iCORE ANNUAL REPORT 2008

iCORE CPE GRANT CPE9

DR. RICHARD SUTTON

REINFORCEMENT LEARNING AND ARTIFICIAL INTELLIGENCE

ICORE ANNUAL REPORT MARCH 31, 2008

EXECUTIVE SUMMARY

A vast number of problems of economic and scientific interest involve sequences of actions where the effect of one action influences the expected utility of subsequent actions. These sequential decision problems include such diverse applications as inventory management, the control of robots and industrial processes, playing backgammon, and planning under uncertainty, all of which are made more challenging because of their sequential and stochastic aspects. Many problems in robotics and artificial intelligence are also of this nature, as indeed are most of the decision-making and planning problems faced by people and animals in their daily lives. Reinforcement learning is a new body of theory and techniques for solving sequential decision problems, based on classical methods such as dynamic programming and inspired by animal learning theory, that enables larger and more such diverse problems to be solved.

The objectives of the reinforcement learning and artificial intelligence (RLAI) research program are to create new methods for reinforcement learning that remove some of the limitations on its widespread application, and to develop reinforcement learning as a model of intelligence that could approach human abilities. These objectives are pursued through mathematics, through computational experiments, through the development of robotic systems, and through the development and testing of computational models of natural learning processes.

The research team consists of about 50 members, 30 of which are graduate students and, of those, 15 of which are recipients of major scholarships. The output of the research program has remained strong, with 32 papers published or accepted for publication in highly-refereed archival venues during the reporting period. Seven PhD and four MSc students were graduated.

The primary focus of the research program has been on how intelligent machines represent their knowledge of the world. The key question is how to organize the knowledge such that it can be verified, learned, and used autonomously without continual tending by human experts. This project has pursued an unusual approach in which knowledge is expressed in terms of the machines' sensors and actuators, thereby enabling it to be compared directly to experiential data. Substantial further progress was made this year toward formalizing the core learning algorithms and developing planning algorithms.

Highlights of the research program this year include participating in the development of the world's best program for Computer Go and creating the software for the next generation of scientific competitions in reinforcement learning. Over 140 teams from around the world have registered and downloaded our software in the run up to the competition and workshop to be held this July in Helsinki, Finland.

RESEARCH PROGRAM OVERVIEW

The iCORE research program in reinforcement learning and artificial intelligence (RLAI) pursues an approach to artificial intelligence and other engineering problems in which they are formulated as Markov Decision Processes (MDPs) and approximately solved using reinforcement learning. Reinforcement learning is a new body of theory and techniques for solving MDPs that has been developed in the last 20 years primarily within the machine learning and optimal-control research communities. Reinforcement learning researchers have developed novel methods to approximate solutions to MDPs that are too large or too ill-defined for classical solution methods such as dynamic programming. For example, reinforcement-learning methods have obtained the best known solutions in such diverse automation applications as helicopter flying, elevator scheduling, playing backgammon, and resource-constrained scheduling. The objectives of the RLAI research program are to create new methods for reinforcement learning that remove some of the limitations on its widespread application and to develop reinforcement learning as a model of intelligence that could approach human abilities.

The iCORE research proposal for the RLAI project identified four focus areas for investigation, on each of which substantial progress was made this year.

- One proposed focus was on the role of state in intelligent systems, in particular, on planning in large state spaces.

  Planning refers to the use of state representations and corresponding models of the world to anticipate the consequences of alternative courses of action and to pick among them. Planning with the kind of function approximation needed for large state spaces is largely an open problem. One appealing approach, called *Dyna*, is to create hypothetical experience using the world model and then learn from it using reinforcement learning algorithms just as if it had actually occurred.

  Last year we developed the first Dyna-style planning methods extended to use linear function approximation. This year we substantially extended and refined that work, including foundational theoretical results on the convergence and stability of linear Dyna-style planning with expectation-based models of the world. These results motivate and validate a new approach to making planning with linear models more computationally efficient. This approach is an extension of a technique known as *prioritized sweeping* such that is applies to features rather than to states, making it applicable to problems with much larger state spaces.

  We have also explored planning in very large state spaces in our application to Computer Go (discussed separately below). This has proved another breakthrough year for Computer Go, and the RLAI project has continued to play a leading role.

- A second proposed focus of the research program was on building powerful, expressive representations of the world, in particular, on the extension of reinforcement learning beyond the flat and low-level representations commonly used

with MDPs to the more flexible, structured, and higher-level representations used in classical artificial intelligence systems. The ability to represent knowledge about possible courses of action at a multiplicity of inter-related temporal scales vastly increases the generality and range of application of reinforcement learning methods. Last year we combined the theory of options (one simple way of formalizing temporal abstraction) with temporal-difference networks and explored corresponding learning algorithms.

