Explorations in efficient reinforcement learning

Wiering, M.

Publication date
1999

Bibliography

Albus, J. S. (1975a). A new approach to manipulator control: The cerebellar model articulation controller (CMAC). *Dynamic Systems, Measurement and Control*, pages 220–227.

Albus, J. S. (1975b). A theory of cerebellar function. *Mathematical Biosciences*, 10:25–61.

Asada, M., Uchibe, E., Noda, S., Tawaratsumida, S., and Hosoda, K. (1994). A vision-based reinforcement learning for coordination of soccer playing behaviors. In *Proceedings of AAAI-94 Workshop on AI and A-life and Entertainment*, pages 16–21.

Atkeson, C. G., Schaal, S. A., and Moore, A. W. (1997). Locally weighted learning. *Artificial Intelligence Review*, 11:11–73.

Axelrod, R. (1984). *The evolution of cooperation.* Basic Books, New York, NY.

Baird, L. (1995). Residual algorithms: Reinforcement learning with function approximation. In Prieditis, A. and Russell, S., editors, *Machine Learning: Proceedings of the Twelfth International Conference*, pages 30–37. Morgan Kaufmann Publishers, San Francisco, CA.

Baluja, S. (1994). Population-based incremental learning: A method for integrating genetic search based function optimization and competitive learning. Technical Report CMU-CS-94-163, Carnegie Mellon University.

Barto, A. G., Bradtke, S. J., and Singh, S. P. (1995). Learning to act using real-time dynamic programming. *Artificial Intelligence*, 72:81–138.

Barto, A. G., Sutton, R. S., and Anderson, C. W. (1983). Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13:834–846.

Baum, E. (1989). A proposal for more powerful learning algortihms. *Neural Computation*, 1(2):201–207.

Baxter, J., Tridgell, A., and Weaver, L. (1997). Knightcap: A chess program that learns by combining TD(λ) with minimax search. Technical report, Australian National University, Canberra.

Bayse, K., Dean, T., and Vitter, J. (1997). Coping with uncertainty in map learning. *Machine Learning*, 29(1):65–88.

Bell, A. and Sejnowski, T. (1998). The "independent components" of natural scenes are edge filters. *Vision Research.* To appear.

Bellman, R. (1961). *Adaptive Control Processes.* Princeton University Press.

Berliner, H. (1977). Experiences in evaluation with BKG - a program that plays backgammon. In *Proceedings of IJCAI*, pages 428–433.

Berry, D. and Fristedt, B. (1985). *Bandit Problems: sequential allocation of experiments.* Chapman and Hall, London/New York.

Bertsekas, D. P. and Tsitsiklis, J. N. (1996). *Neuro-dynamic Programming.* Athena Scientific, Belmont, MA.

Boutilier, C. and Poole, D. (1996). Computing optimal policies for partially observable decision processes using compact representations. In *AAAI-1996: Proceedings of the Thirteenth National Conference on Artificial Intelligence*, pages 1168–1175, Portland, OR.

Box, G., Jenkins, G. M., and Reinsel, H. C. (1994). *Time series analysis: forecasting and control.* Prentice Hall.

Boyan, J. A. (1992). Modular neural networks for learning context-dependent game strategies. Master's thesis, University of Chicago.

Boyan, J. A. and Moore, A. W. (1995). Generalization in reinforcement learning: Safely approximating the value function. In Tesauro, G., Touretzky, D. S., and Leen, T. K., editors, *Advances in Neural Information Processing Systems 7*, pages 369–376. MIT Press, Cambridge MA.

Boyan, J. A. and Moore, A. W. (1997). Using prediction to improve combinatorial optimization search. In *Proceedings of the Sixth International Workshop on Artificial Intelligence and Statistics (AISTATS)*, page 14.

Bradtke, S. J. and Barto, A. G. (1996). Linear least-squares algorithms for temporal difference learning. *Machine Learning*, 22:33–57.

Caironi, P. V. C. and Dorigo, M. (1994). Training Q-agents. Technical Report IRIDIA-94-14, Université Libre de Bruxelles.

Campos, L. M. D., Huete, J. P., and Moral, S. (1994). Probability intervals: A tool for uncertain reasoning. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2 (2):167–196.

Cassandra, A. R. (1998). *Exact and Approximate Algorithms for Partially Observable Markov Decision Processes.* PhD thesis, Brown University, Providence, RI.

Cassandra, A. R., Kaelbling, L. P., and Littman, M. L. (April 1994). Acting optimally in partially observable stochastic domains. Technical Report CS-94-20, Brown University, Providence RI.

Chapman, D. and Kaelbling, L. P. (1991). Input generalization in delayed reinforcement learning. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI)*, volume 2, pages 726–731. Morgan Kaufman.

Chrisman, L. (1992). Reinforcement learning with perceptual aliasing: The perceptual distinctions approach. In *Proceedings of the Tenth International Conference on Artificial Intelligence*, pages 183–188. AAAI Press, San Jose, California.

