2006).

#### 4. O Mecanismo de Aprendizagem CALM

CALM constrói árvores de esquemas para representar as regularidades que ele percebe no ambiente a partir de suas interações, uma árvore para cada propriedade a ser antecipada. Cada esquema representa uma transformação confirmada pelo agente. Um esquema é uma 3-upla  $\Xi = \{x \wedge a \rightarrow x_i'\}$ , onde  $x \in X$  é um vetor que representa o contexto (baseado na combinação de propriedades perceptivas),  $a \in A$  é uma ação, e  $x_i'$  é uma expectativa, ou seja, o valor previsto para uma determinada propriedade  $i$  em  $X$  no instante seguinte, após o agente realizar a ação  $a$  estando na situação  $x$ . Os vetores de contexto e ação podem representar situações generalizadas, pois são definidos somente em relação a propriedades consideradas relevantes para a antecipação.

Uma árvore de esquemas aprofunda-se pela especialização dos esquemas. O nó raiz define contexto e ação completamente generalizados, onde os elementos são marcados como “#” (indefinidos). Cada nível adicionado à árvore diferencia um elemento, seja em  $x$  ou em  $a$ , criando ramos mais especializados. Transformações descobertas como não-determinísticas são representadas por esquemas cuja expectativa é indefinida.

Durante o desempenho de sua atividade, a cada ciclo de tempo  $t$  o agente percebe uma determinada situação  $x$ , que excita o subconjunto de esquemas compatíveis com a situação, entre o conjunto total de esquemas. Os esquemas excitados descrevem possíveis mudanças no ambiente, em função de ações alternativas. CALM analisa entre

os esquemas excitados e escolhe as ações que irá executar, ativando a combinação de esquemas que supostamente o leva a maximizar o valor esperado das recompensas.

CALM opera de forma incremental. A partir de um FPOMDP inicial, o mecanismo gradativamente refina o modelo a cada nova experiência que o ponha em desequilíbrio. Quando a expectativa de um esquema falha, então o mecanismo corrige a árvore correspondente, geralmente especializando-a mais, ou então indefinindo a expectativa do esquema se sua especialização não é possível ou se a árvore já atingiu um limite de especialização máxima, (para evitar *overfitting*).

O agente tem uma árvore para cada propriedade que deseja ser antecipada. Como a função de recompensa é fatorada, inicialmente o agente deseja prever apenas aquelas propriedades que estão associadas a recompensas não nulas (positivas ou negativas), senão, aquelas que ajudem a antecipar o valor dessas primeiras. Estas árvores iniciais possuem apenas um esquema raiz, que define uma situação completamente generalizada, onde  $x = \#$  e  $a = \#$ . Entretanto, a expectativa será inicializada supondo a possibilidade de uma antecipação definida, criada ingenuamente a partir da primeira experiência compatível com o esquema. Dito de outro modo, a expectativa de um esquema é iniciada como o espelho do primeiro resultado observado por ele.

Os vetores de contexto e ação vão sendo gradativamente especializados, criando-se novos ramos na árvore, que a princípio devem incluir propriedades identificadas como relevantes para a descrição da transformação. O crescimento da árvore acontece enquanto é possível preservar o determinismo de uma antecipação. No limite em que isso não é mais possível, ou porque uma transformação não é determinística, ou porque a árvore atingiu um limite máximo de especialização, então a expectativa é generalizada, e eventualmente, a árvore pode ser reduzida. Três métodos básicos compõem a função de aprendizagem do CALM: diferenciação, ajuste e integração, conforme ilustrados nas figuras 1, 2 e 3.

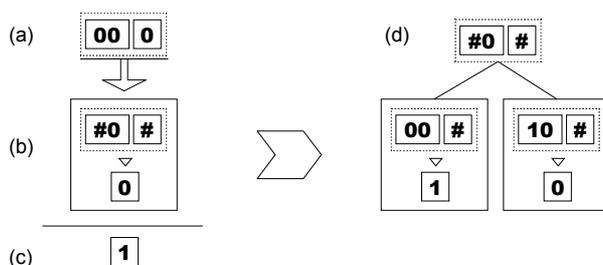


Figura 1: Diferenciação. (a) contexto perceptivo e ação executada; (b) esquema ativado; (c) resultado observado; (d) transformação implementada.

