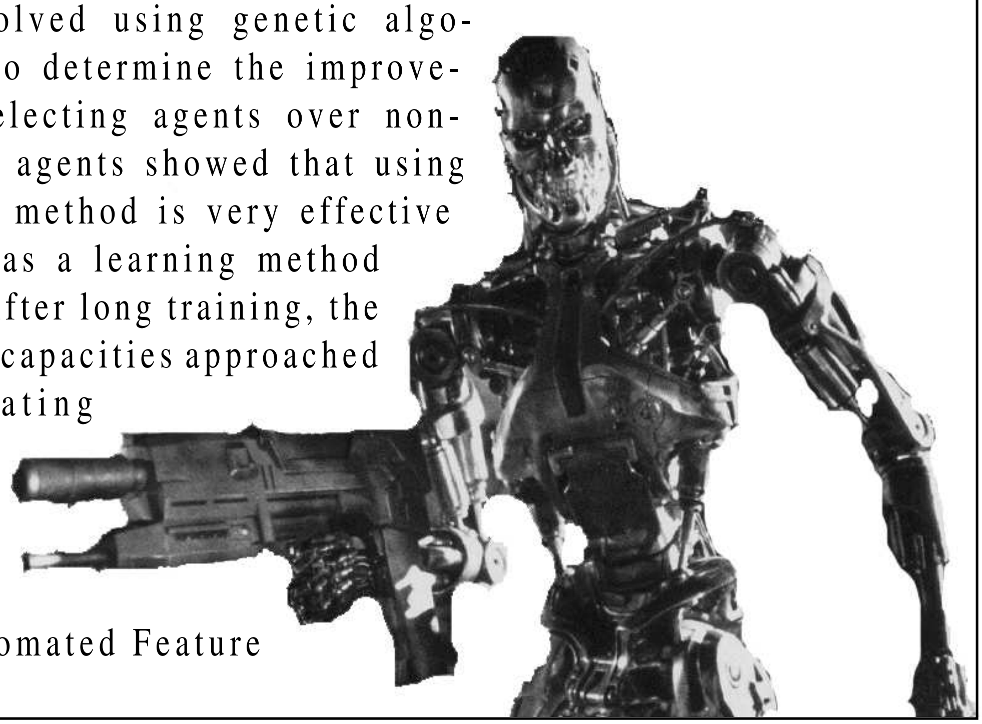
Automated Feature Selection to Maximize Learning in Artificial Intelligence

Joseph P. Turian

ABSTRACT

Traditional artificial intelligence computer programs reach a plateau beyond which learning stops. Game-playing artificial intelligence agents play at a level anywhere from beginner to expert but are never optimal, that is, they do not always make the best choice. The purpose of this project was to improve computer learning techniques so that plateau learning ability is increased. Automated Feature Selection (AFS), a new general technique, was created and implemented in two algorithms, Randomly Generating Operations (RGO) and Incremental Addition (IA). A key component of these algorithms is that neural network units which represent features are added into the training process. Feature selection (learning capacity maximization by using features in analysis) was entirely automated. Human interaction was not required once the program was started and hand-crafted features were not adopted. The agent was an independent entity that used learning methods which can be applied to any similar learning system. In RGO, features were randomly added to the neural networks. In IA, features were randomly added to the neural networks, and the most fit neural networks were evolved using genetic algorithms. Tests run to determine the improvement of feature selecting agents over non-featuring selecting agents showed that using RGO as a learning method is very effective and that using IA as a learning method may be effective. After long training, the programs' learning capacities approached optimality, indicating that intelligence maximization with minimal domain knowledge may be possible using Automated Feature Selection.