Cichosz, P. (1995). Truncating temporal differences: On the efficient implementation of TD(λ) for reinforcement learning. *Journal on Artificial Intelligence*, 2:287–318.

Cliff, D. and Ross, S. (1994). Adding temporary memory to ZCS. *Adaptive Behavior*, 3:101–150.

Cohn, D. A. (1994). Neural network exploration using optimal experiment design. In Cowan, J., Tesauro, G., and Alspector, J., editors, *Advances in Neural Information Processing Systems 6*, pages 679–686. San Mateo, CA: Morgan Kaufmann.

Cramer, N. L. (1985). A representation for the adaptive generation of simple sequential programs. In Grefenstette, J., editor, *Proceedings of an International Conference on Genetic Algorithms and Their Applications*, pages 183–187, Hillsdale NJ. Lawrence Erlbaum Associates.

Crites, R. and Barto, A. (1996). Improving elevator performance using reinforcement learning. In Touretzky, D., Mozer, M., and Hasselmo, M., editors, *Advances in Neural Information Processing Systems 8*, pages 1017–1023, Cambridge MA. MIT Press.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Math. Control Signals Systems*, 2:303–314.

D'Ambrosio, B. (1989). POMDP learning using qualitative belief spaces. Technical report, Oregon State University, Corvallis.

Davies, S., Ng, A. Y., and Moore, A. W. (1998). Applying online search techniques to continuous-state reinforcement learning. In *Proceedings of the AAAI'98.*

Dayan, P. (1992). The convergence of TD(λ) for general lambda. *Machine Learning*, 8:341–362.

Dayan, P. and Hinton, G. (1993). Feudal reinforcement learning. In Lippman, D. S., Moody, J. E., and Touretzky, D. S., editors, *Advances in Neural Information Processing Systems 5*, pages 271–278. San Mateo, CA: Morgan Kaufmann.

Dayan, P. and Sejnowski, T. (1994). TD(λ): Convergence with probability 1. *Machine Learning*, 14:295–301.

Dayan, P. and Sejnowski, T. J. (1996). Exploration bonuses and dual control. *Machine Learning*, 25:5–22.

D'Epenoux, F. (1963). A probabilisitc production and inventory problem. *Management Science*, 10:98–108.

Di Caro, G. and Dorigo, M. (1998). An adaptive multi-agent routing algorithm inspired by ants behavior. In *Proceedings of PART98 - Fifth Annual Australasian Conference on Parallel and Real-Time Systems.*

Dietterich, T. (1997). Hierarchical reinforcement learning with the MAXQ value function decomposition. Technical report, Oregon State University.

Digney, B. (1996). Emergent hierarchical control structures: Learning reactive/hierarchical relationships in reinforcement environments. In Maes, P., Mataric, M., Meyer, J.-A., Pollack, J., and Wilson, S. W., editors, *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior, Cambridge, MA*, pages 363–372. MIT Press, Bradford Books.

Dijkstra, E. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271.

Dodge, Y., Fedorov, V. V., and Wynn, H. P., editors (1988). *Optimal Design and Analysis of Experiments: Proceedings of First International Conference on Optimal Design and Analysis of Experiments.* Elsevier Publishers.

Dorigo, M. and Colombetti, M. (1997). *Robot Shaping: An Experiment in Behavior Engineering.* MIT Press/Bradford Books. in press.

Dorigo, M. and Gambardella, L. M. (1997). Ant colony system: A cooperative learning approach to the traveling salesman problem. *Evolutionary Computation*, 1(1):53–66.

Dorigo, M., Maniezzo, V., and Colorni, A. (1996). The ant system: Optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics-Part B*, 26(1):29–41.

Fedorov, V. V. (1972). *Theory of optimal experiments.* Academic Press.

Friedman, J., Bentley, J., and Finkel, R. (1977). An algorithm for finding best matches in logarithmic expected time. *AMC Transactions on Mathematical Software*, 3(3):209–226.

Fritzke, B. (1994). Supervised learning with growing cell structures. In Cowan, J., Tesauro, G., and Alspector, J., editors, *Advances in Neural Information Processing Systems 6*, pages 255–262. San Mateo, CA: Morgan Kaufmann.

Gambardella, L. M., Taillard, E., and Dorigo, M. (1997). Ant colonies for the QAP. Technical Report IDSIA-4-97, IDSIA, Lugano, Switzerland. Submitted to: Journal of the Operational Research Society.

Gittins, J. C. (1989). *Multi-armed Bandit Allocation Indices.* Wiley, Chichester, NJ.

Givan, R., Leach, S., and Dean, T. (1998). Bounded parameter Markov decision processes. Technical report. Retrievable from http://www.cs.brown.edu/people/tld/home.html.

Glover, F. and Laguna, M. (1997). *Tabu Search.* Kluwer Academic Publishers.