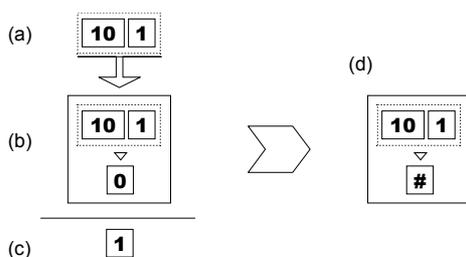


Figura 2: Ajuste.

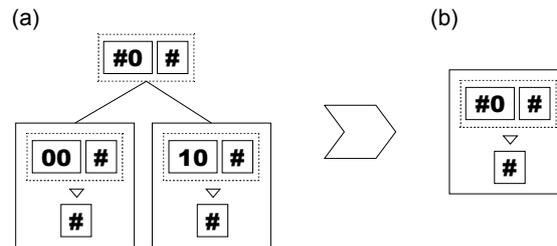


Figura 3: Integração. (a) sub-árvore original; (b) esquema integrado.

O método de *diferenciação* transforma um esquema que não teve sua expectativa correspondida, em uma sub-árvore de dois níveis, especializando os ramos. A diferenciação identifica algum elemento da ação ou do contexto capaz de diferenciar as situações até então representadas no esquema, da situação atual. O esquema é então dividido, e cada um dos novos esquemas mais especializados se encarrega da parte do domínio em que sua antecipação permanece válida. Em outras palavras, a nova situação que evidenciou uma transformação diferente é isolada das outras situações para as quais até então o esquema vinha funcionando bem.

É a diferenciação que promove o crescimento das árvores, adicionando propriedades consideradas relevantes como diferenciadores nas árvores. A escolha de uma propriedade diferenciadora (*feature selection*) constitui um problema delicado envolvido neste método. Para realizá-la, é necessário de que os esquemas guardem uma memória generalizada representando situações passadas, e possibilitando assim diferenciá-las da situação presente.

Quando nenhum tipo de diferenciação é mais possível, ou quando a diferenciação da árvore chegou a um limite de superespecialização, então o mecanismo executa o método de *ajuste*, que reequilibra o esquema pela redução de suas expectativas. Reduzir a expectativa de um esquema significa simplesmente indefinir sua antecipação. A combinação de sucessivos ajustes pode revelar diferenciações desnecessárias. Assim, o método de *integração* tem a função de constantemente verificar a redundância entre ramos da árvore, e unificá-los, reduzindo o tamanho da árvore.

Dessa forma, o percurso realizado pelo conhecimento pode ser imaginado como um primeiro momento onde predominam as diferenciações, quando, portanto as árvores de esquemas tendem a crescer e se especializar, e em seguida, um segundo momento onde predominam os ajustes nas expectativas e as integrações de esquemas, quando então a árvore tende a se reduzir, atingindo o ponto em que represente apenas as regularidades realmente verificadas no ambiente.

## 5. Tratando a Observação Parcial

Quando não é mais possível diferenciar as situações por elementos perceptivo-sensoriais, então o mecanismo CALM tenta induzir a existência de *propriedades ocultas*, através da introdução de elementos sintéticos nos vetores de contexto dos esquemas. Esses novos elementos representam variáveis que não são diretamente observáveis pelo agente, mas que podem ter seus valores deduzidos através da análise do encadeamento histórico dos contextos perceptivos, e assim, podem ajudar a realizar novas antecipações. A descoberta de propriedades ocultas insere-se depois da diferenciação sensorial, e antes do ajuste.

Se um esquema falha e não há mais alternativa de diferenciação por propriedades sensorio-motoras, mesmo assim o mecanismo não indefine ainda sua expectativa, o que implicaria na suposição de que aquela transformação não seria determinística. O fato é que ela pode ser determinística desde que consideradas certas condições que, entretanto, não são diretamente observáveis. É para este fim que o CALM recorre a um método para construir propriedades abstratas, representadas através de *elementos sintéticos* incluídos no vetor de contexto dos esquemas, e que podem servir para ampliar o horizonte de diferenciações.

