

Árvores de Decisão para Descoberta de Padrões Comportamentais de Agentes

Josimara de Á. Silveira, Thiago L. Paes, Leonardo R. Emmendorfer e Diana F. Adamatti

Centro de Ciências Computacionais – C3
Universidade Federal do Rio Grande (FURG)
Av. Itália, km 08, Carreiros
96201-090 – Rio Grande -RS - BRASIL

josi_silver@hotmail.com, thiagopaes@gmail.com, leonardo.emmendorfer@gmail.com, dianaadamatti@furg.br

1. RESUMO

Uma questão muito freqüente em aprendizado artificial diz respeito aos limites associados a cada uma das técnicas. Em outras palavras, qual o escopo de aplicação de cada técnica de aprendizado artificial? Com isso em mente, surge a idéia de conduzir experimentos de modo a responder esta questão e procurar descobrir até que ponto as árvores de decisão são suficientes para detectar padrões de comportamentos de agentes e gerar descrições aproximadas da sua personalidade. O trabalho tomou como base um modelo multiagente do tipo presa-predador[1][2] implementado na versão 4.04 do NetLogo[4]. Foi criada uma versão estendida da original, na qual foram adicionados comportamentos entre os agentes, que antes se moviam aleatoriamente. Foram escolhidos empiricamente conjuntos de parâmetros para os quais o modelo é estável, representando um ecossistema auto-sustentável.

Os registros provenientes do NetLogo foram tratados para servir de entrada para o software WEKA[5], v3.7, o qual utiliza uma coleção de algoritmos de aprendizado de máquina para realizar mineração de dados. Os registros dos agentes foram classificados utilizando o algoritmo de árvore de decisão J48, uma variante do algoritmo C4.5[3]. A classificação dos dados no WEKA mostrou que as árvores de decisão tiveram poder suficiente para capturar e explicitar comportamentos que estavam implícitos no código.

Como resultado, as técnicas utilizadas mostraram que, para o modelo em questão, o comportamento efetivo dos agentes (detectado pelas árvores) pode ser representado por um conjunto pequeno de variáveis se comparado com o comportamento programado (explícito no código) e o erro é baixo (5,1% e 5,8% para lobos e ovelhas, respectivamente). Para ambos os agentes, tendo em vista as árvores geradas, 4 (quatro) é o número máximo de variáveis que precisam ser analisadas para sabermos qual a ação de cada agente. Em contrapartida, ao olhar o código referente à simulação, o número de variáveis é consideravelmente maior.

REFERÊNCIAS

- [1] F. Hoppensteadt, "Predator-prey model", *Scholarpedia*, 1(10), 1563, 2006
- [2] A. J. Lotka, Lotka on population study, ecology, and evolution. *Population and Development Review*, 15(3), 539–550, 1989
- [3] JR. Quinlan, C4.5: Programs for Machine Learning. *Morgan Kaufmann Publishers*, 1993.
- [4] U. Wilensky. NetLogo Wolf Shee Predation model. *Center for Connected Learning and Computer-Based Modeling*, Northwestern University, Evanston, IL.1997
- [5] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten . The WEKA Data Mining Software: An Update; *SIGKDD Explorations*, Volume 11, Issue 1, 2009