Gordon, G. (1995a). Stable function approximation in dynamic programming. Technical Report CMU-CS-95-103, School of Computer Science, Carnegie Mellon University, Pittsburgh.

Gordon, G. (1995b). Stable function approximation in dynamic programming. In Prieditis, A. and Russell, S., editors, *Machine Learning: Proceedings of the Twelfth International Conference*, pages 261–268. Morgan Kaufmann Publishers, San Francisco, CA.

Heger, M. (1994). Consideration of risk in reinforcement learning. In *Machine Learning: Proceedings of the 11th International Conference*, pages 105–111. Morgan Kaufmann Publishers, San Francisco, CA.

Hihi, S. E. and Bengio, Y. (1996). Hierarchical recurrent neural networks for long-term dependencies. In Touretzky, D. S., Mozer, M. C., and Hasselmo, M. E., editors, *Advances in Neural Information Processing Systems 8*, pages 493–499. MIT Press, Cambridge MA.

Hinton, G. and Sejnowski, T. (1983). Optimal perceptual inference. In *Proceedings of the 1983 IEEE Conference on Computer Vision and Pattern Recognition*, pages 448–453. New york: IEEE.

Hochreiter, S. and Schmidhuber, J. H. (1997). Long short-term memory. *Neural Computation*, 9:1681–1726.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems.* University of Michigan Press, Ann Arbor.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79:2554–2558.

Humphrys, M. (1996). Action selection methods using reinforcement learning. In Maes, P., Mataric, M., Meyer, J.-A., Pollack, J., and Wilson, S. W., editors, *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior, Cambridge, MA*, pages 135–144. MIT Press, Bradford Books.

Jaakkola, T., Jordan, M. I., and Singh, S. P. (1994). On the convergence of stochastic iterative dynamic programming algorithms. *Neural Computation*, 6:1185–1201.

Jaakkola, T., Singh, S. P., and Jordan, M. I. (1995). Reinforcement learning algorithm for partially observable Markov decision problems. In Tesauro, G., Touretzky, D. S., and Leen, T. K., editors, *Advances in Neural Information Processing Systems 7*, pages 345–352. MIT Press, Cambridge MA.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., and Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Computation*, 3(1):79–87.

Jollife, I. T. (1986). *Principal Component Analysis.* New York: Springer Verlag.

Jordan, M. I. and Jacobs, R. A. (1992). Hierarchies of adaptive experts. In Moody, J. E., Hanson, S. J., and Lippmann, R. P., editors, *Advances in Neural Information Processing Systems 4*, pages 985–993. Morgan Kauffmann.

Jordan, M. I. and Rumelhart, D. E. (1990). Supervised learning with a distal teacher. Technical Report Occasional Paper #40, Center for Cognitive Science, Massachusetts Institute of Technology.

Judd, J. (1990). *Neural Network Design and the Complexity of Learning.* The MIT press, Cambridge.

Kaelbling, L. P. (1993). *Learning in Embedded Systems.* MIT Press.

Kaelbling, L. P., Littman, M. L., and Cassandra, A. R. (1995). Planning and acting in partially observable stochastic domains. Unpublished report.

Kaelbling, L. P., Littman, M. L., and Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4:237–285.

Kearns, M. and Singh, S. P. (1998). Near-optimal performance for reinforcement learning in polynomial time. Retrievable from http://www.research.att.com/~mkearns/.

Kitano, H., Asada, M., Kuniyoshi, Y., Noda, I., and Osawa, E. (1997). Robocup: The robot world cup initiative. In *Proceedings of the First International Conference on Autonomous Agents (Agents-97)*. The ACM Press.

Koenig, S. and Simmons, R. G. (1992). Complexity analysis of real-time exploration learning applied to finding shortest paths in deterministic domains. Technical Report CMU-CS-93-106, School of Computer Science, Carnegie Mellon University.

Koenig, S. and Simmons, R. G. (1996). The effect of representation and knowledge on goal-directed exploration with reinforcement learning algorithms. *Machine Learning*, 22:228–250.

Kohonen, T. (1988). *Self-Organization and Associative Memory*. Springer, second edition.

Koza, J. R. (1992). Genetic evolution and co-evolution of computer programs. In Langton, C., Taylor, C., Farmer, J. D., and Rasmussen, S., editors, *Artificial Life II*, pages 313–324. Addison Wesley Publishing Company.

Kröse, B. J. A. and van Dam, J. W. M. (1992). Adaptive state space quantisation : Adding and removing neurons. In Aleksander, I. and Taylor, J., editors, *Artificial Neural Networks, 2*, pages 619–624. North-Holland/Elsevier Science Publishers, Amsterdam.

Kröse, B. J. A. and Van de Smagt, P. (1993). An introduction to neural networks. Autonomous Systems, University of Amsterdam.

Landelius, T. (1997). *Reinforcement Learning and distributed Local Model Synthesis*. PhD thesis, Linköping Universtity, Sweden.