Dessa forma, quando lidando com ambientes parcialmente observáveis, CALM tem dois desafios adicionais: (a) inferir a existência de características não-observáveis do ambiente, que serão representadas como elementos sintéticos do contexto, e (b) incluir estes novos elementos no seu modelo preditivo. Uma boa estratégia para realizar esta tarefa é observar numa janela mais ampla a história das interações entre agente e ambiente. É possível encontrar pedaços suficientemente pequenos de história capazes de distinguir e identificar a evolução de características não-observáveis do mundo (Holmes; Isbell, 2006).

Quando a diferenciação sensorial é chamada e não encontra solução, então CALM executa a diferenciação abstrata, que simplesmente cria um novo elemento sintético booleano, o qual é então incluído entre os termos de contexto do agente. Imediatamente, esse elemento é usado como diferenciador para o esquema em desequilíbrio, separando a situação atual das anteriores pela atribuição de diferentes valores (definidos de forma arbitrária) para essa nova propriedade. Esses valores representam a presença ou ausência de alguma condição não-observável do ambiente, necessária para corrigir o modelo antecipatório do agente com relação à dada situação. Uma vez que um novo elemento sintético é criado, pode ser usado em diferenciações futuras. Para evitar que o CALM crie elementos sintéticos indefinidamente, é possível parametrizar um limite máximo, depois do qual as situações problemáticas são resolvidas através do ajuste. O método é ilustrado na figura 4, onde os elementos sintéticos estão representados por naipes de baralho.

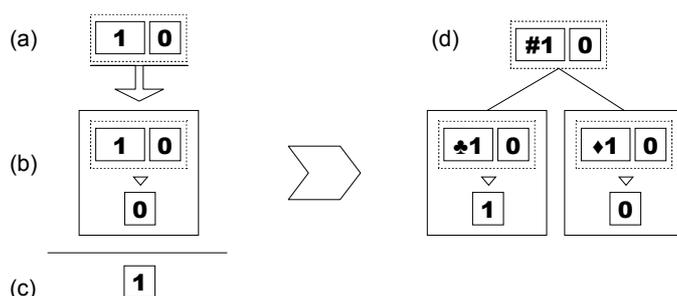


Figura 4: Criação do Elemento Sintético. (a) situação experimentalada. (b) esquema ativado. (c) resultado realmente observado. (d) diferenciação abstrata.

Como o elemento sintético não está associado à percepção sensorial do agente, seu valor não pode ser diretamente observado. Inicialmente, logo após ser incluído no repertório de propriedades do contexto, o valor de um elemento sintético é verificável apenas *a posteriori*, ou seja, depois de se executar uma ação em uma situação ambígua, e então, verificando o resultado e deduzindo daí o valor necessário do elemento sintético. Descobrir o valor de um elemento sintético depois que a circunstância onde a

informação era necessária pode parecer inútil, mas de fato, esta dedução tardia serve de parâmetro para a realização da *antecipação abstrata*.

Se a propriedade não-observável representada pelo elemento sintético possui um comportamento regular, então o mecanismo pode retropropagar o valor deduzido para o esquema ativado no instante anterior, que vai incluí-lo como uma antecipação. Por exemplo, no tempo  $t_1$  CALM ativa o esquema  $\Xi_1 = (\#0 \wedge w \rightarrow \#)$ , que faz uma antecipação indefinida de um elemento sintético. Isto o leva em  $t_2$  a uma situação ‘#’ ambígua, que irá excitar esquemas antagônicos  $\Xi_2 = (\clubsuit\# \wedge w \rightarrow 0)$  e  $\Xi_3 = (\diamond\# \wedge w \rightarrow 1)$ . O mecanismo não pode saber em  $t_2$  o valor do elemento sintético, crucial para determinar qual é a real situação que está sendo enfrentada. Entretanto, após executar a ação ‘w’ em  $t_2$ , o agente observará os resultados de sua ação, e irá perceber, por exemplo, um ‘0’ em  $t_3$ . Agora, estando em  $t_3$ , o agente pode deduzir que a situação que se apresentava de fato em  $t_2$  era ‘ $\clubsuit$ ’, e ele pode incluir essa informação no esquema que havia sido ativado em  $t_1$ , transformado então para  $\Xi_1 = (\#0 \wedge w \rightarrow \clubsuit)$ .

