Proposta de pesquisa para doutorado

Nome: Nicolau Leal Werneck
Título: Determinação de hierarquias computacionais em MDPs.
Orientação: profa. Anna Helena Reali Costa
Período de seleção: 3^o período de 2007.

1 Objetivos

Muitos estudos em inteligência artificial envolvem sistemas onde ambientes são povoados por agentes que interagem com outras entidades [1, 2]. Estes agentes devem tomar decisões a cada momento, e podem ser direcionados a buscar objetivos que são usualmente definidos como a otimização de uma recompensa numérica determinada a partir do estado do sistema.

Podemos analisar estes sistemas tomando-os como processos de decisão markovianos (*Markov Decision Process*, MDP) [3]. Este modelo permite a aplicação da técnica de aprendizado por reforço, que em condições adequadas garante o aprendizado de uma política de atuação ótima.

Existem diversas extensões pertinentes do MPD. Podemos estudar sistemas onde múltiplos agentes podem interagir com diferentes objetivos [4]. Outra extensão que pretendemos abordar é o POMDP, onde se considera que existem variáveis no sistema que não são acessíveis pelos agentes [5].

Uma extensão além é o modelo descentralizado, DEC-POMDP, onde os agentes atuam de forma solitária, sem contato irrestrito a informações disponíveis para outros agentes [6]. É possível ainda considerar variáveis contínuas nos sistemas [7].

O maior problema no uso de MDPs e aprendizagem por reforço é que conforme os sistemas crescem, aumenta rapidamente o número de estados possíveis apresentados aos agentes. Isto faz com que o custo computacional do processo de aprendizado comece a torná-lo impraticável.

Diferentes soluções para este problema vêm sendo propostas na literatura. O que se faz é encontrar maneiras de reduzir na prática o número de estados considerados no treinamento. Por exemplo, pode ser possível levar em consideração que certos grupos de estados são semelhantes [8].

Alguns algoritmos de aprendizado procuram no sistema estruturas que possam ser exploradas para se realizar simplificações automaticamente [9,

10]. Simplificações também podem ser obtidas com a consideração do funcionamento do sistema em diferentes escalas de tempo, e com a construção de estruturas hierárquicas de controle [11, 12].

Nosso objetivos serão investigar formas de identificar estruturas internas nos sistemas para produzir representações mais eficientes de estados, e ainda tentar identificar estruturas hierárquicas no funcionamento dos sistemas que possam beneficiar suas caracterizações [12].

Buscaremos em nosso trabalho tentar relacionar esta tentativa de encontrar estruturas intrínsecas aos sistemas com os conceitos surgidos nas últimas décadas de autopoiese [13, 14] e de biologia relacional [15].

2 Metodologia

Realizaremos testes onde diferentes algoritmos de aprendizagem serão empregados para treinar agentes atuando em diferentes problemas. Dentre os algoritmos estarão os que procuram de forma automática estruturas hierárquicas no funcionamento dos sistemas.

Vamos comparar o desempenho destes algoritmos em diferentes problemas. Esperamos poder observar de que forma mudanças sutis na definição dos problemas afetam as estruturas de computação encontradas com os treinamentos. Poderemos ainda propor novos algoritmos de identificação de hierarquias e estruturas computacionais baseadas em componentes.

Entre os problemas de teste que desejamos utilizar estão jogos clássicos e bem-conhecidos como o Sokoban [16], e mais recentes como o Arimaa [17].

3 Recursos

Os trabalhos a serem executados dependem apenas da disponibilidade de computadores. Alguns testes poderão demandar a execução de programas muito longos. Seria interessante portanto o acesso temporário a algum *cluster* que possa realizar em menos tempo estas pesquisas mais custosas.

A plataforma de desenvolvimento preferida seria composta apenas por softwares livres, como o sistema operacional GNU/Linux, compilador de C++ gcc e o software Octave (que é compatível com Matlab).