Lauritzen, S. and Wermuth, N. (1989). Graphical models for associations between variables some of which are qualitative and some quantitative. *Annals of Statistics*, 17:31–57.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989). Back-propagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551.

Levin, L. A. (1973). Universal sequential search problems. *Problems of Information Transmission*, 9(3):265–266.

Li, M. and Vitányi, P. M. B. (1993). *An Introduction to Kolmogorov Complexity and its Applications*. Springer.

Lin, C.-S. and Chiang, C.-T. (1997). Learning convergence of CMAC technique. *IEEE Transactions on Neural Networks*, 8(6):1281–1292.

Lin, L.-J. (1993). *Reinforcement Learning for Robots Using Neural Networks.* PhD thesis, Carnegie Mellon University, Pittsburgh.

Lin, T., Horne, B., and Giles, C. (1996). How embedded memory in recurrent neural network architectures helps learning long-term temporal dependencies. Technical Report CS-TR-3626 and UMIACS-TR-96-28, University of Maryland, College Park MD 20712.

Lindgren, K. and Nordahl, M. G. (1994). Cooperation and community structure in artificial ecosystems. *Artificial Life*, 1(1/2):15–37.

Littman, M. L. (1994a). Markov games as a framework for multi-agent reinforcement learning. In Prieditis, A. and Russell, S., editors, *Machine Learning: Proceedings of the Eleventh International Conference*, pages 157–163. Morgan Kaufmann Publishers, San Francisco, CA.

Littman, M. L. (1994b). Memoryless policies: Theoretical limitations and practical results. In Cliff, D., Husbands, P., Meyer, J. A., and Wilson, S. W., editors, *Proceedings of the International Conference on Simulation of Adaptive Behavior: From Animals to Animats 3*, pages 297–305. MIT Press/Bradford Books.

Littman, M. L. (1996). *Algorithms for Sequential Decision Making.* PhD thesis, Brown University.

Littman, M. L., Cassandra, A. R., and Kaelbling, L. P. (1995a). Learning policies for partially observable environments: Scaling up. In Prieditis, A. and Russell, S., editors, *Machine Learning: Proceedings of the Twelfth International Conference*, pages 362–370. Morgan Kaufmann Publishers, San Francisco, CA.

Littman, M. L., Dean, T. L., and Kaelbling, L. P. (1995b). On the complexity of solving Markov decision problems. In *Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence (UAI-95).*

Lovejoy, W. S. (1991). A survey of algorithms methods for partially observable Markov decision processes. *Annals of Operations Research*, 28:47–66.

Luke, S., Hohn, C., Farris, J., Jackson, G., and Hendler, J. (1997). Co-evolving soccer softbot team coordination with genetic programming. In *Proceedings of the First International Workshop on RoboCup, at the International Joint Conference on Artificial Intelligence (IJCAI-97).*

Mahadevan, S. (1996). Sensitive discount optimality: Unifying discounted and average reward reinforcement learning. In Saitta, L., editor, *Machine Learning: Proceedings of the Thirteenth International Conference*, pages 328–336. Morgan Kaufmann Publishers, San Francisco, CA.

Martin, M. (1998). *Reinforcement Learning for Embedded Agents facing Complex tasks.* PhD thesis, Universitat Politecnica de Catalunya, Barcelona.

Martinetz, T. and Schulten, K. (1991). A "neural-gas" network learns topologies. In Kohonen, T., Mäkisara, K., Simula, O., and Kangas, J., editors, *Artificial Neural Networks*, pages 397–402. Elsevier Science Publishers B.V., North-Holland.

Mataric, M. J. (1994). *Interaction and Intelligent Behavior.* PhD thesis, Massacusetts institute of Technology.

Matsubara, H., Noda, I., and Hiraki, K. (1996). Learning of cooperative actions in multi-agent systems: a case study of pass play in soccer. In Sen, S., editor, *Working Notes for the AAAI-96 Spring Symposium on Adaptation, Coevolution and Learning in Multi-agent Systems*, pages 63–67.

McCallum, R. A. (1993). Overcoming incomplete perception with utile distinction memory. In *Machine Learning: Proceedings of the Tenth International Conference*, pages 190–196. Morgan Kaufmann, Amherst, MA.

McCallum, R. A. (1996). Learning to use selective attention and short-term memory in sequential tasks. In Maes, P., Mataric, M., Meyer, J.-A., Pollack, J., and Wilson, S. W., editors, *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior, Cambridge, MA*, pages 315–324. MIT Press, Bradford Books.

McDonald, M. A. F. and Hingston, P. (1994). Approximate discounted dynamic programming is unreliable. Technical Report 94/6, Department of Computer Science, The University of Western Australia, Crawley, WA.

Moore, A. W. (1991). *Efficient Memory-based Learning for Robot Control.* PhD thesis, University of Cambridge.

Moore, A. W. (1998). Personal communication.

Moore, A. W. and Atkeson, C. G. (1993). Prioritized sweeping: Reinforcement learning with less data and less time. *Machine Learning*, 13:103–130.