## 6. Resultados Experimentais

O mecanismo CALM já havia sido testado no âmbito de problemas completamente observáveis através de experimentos que, embora simples, demonstraram a viabilidade e a capacidade de convergência do método (Perotto; Álvares, 2006). Para testar sua funcionalidade no caso parcialmente observável, CALM foi testado com o problema flip (Holmes; Isbell 2006). Trata-se de um agente vivendo num mundo de dois estados, os quais são para ele ocultos. Ele tem um atuador com três ações possíveis (l, r, u) e uma percepção com dois valores possíveis (0, 1). O agente percebe ‘1’ quando o estado subjacente troca, e ‘0’ se permanece no mesmo estado. A ação u não faz nada, a ação l faz uma transição determinística para o estado da esquerda, e r do mesmo modo para a direita, conforme a figura 5.

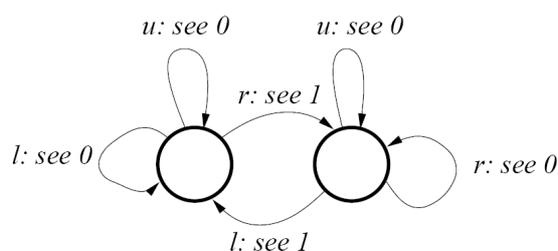


Figura 5: O problema *flip*.

CALM é capaz de resolver o problema através da criação de um novo elemento sintético para representar os estados subjacentes esquerdo ( $\clubsuit$ ) e direito ( $\diamond$ ). Durante uma caminhada típica do agente pelo ambiente, os primeiros passos não indicam a existência de propriedades não-observáveis. Entretanto, em algum momento um esquema desequilibrado exigirá uma diferenciação abstrata, uma nova árvore para antecipar o novo elemento sintético. As árvores finais, estáveis, são alcançadas ao fim das antecipações e diferenciações abstratas, onde um único elemento sintético é necessário. A figura 6 mostra as árvores definitivas.

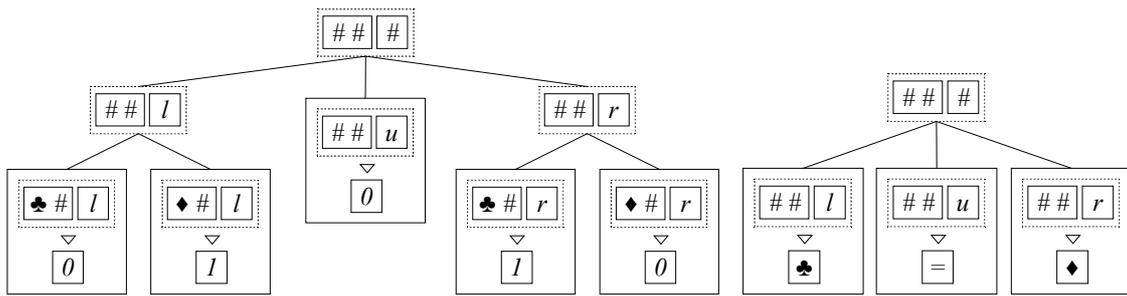


Figura 6: solução para o problema *flip*: à esquerda a árvore para antecipação da percepção, e à direita a árvore para antecipação do estado oculto.

## 7. Trabalhos Relacionados, Conclusões e Próximos Passos

**CALM** é um mecanismo original que permite a um agente construir incrementalmente um modelo de mundo durante o curso de suas interações com ele. O mecanismo descobre regularidades em ambientes parcialmente determinísticos e parcialmente observáveis, a partir de um método diretamente indutivo.

O trabalho pioneiro apresentado por Gary Drescher (1991) pode ser considerado o primeiro mecanismo de IA realmente inspirado pela descrição construtivista da inteligência, incluindo a descoberta de propriedades não-observáveis. Entretanto o método de aprendizagem proposto naquele trabalho baseia-se em análise exaustiva de correlações estatísticas, obrigando a representar as combinações de propriedades, o que faz dele um método pouco eficiente em termos computacionais conforme se aumenta o tamanho do problema (Franklin, 1995). Outros autores tentaram dar continuidade a seu trabalho, sem no entanto atingir resultados de impacto.