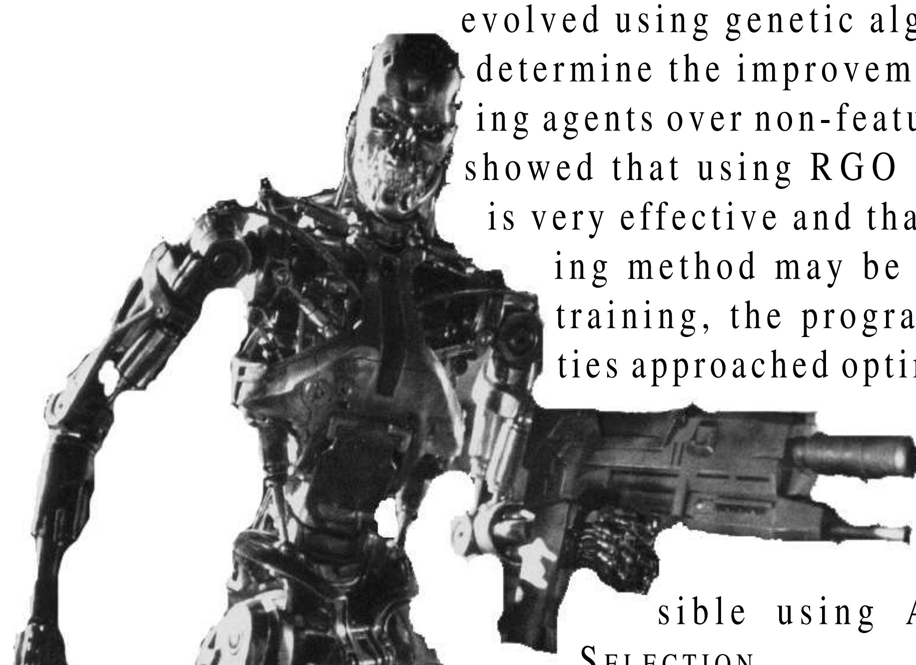


Automated Feature Selection to Maximize Learning in Artificial Intelligence

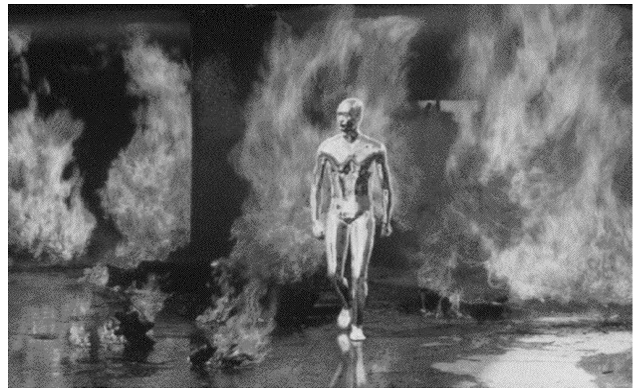
Joseph P. Turian

ABSTRACT

Traditional artificial intelligence computer programs reach a plateau beyond which learning stops. Game-playing artificial intelligence agents play at a level anywhere from beginner to expert but are never optimal, that is, they do not always make the best choice. The purpose of this project was to improve computer learning techniques so that plateau learning ability is increased. AUTOMATED FEATURE SELECTION (AFS), a new general technique, was created and implemented in two algorithms, RANDOMLY GENERATING OPERATIONS (RGO) and Incremental Addition (IA). A key component of these algorithms is that neural network units which represent features are added into the training process. Feature selection (learning capacity maximization by using features in analysis) was entirely automated. Human interaction was not required once the program was started and hand-crafted features were not adopted. The agent was an independent entity that used learning methods which can be applied to any similar learning system. In RGO, features were randomly added to the neural networks. In IA, features were randomly added to the neural networks, and the most fit neural networks were evolved using genetic algorithms. Tests run to determine the improvement of feature selecting agents over non-featuring selecting agents showed that using RGO as a learning method is very effective and that using IA as a learning method may be effective. After long training, the programs' learning capacities approached optimality, indicating that intelligence maximization with minimal domain knowledge may be possible using AUTOMATED FEATURE SELECTION.



Bibliography



[CARU94] Caruana, Rich <carauna@cs.cmu.edu> and Dayne Freitag <dayne@cs.cmu.edu>. “Greedy Attribute Selection”. Proceedings of the Eleventh International Conference on Machine Learning, pp. 28-36. http://www.cs.cmu.edu/afs/cs.cmu.edu/user/caruana/pub/ papers/ml94.ps. 1994.

[DAMI95] Damish, Mark <damish@ll.mit.edu>. “Backgammon—Frequently Asked Questions”. Last Modified: 15DEC95.