Moore, A. W. and Atkeson, C. G. (1995). The parti-game algorithm for variable resolution reinforcement learning in multidimensional state-spaces. *Machine Learning*, 21:3:199–233.

Moore, A. W., Atkeson, C. G., and Schaal, S. A. (1997). Locally weighted learning for control. *Artificial Intelligence Review*, 11:75–113.

Munos, R. (1996). A convergent reinforcement learning scheme in the continuous case: the finite element reinforcement learning. In Saitta, L., editor, *Machine Learning: Proceedings of the Thirteenth International Conference*, pages 337–345. Morgan Kaufmann Publishers, San Francisco, CA.

Myerson, R. (1991). *Game Theory.* Harvard University Press.

Nguyen and Widrow, B. (1989). The truck backer-upper: An example of self learning in neural networks. In *IEEE/INNS International Joint Conference on Neural Networks, Washington, D.C.*, volume 1, pages 357–364.

Nilsson, N. J. (1971). *Problem-Solving Methods in Artificial Intelligence.* McGraw-Hill.

Nowlan, S. (1991). *Soft Competitive Adaption: Neural Network Learning Algorithms based on Fitting Statistical Mixtures.* PhD thesis, Carnegie Mellon University, Pittsburgh.

Oja, E. and Karhunen, J. (1995). Signal separation by nonlinear hebbian learning. In Palaniswami, M., Attikiouzel, Y., Marks II, R., Fogel, D., and Fukuda, T., editors, *Computational Intelligence - a Dynamic System Perspective*, pages 83–97. IEEE Press, New York.

Okabe, A., Boots, B., and Sugihara, K. (1990). *Spatial Tesselations - Concepts and applications of Voronoi diagrams.* Wiley and Sons, New York.

Omohundro, S. M. (1988). Foundations of geometric learning. Technical Report UIUCDCS-R-88-1 408, University of Illinois, Department of Computer Science.

Omohundro, S. M. (1989). Five balltree construction algorithms. Technical Report TR-89-063, International Computer Science Institute, Berkeley, CA.

Omohundro, S. M. (1991). Bumptrees for efficient function, constraint, and classification learning. In Lippman, D. S., Moody, J. E., and Touretzky, D. S., editors, *Advances in Neural Information Processing Systems 3*, pages 693–699. San Mateo, CA: Morgan Kaufmann.

Parr, R. and Russell, S. (1995). Approximating optimal policies for partially observable stochastic domains. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-95)*, pages 1088–1094. Morgan Kaufmann.

Parr, R. and Russell, S. (1997). Reinforcement learning with hierarchies of machines. In *Advances in Neural Information Processing Systems 11.*

Peek, N. B. (1997). Predictive probabilistic models for treatment planning in paediatric cardiology. In *Proceedings of CESA'98 IMACS Multiconference (Computational Engineering in Systems Applications), Symposium on Signal Processing and Cybernetics.*

Peng, J. and Williams, R. (1996). Incremental multi-step Q-learning. *Machine Learning*, 22:283–290.

Peng, J. and Williams, R. J. (1993). Efficient learning and planning with the DYNA framework. *Adaptive Behavior*, 1:437–454.

Pineda, F. (1997). Mean-field theory for batched TD(λ). *Neural Computation*, 9(7):1404–1419.

Pollack, J. and Blair, A. (1996). Why did TD-Gammon work. In Touretzky, D., Mozer, M., and Hasselmo, M., editors, *Advances in Neural Information Processing Systems 8*, pages 10–16, Cambridge MA. MIT Press.

Precup, D. and Sutton, R. (1998). Theoretical results on reinforcement learning with temporally abstract options. In *Proceedings of the Tenth European Conference on Machine Learning (ECML'98).*

Preparate, F. P. and Shamos, M. I. (1985). *Computational Geometry: an Introduction.* Springer Verlag, New York.

Prescott, T. (1994). *Exploration in Reinforcement and Model-based Learning.* PhD thesis, University of Sheffield.

Press, W., Teukolsky, S., Vettering, W., and Flannery, B. (1988). *Numerical recipes in C.* Cambridge University Press.

Rao, C. and Mitra, S. (1971). *Generalized Inverse of Matrices and Its Applications.* Wiley, New York.

Rechenberg, I. (1971). Evolutionsstrategie - Optimierung technischer Systeme nach Prinzipien der biologischen Evolution. Dissertation. Published 1973 by Fromman-Holzboog.

Rechenberg, I. (1989). Evolution strategy: Nature's way of optimization. In Bergmann, editor, *Methods and Applications, Possibilities and Limitations*, pages 106–126. Lecture notes in Engineering.

Resnick, S. (1992). *Adventures in stochastic processes.* Birkhaeuser Verlag.

Ring, M. B. (1994). *Continual Learning in Reinforcement Environments.* PhD thesis, University of Texas, Austin, Texas.