O mecanismo CALM, proposto em nosso trabalho, compara-se de forma mais clara com os mecanismos apresentados em (Holmes; Isbell, 2006), que constrói uma árvore de sufixos (porém tratando *perceptual aliasing* através de sequências históricas e não pela indução de propriedades abstratas), e em (Degris et al., 2006), que constrói a estrutura de um FMDP pela observação (porém fazendo-o de forma não-incremental, e para ambientes completamente observáveis).

O uso de elementos sintéticos permite ultrapassar os limites sensório-motores do conhecimento. No CALM, elementos sintéticos podem representar 5 tipos de propriedades não-observáveis: (1) propriedades ocultas em ambientes parcialmente observáveis, (2) sub-ambientes em mundos não-estacionários discretos, (3) marcadores de etapas em sequências de ações, (4) marcadores de posição relativos à diferentes pontos de vista, e (5), propriedades de fato abstratas, que podem se tornar ferramentas poderosas de compreensão, habilitando o agente a organizar o universo em níveis conceituais mais altos.

O CALM representa uma solução interessante para problemas onde o ambiente é visto pelo agente como parcialmente observável e parcialmente determinístico, desde que este ambiente seja bem estruturado, ou seja, que o acesso perceptivo a propriedades relevantes seja razoável, que uma parcela grande das transformações seja determinística, ao menos no que diz respeito ao modelo subjacente não-observável das transformações, e que as transformações relativas a cada propriedade dependam apenas de um pequeno subconjunto de outras propriedades, e não da descrição completa dos estados.

A sequência do trabalho inclui a experimentação do CALM em outros problemas, sobretudo os de mais complexidade, além da extensão do CALM para aprender regularidades não-determinísticas, e também tratar variáveis contínuas.

## Referências

- Beer, R.D. (1995). A dynamical systems perspective on agent-environment interactions. *Artificial Intelligence*. Elsevier, v.72, p.173-215.
- Boutilier, C.; Dearden, R.; Goldszmidt, M. (2000). *Stochastic dynamic programming with factored representations*. *Artificial Intelligence*. Elsevier, v.121, p.49-107.
- Crook P.; Hayes G. (2003). *Learning in a State of Confusion: Perceptual Aliasing in Grid World Navigation*. In: Proc. Towards Intelligent Mobile Robots, Bristol: UWE.
- Degrís, T.; Sigaud, O.; Willemin, P-H. (2006). *Learning the Structure of Factored Markov Decision Processes in Reinforcement Learning Problems*. Proc. 23<sup>th</sup> ICML.
- Drescher, G. (1991). *Made-Up Minds: a constructivist approach to artificial intelligence*. Cambridge: MIT Press.
- Franklin, S. (1995). *Artificial Minds*. Cambridge: MIT Press.
- Guestrin, C.; Koller, D.; Parr, R.; Venkataraman, S. (2003). *Efficient Solution Algorithms for Factored MDPs*. *Journal of Artificial Intelligence Research*. Morgan Kaufmann, v.19, p.399-468.
- Holmes, M. and Isbell, C. (2006). *Looping Suffix Tree-Based Inference of Partially Observable Hidden State*. Proc. 23<sup>th</sup> ICML.
- Perotto, F.S., Alvares, L.O. (2006). *Incremental Inductive Learning in a Constructivist Agent*. Proc. of SGAI. London: Springer-Verlag, p.129-144.
- Piaget, J. (1957). *Construction of Reality in the Child*. London: R. Kegan Paul.
- Shani, G.; Poupart, P.; Brafman, R.; Shimony, S. (2008). *Efficient ADD Operations for Point-Based Algorithms*. Proc. ICAPS, p.330-337.
- Suchman, L.A. (1987). *Plans and Situated Actions*. Cambridge University Press.
- Thornton, C. (2003). *Indirect sensing through abstractive learning*. *Intelligent Data Analysis*, v.7, n.3, p.1-16.