[JOHN94] John, George H. <gjohn@CS.Stanford.EDU>, Ron Kohavi <ronnyk@CS.Stanford.EDU>, and Karl Pfleger <kpfleger@CS.Stanford.EDU>. “Irrelevant Features and the Subset Selection Problem”. In Machine Learning: Proceedings of the Eleventh International Conference (ICML-94), pp. 121-129. ftp://starry.stanford.edu/pub/gjohn/papers/relevance4.ps. 1994.

[KOHA94] Kohavi, Ron <ronnyk@CS.Stanford.EDU> and Brian Frasca <brianf@CS.Stanford. EDU>. “Useful Feature Subsets and Rough Set Reduction”. To appear in The Third International Workshop on Rough Sets and Soft Computing (RSSC 94). ftp://starry.stanford.edu/pub/ronnyk/rough.ps. 1994.

[KOHA95] Kohavi, Ron <ronnyk@CS.Stanford.EDU>. “The Power of Decision Tables”. The European Conference on Machine Learning (ECML-95), 1995. ftp://starry.stanford.edu/pub/ronnyk/tables.ps. 1995.

[KOHA96] Kohavi, Ron <ronnyk@sgi.com> and George G. John <gjohn@CS.Stanford.EDU>. “Wrappers for Feature Selection”. Technical Report, Computer Science Department, Stanford University. ftp://starry.stanford.edu/pub/ronnyk/wrappers.ps. 14AUG96.

[KRUS91] Kruse, Robert L. et al. Data Structures and Program Design in C. Prentice Hall, Engelwood Cliffs. 1991.

[LIU95] Liu, Huan <liuh@iscs.nus.sg> and Rudy Setiono <rudys@iscs.nus.sg>. “Chi2: Feature Selection and Discretization of Numeric Attributes”. Proceedings of the Seventh IEEE International Conference on Tools with Artificial Intelligence. http://www.iscs.nus.sg/~liuh/tai95.ps. 1995.

[LIU196] Liu, Huan <liuh@iscs.nus.sg> and Rudy Setiono <rudys@iscs.nus.sg>. “Feature Selection and Classification — A Probabilistic Wrapper Approach”. Proceedings of the Ninth International Conference on Industrial and Engineering Applications of AI and ES. http://www.iscs.nus.sg/~liuh/wrapper.ps. 1996.

[LIU296] Liu, Huan <liuh@iscs.nus.sg> and Rudy Setiono <rudys@iscs.nus.sg>. “A Probabilistic Approach to Feature Selection — A Filter Approach”. Thirteenth International Conference on Machine Learning (ICML ’96). http://www.iscs.nus.sg/~liuh/ml96.ps. JUL96.

[PFAH95] Pfahringer, Bernhard <bernhard@ai.univie.ac.at>. “Compression-Based Feature Subset Selection”. Proceedings of the IJCAI-95 Workshop on Data Engineering for Inductive Learning, pp. 109-119. http://ai.iit.nrc.ca/DEIL/pfahringer.ps.Z. 1995.

[PIRA95] Piramuthu, Selwyn. “Feature Selection and Backpropagation Neural Networks”. International Workshop on Neural Networks — Proceedings Volume I/II (1994/1995). http://www.leidenuniv.nl/medfac/fff/groepc/chaos/helnet/selwyn.htm. 1995.

[RUSS95] Russell, Stuart and Peter Norvig. Artificial Intelligence: A Modern Approach. Prentice Hall, Engelwood Cliffs. 1995.

[SUTT88] Sutton, Richard S. “Learning to Predict by the Methods of Temporal Differences.” Machine Learning, Vol. 3, pp. 9-44. ftp://ftp.cs.umass.edu/pub/anw/pub/sutton/sutton-88.ps.gz. 4FEB1988.

[TESA91] Tesauro, Gerald J. “Practical Issues in Temporal Difference Learning”. Machine Learning, Vol. 8, No. 3-4, pp. 257-277. MAY92.

[TESA95] Tesauro, Gerald J. “Temporal Difference Learning and TD-Gammon”. Communications of the ACM, Vol. 38, No. 3. http://www.research.ibm.com/massdist/tdl.html. MAR95.

[VAFA93] Vafaie, Haleh and Kenneth De Jong. “Robust Feature Selection Algorithms.” Proceedings of the 5th IEEE International Conference on Tools for Artificial Intelligence, pp. 356-363. http://www.cs.gmu.edu/research/gag/papers/TAI93.ps. 1993.