Ron, D., Singer, Y., and Tishby, N. (1994). Learning probabilistic automata with variable memory length. In Aleksander, I. and Taylor, J., editors, *Proceedings Computational Learning Theory.* ACM Press.

Roth, A. and Erev, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior*, 8:164–212. Special issue: Nobel Symposium.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning internal representations by error propagation. In *Parallel Distributed Processing*, volume 1, pages 318–362. MIT Press.

Rummery, G. and Niranjan, M. (1994). On-line Q-learning using connectionist sytems. Technical Report CUED/F-INFENG-TR 166, Cambridge University, UK.

Sahota, M. (1993). Real-time intelligent behaviour in dynamic environments: Soccer-playing robots. Master's thesis, University of British Columbia.

Sałustowicz, R. P. and Schmidhuber, J. H. (1997). Probabilistic incremental program evolution. *Evolutionary Computation*, 5(2):123–141.

Sałustowicz, R. P., Wiering, M. A., and Schmidhuber, J. H. (1997a). Evolving soccer strategies. In *Proceedings of the Fourth International Conference on Neural Information Processing (ICONIP'97)*, pages 502–506. Springer-Verlag Singapore.

Sałustowicz, R. P., Wiering, M. A., and Schmidhuber, J. H. (1997b). On learning soccer strategies. In Gerstner, W., Germond, A., Hasler, M., and Nicoud, J.-D., editors, *Proceedings of the Seventh International Conference on Artificial Neural Networks (ICANN'97)*, volume 1327 of *Lecture Notes in Computer Science*, pages 769–774. Springer-Verlag Berlin Heidelberg.

Sałustowicz, R. P., Wiering, M. A., and Schmidhuber, J. H. (1998). Learning team strategies: Soccer case studies. *Machine Learning*, 33(2/3).

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal on Research and Development*, 3:210–229.

Sandholm, T. W. and Crites, R. H. (1995). On multiagent Q-learning in a semi-competitive domain. In Weiss, G. and Sen, S., editors, *IJCAI'95 Workshop: Adaption and Learning in Multi-Agent Systems*, pages 164–176. Springer-Verlag.

Santamaria, J. C., Sutton, R. S., and Ram, A. (1996). Experiments with reinforcement learning in problems with continuous state and action spaces. Technical Report COINS 96-088, Georgia Institute of Technology, Atlanta.

Schmidhuber, J. H. (1991a). Curious model-building control systems. In *Proceedings of the International Joint Conference on Neural Networks, Singapore*, volume 2, pages 1458–1463. IEEE.

Schmidhuber, J. H. (1991b). Learning to generate sub-goals for action sequences. In Kohonen, T., Mäkisara, K., Simula, O., and Kangas, J., editors, *Artificial Neural Networks*, pages 967–972. Elsevier Science Publishers B.V., North-Holland.

Schmidhuber, J. H. (1991c). A possibility for implementing curiosity and boredom in model-building neural controllers. In Meyer, J. A. and Wilson, S. W., editors, *Proceedings of the International Conference on Simulation of Adaptive Behavior: From Animals to Animats*, pages 222–227. MIT Press/Bradford Books.

Schmidhuber, J. H. (1991d). Reinforcement learning in Markovian and non-Markovian environments. In Lippman, D. S., Moody, J. E., and Touretzky, D. S., editors, *Advances in Neural Information Processing Systems 3*, pages 500–506. San Mateo, CA: Morgan Kaufmann.

Schmidhuber, J. H. (1992). Learning complex, extended sequences using the principle of history compression. *Neural Computation*, 4(2):234–242.

Schmidhuber, J. H. (1996). A general method for incremental self-improvement and multi-agent learning in unrestricted environments. In Yao, X., editor, *Evolutionary Computation: Theory and Applications*. Scientific Publ. Co., Singapore.

Schmidhuber, J. H. (1997). What's interesting? Technical Report IDSIA-35-97, IDSIA.

Schmidhuber, J. H., Zhao, J., and Schraudolph, N. N. (1997a). Reinforcement learning with self-modifying policies. In Thrun, S. and Pratt, L., editors, *Learning to learn*. Kluwer.

Schmidhuber, J. H., Zhao, J., and Wiering, M. A. (1996). Simple principles of metalearning. Technical Report IDSIA-69-96, IDSIA.

Schmidhuber, J. H., Zhao, J., and Wiering, M. A. (1997b). Shifting inductive bias with success-story algorithm, adaptive Levin search, and incremental self-improvement. *Machine Learning*, 28:105–130.

Schneider, J. G. (1997). Exploiting model uncertainty estimates for safe dynamic control learning. In Mozer, M. C., Jordan, M. I., and Petsche, T., editors, *Advances in Neural Information Processing Systems 9*, pages 1047–1053. MIT Press/Bradford Books, Cambridge.