4 Cronograma

	ANO E QUADRIMESTRE					
	2007	2008			2009	
ATIVIDADE	3^o	1^o	2^o	3^o	1^o	2^o
Duas disciplinas.	×					
Duas disciplinas.		×				
Uma disciplina.			×			
Exame de redação em inglês				×[a]		
Estudos sobre MDP, POMDP, DEC-MDP e DEC-POMDP.	×	×	×	×	×	×
Desenvolvimento de ambiente para experimentos.		×	×			
Testes de algoritmos para problemas DEC-MDP/POMDP.				×	×	
Elaboração de artigo para congresso nacional[b]					×	
Desenvolvimento de proposta de tese.					×	×
Exame de qualificação						×

	2009	2010			2011	
ATIVIDADE	3^o	1^o	2^o	3^o	1^o	2^o
Aprimoramento da proposta de tese.	×	×	×			
Implementação e testes preliminares da proposta.	×	×				
Elaboração de artigo para congresso internacional[c]	×	×				
Implementação e análise de resultados da proposta				×	×	×
Elaboração de artigo para revista internacional				×	×	
Redação da tese					×	×
Defesa de tese						×

[a] Em Agosto, ou assim que for oferecido pelo programa.
[b] Provavelmente o VII Encontro Nacional de Inteligência Artificial.
[c] Provavelmente ICML ou ECAI

Referências

[1] R. D. Beer, "A dynamical systems perspective on agent-environment interaction," *Artificial Intelligence*, vol. 72, 1995.

[2] R. Brooks, "Elephants don't play chess," *Robotics and Autonomous Systems*, vol. 6, pp. 3–15, 1990.

[3] M. L. Putterman, *Markov Deciosion Processes: Discrete Stochastic Dynamic Programming*. Wiley, 1994.

[4] M. L. Littman, "Markov games as a framework for multi-agent reinforcement learning," in *Proceedings of the Eleventh International Conference on Machine Learning*, pp. 157–163, Morgan Kaufmann, 1994.

[5] W. Zhang, *Algorithms for Partially Observable Markov Decision Processes*. PhD thesis, Hong Kong University of Science and Technology, 2001.

[6] R. Nair, M. Tambe, M. Yokoo, D. Pynadath, and S. Marsella, "Taming decentralized pomdps: Towards efficient policy computation for multiagent settings," 2003.

[7] C. Guestrin, M. Hauskrecht, and B. Kveton, "Solving factored mdps with continuous and discrete variables," 2004.

[8] R. Pegoraro and A. H. R. Costa, "Agilizando aprendizagem por reforço através da utilização de similaridades entre estados," in *International Joint Conference IBERAMIA'2000 and SBIA'2000, Workshop Proceedings, Meeting on Multi-Agent Collaborative and Adversarial Perception, Planning, Execution, and Learning* (L. N. Barros, R. M. C. Jr., F. G. Cozman, and A. H. R. Costa, eds.), pp. 175–184, November 2000.

[9] C. Boutilier, R. Dearden, and M. Goldszmidt, "Exploiting structure in policy construction," in *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence* (C. Mellish, ed.), (San Francisco), pp. 1104–1111, Morgan Kaufmann, 1995.

[10] S. Thrun and A. Schwartz, "Finding structure in reinforcement learning," in *Advances in Neural Information Processing Systems 7*, MIT Press., 1995.

[11] M. Hauskrecht, N. Meuleau, L. P. Kaelbling, T. Dean, and C. Boutilier, "Hierarchical solution of markov decision processes using macroactions.," in *UAI* (G. F. Cooper and S. Moral, eds.), pp. 220–229, Morgan Kaufmann, 1998.

[12] A. Barto and S. Mahadevan, "Recent advances in hierarchical reinforcement learning," 2003. Discrete event systems (2003, to appear).

[13] R. D. Beer, "Autopoiesis and cognition in the game of life," *Artificial Life*, vol. 10, pp. 309–326, 2004.

[14] F. J. Varela and H. R. Maturana, *Autopoiesis and Cognition: The Realization of the Living*. Springer, 1991.

[15] R. Rosen, *Life Itself*. Columbia University Press, 1991.

[16] D. Dor and U. Zwick, "Sokoban and other motion planning problems," *Computational Geometry*, vol. 13, no. 4, pp. 215–228, 1999.

[17] O. Syed and A. Syed, "Arimaa — a new game designed to be difficult for computers," *International Computer Games Association Journal*, vol. 26, pp. 138–139, 2003.

Nicolau Leal Werneck
Candidato

Professora Anna Helena Reali Costa