Schraudolph, N. N., Dayan, P., and Sejnowski, T. J. (1994). Temporal difference learning of position evaluation in the game of go. In Cowan, J. D., Tesauro, G., and Alspector, J., editors, *Advances in Neural Information Processing Systems*, volume 6, pages 817–824. Morgan Kaufmann, San Francisco.

Schwartz, A. (1993). A reinforcement learning method for maximizing undiscounted rewards. In *Machine Learning: Proceedings of the Tenth International Conference*, pages 298–305. Morgan Kaufmann, Amherst, MA.

Singh, S. P. (1992). The efficient learning of multiple task sequences. In Moody, J., Hanson, S., and Lippman, R., editors, *Advances in Neural Information Processing Systems 4*, pages 251–258, San Mateo, CA. Morgan Kaufmann.

Singh, S. P. and Sutton, R. S. (1996). Reinforcement learning with replacing elibibility traces. *Machine Learning*, 22:123–158.

Singh, S. P. and Yee, R. C. (1994). An upper bound on the loss from approximate optimal-value functions. *Machine Learning*, 16.

Sondik, E. J. (1971). *The Optimal Control of Partially Observable Markov Decision Processes.* PhD thesis, Standford, California.

Steels, L. (1997). Constructing and sharing perceptual distinctions. In van Someren, M. and Widmer, G., editors, *Machine Learning: Proceedings of the ninth European Conference*, pages 4–13. Springer-Verlag, Berlin Heidelberg.

Stone, P. and Veloso, M. (1996). Beating a defender in robotic soccer: Memory-based learning of a continuous function. In Tesauro, G., Touretzky, D. S., and Leen, T. K., editors, *Advances in Neural Information Processing Systems 8*, pages 896–902. MIT Press, Cambridge MA.

Stone, P. and Veloso, M. (1998). Team-partitioned opaque-transition reinforcement learning. In *Proceedings of the Conference on automated learning and discovery (CONALD'98): Robot Exploration and Learning.* Carnegie Mellon University, Pittsburgh.

Storck, J., Hochreiter, S., and Schmidhuber, J. H. (1995). Reinforcement driven information acquisition in nondeterministic environments. In *Proceedings of the International Conference on Artificial Neural Networks*, volume 2, pages 159–164. EC2 & Cie, Paris.

Sutton, R. S. (1984). *Temporal Credit Assignment in Reinforcement Learning.* PhD thesis, University of Massachusetts, Dept. of Comp. and Inf. Sci.

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Machine Learning*, 3:9–44.

Sutton, R. S. (1990). Integrated architectures for learning, planning and reacting based on dynamic programming. In *Machine Learning: Proceedings of the Seventh International Workshop.*

Sutton, R. S. (1995). TD models: Modeling the world at a mixture of time scales. In Prieditis, A. and Russell, S., editors, *Machine Learning: Proceedings of the Twelfth International Conference*, pages 531–539. Morgan Kaufmann Publishers, San Francisco, CA.

Sutton, R. S. (1996). Generalization in reinforcement learning: Successful examples using sparse coarse coding. In Touretzky, D. S., Mozer, M. C., and Hasselmo, M. E., editors, *Advances in Neural Information Processing Systems 8*, pages 1038–1045. MIT Press, Cambridge MA.

Sutton, R. S., Precup, D., and Singh, S. P. (1998). Between MDPs and semi-MDPs: Learning, planning, learning and sequential decision making. Technical Report COINS 89-95, University of Massachusetts, Amherst.

Teller, A. (1994). The evolution of mental models. In Kinnear, Jr., K. E., editor, *Advances in Genetic Programming*, pages 199–219. MIT Press.

Tesauro, G. (1992). Practical issues in temporal difference learning. In Lippman, D. S., Moody, J. E., and Touretzky, D. S., editors, *Advances in Neural Information Processing Systems 4*, pages 259–266. San Mateo, CA: Morgan Kaufmann.

Tesauro, G. (1995). Temporal difference learning and TD-Gammon. *Communications of the ACM*, 38:58–68.

Tham, C. (1995). Reinforcement learning of multiple tasks using a hierarchical CMAC architecture. *Robotics and Autonomous Systems*, 15(4):247–274.

Thrun, S. (1992). Efficient exploration in reinforcement learning. Technical Report CMU-CS-92-102, Carnegie-Mellon University.

Thrun, S. (1995). Learning to play the game of chess. In Tesauro, G., Touretzky, D., and Leen, T., editors, *Advances in Neural Information Processing Systems 7*, pages 1069–1076. San Fransisco, CA: Morgan Kaufmann.

Thrun, S. (1998). Learning metric-topological maps for indoor mobile robot navigation. *Artificial Intelligence Journal*, 99(1):21–71.

Thrun, S. and Möller, K. (1992). Active exploration in dynamic environments. In Lippman, D. S., Moody, J. E., and Touretzky, D. S., editors, *Advances in Neural Information Processing Systems 4*, pages 531–538. San Mateo, CA: Morgan Kaufmann.

Thrun, S. and Schwartz, A. (1995). Finding structure in reinforcement learning. In Tesauro, G., Touretzky, D. S., and Leen, T. K., editors, *Advances in Neural Information Processing Systems 7*, pages 385–392. MIT Press, Cambridge MA.

Trovato, K. (1996). *A* Planning in Discrete Configuration Spaces of Autonomous Systems.* PhD thesis, University of Amsterdam.

Tsitsiklis, J. N. (1994). Asynchronous stochastic approximation and Q-learning. *Machine Learning*, 16:185–202.

Tsitsiklis, J. N. and Van Roy, B. (1996). An analysis of temporal-difference learning with function approximation. Technical Report LIDS-P-2322, Cambrdge,MA: MIT Laboratory for Information and Decision Systems.

Van Dam, J. W. M. (1998). *Environmental Modelling for Mobile Robots: Neural Learning for Sensor Fusion.* PhD thesis, University of Amsterdam, The Netherlands.

Van de Smagt, P. (1995). *Visual robot arm guidance using neural networks.* PhD thesis, University of Amsterdam, The Netherlands.

Van der Wal, J. (1981). *Stochastic Dynamic Programming.* Number 139 in Mathematical Centre tracts. Mathematisch Centrum, Amsterdam.

Van Emde Boas, P., Kaas, R., and Zijlstra, E. (1977). Design and implementation of an efficient priority queue. *Mathematical Systems Theory*, 10:99–127.

Vennix, J. A. M. (1996). Systeemdynamica methode for strategie-ontwikkeling. Technical report, Faculteit der Beleidswetenschappen, Katholieke Universiteit Nijmegen.

Watkins, C. J. C. H. (1989). *Learning from Delayed Rewards.* PhD thesis, King's College, Cambridge, England.

Watkins, C. J. C. H. and Dayan, P. (1992). Q-learning. *Machine Learning*, 8:279–292.

Werbos, P. J. (1974). *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences.* PhD thesis, Harvard University.

Whitehead, S. (1992). *Reinforcement Learning for the adaptive control of perception and action.* PhD thesis, University of Rochester.

Widrow, B. and Hoff, M. E. (1960). Adaptive switching circuits. *1960 IRE WESCON Convention Record*, 4:96–104. New York: IRE. Reprinted in Anderson and Rosenfeld [1988].

Wiering, M. A. (1995). *TD Learning of Game Evaluation Functions with Hierarchical Neural Architectures.* Master's thesis, Department of Computer Systems, University of Amsterdam.

Wiering, M. A. and Dorigo, M. (1998). Learning to control forest fires. In Haasis, H.-D. and Ranze, K. C., editors, *Proceedings of the 12th international Symposium on "Computer Science for Environmental Protection"*, volume 18 of *Umweltinformatik Aktuell*, pages 378–388, Marburg. Metropolis Verlag.

Wiering, M. A., Salustowicz, R. P., and Schmidhuber, J. H. (1998). CMAC models learn to play soccer. In Niklasson, L., Bodén, M., and Ziemke, T., editors, *Proceedings of the 8th International Conference on Artificial Neural Networks (ICANN'98)*, volume 1, pages 443–448. Springer-Verlag, London.

Wiering, M. A. and Schmidhuber, J. H. (1996). Solving POMDPs with Levin search and EIRA. In Saitta, L., editor, *Machine Learning: Proceedings of the Thirteenth International Conference*, pages 534–542. Morgan Kaufmann Publishers, San Francisco, CA.

Wiering, M. A. and Schmidhuber, J. H. (1997). HQ-learning. *Adaptive Behavior*, 6(2):219–246.

Wiering, M. A. and Schmidhuber, J. H. (1998a). Efficient model-based exploration. In Meyer, J. A. and Wilson, S. W., editors, *Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior: From Animals to Animats 6*, pages 223–228. MIT Press/Bradford Books.

Wiering, M. A. and Schmidhuber, J. H. (1998b). Fast online Q(λ). *Machine Learning Journal.*

Williams, R. J. and Baird, L. C. (1993). Tight performance bounds on greedy policies based on imperfect value function. Technical Report NU-CCS-93-14, College of Computer Science, Northeastern University, Boston, MA.

Wilson, S. (1994). ZCS: A zeroth level classifier system. *Evolutionary Computation*, 2:1–18.

Wilson, S. (1995). Classifier fitness based on accuracy. *Evolutionary Computation*, 3(2):149–175.

Zhang, N. L. and Liu, W. (1996). Planning in stochastic domains: Problem characteristics and approximation. Technical Report HKUST-CS96-31, Hong Kong University of Science and Technology.

Zhao, J. and Schmidhuber, J. H. (1996). Incremental self-improvement for life-time multi-agent reinforcement learning. In Maes, P., Mataric, M., Meyer, J.-A., Pollack, J., and Wilson, S. W., editors, *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior, Cambridge, MA*, pages 516–525. MIT Press, Bradford Books.