Anna Koop's MSc thesis summarizes our main areas of progress toward this focus. Foremost among these is the exploration of the interplay of state and learning. If the state space is large and the resources of the learning system are limited, then it can pay to learn continually, tracking the locally best behavior rather than learning a single universal solution. We show that this idea, which we term *temporal coherence*, can lead to computationally superior learning algorithms; it may also shed light on the development of the object concept in infants.

- A third proposed focus of the research program was broadly concerned with approximation and generalization in reinforcement learning. Approximation is required in all large-scale applications, yet is incompletely understood in both theory and practice. In addition to foundational theoretical and algorithmic work, much of our work to date has focused on developing software to improve practice, including the development of RL-Glue, the RL toolkit, and the RL-Library.

  This year our software efforts have continued and intensified in almost year-long preparation for the next international reinforcement learning competition, to be held this summer in Helsinki, Finland. The RLAI project will participate to some extent as a competitor, but our primary role will be as the developer of the competition software and test problems. For the first time, the test problems are being made available a full six months before the competition, which greatly increases potential participation. We are committed to making this year's competition the best ever. Our competition software builds on our previous work with the RL-Glue interface and the RL-Library research repository, and helps reaffirm these as de-facto standards.

- The fourth proposed area of investigation for the RLAI project was the demonstration of advances in robotics applications. Our robotics laboratory has a variety of robots, including *Kato,* a robot Segway. In past years we have used Kato to demonstrate robot *geo-caching*—the locating, usually by people, of a hidden cache through knowledge only of its GPS coordinates—and have explored the use of *sensor bootstrapping*—defining the meaning of one sensor in terms of its dynamic, causal relationship to another sensor.

  This year we have built on past work with Kato to produce an integrated system capable of autonomous navigation over several kilometers without terrain preparation (other than planting the cache). Kato is able to thread its way around a wide variety of obstacles, into and out of dead-ends, while avoiding collisions with pedestrians. Kato uses a combination of color vision, laser range-finding, GPS, and dead reckoning.

RESEARCH PROJECTS

This section describes in more detail a few of the research projects making up the research program.

Computer Go

The ancient oriental game of Go has long been a challenge to artificial intelligence. The techniques that have worked so well in chess and in so many real-world applications seemed to have no traction in Go because of its large branching factor, which makes traditional search impossible. After decades of research, the best Computer Go programs were still no challenge to weak amateur players. Two years ago it looked likely that computers might never play a strong game of Go. In these last two years, however, the world of Computer Go has undergone a revolution, in large part because of the incorporation of sample-based search (planning) and learning methods influenced by reinforcement learning research. RLAI project members have played key roles in these developments.

The revolution in Computer Go research was initiated by research by principal investigator Csaba Szepesvari, with Kocsis Levente, just prior to joining the RLAI team. Their algorithm, known as UCT, is a sampling-based search algorithm for large, discrete action spaces with a hierarchical structure. UCT was applied to Computer Go by many researchers, leading to a huge jump in performance. In competitions on the standard Computer Go Server, all of the top ten programs are now based on UCT, and the best non-UCT program is 300 rating points worse than the tenth best UCT program, a difference corresponding to about a ten-to-one chance of winning. The Computer Go revolution and Szepesvari's role in it was recognized in an article in Scientific American ("Silicon Smackdown," by Karen A. Frenkel, June 2007, Vol. 296, Issue 6).

This year the program *MoGo*, developed by Sylvain Gelly (University of Paris South) with contributions by RLAI team member David Silver, reached a major landmark in Computer Go by becoming the first ever program to win a game against a professional Go player under tournament conditions. The win was against Guo Juan, a 5-dan professional player and probably the strongest Go player in Europe. Two years ago no one would have imagined that this could possibly happen for decades. *MoGo*'s victory against Guo Juan was in a smaller version of Go played on a 9x9 board. This form of Go is commonly played by people and is the standard in Computer Go research. *MoGo* also won the Gold medal at the 2007 Computer Olympiad for full 19x19 Go.

Computational Models of Animal Learning

Reinforcement learning is studied in psychology and neuroscience as well as in engineering and artificial intelligence. In psychology, reinforcement-learning methods are

important models of elemental learning processes in animals, such as classical conditioning. In neuroscience, reinforcement-learning methods are the dominant models of reward systems in the brain, in particular of the dopamine system. Working with Professor Jim Kehoe of the University of New South Wales, one of the world's foremost experts in animal learning psychology, we have been exploring ways in which natural learning systems might provide insights into computational reinforcement learning algorithms, as has happened in the past, leading to some of the most effective modern algorithms such as TD(λ) and Q-learning.

This year we have developed a new computational model of classical conditioning and related it to both new behavioral data obtained in Kehoe's laboratory and to neurophysiological data from the literature on the dopamine reward system. The animal experiments and models are meant to shed light on a fundamental problem in reinforcement learning as it might be used in real-time applications: how is the time series of inputs (sensor data or stimuli) represented to the learning algorithm? For any punctate external signal there must be a temporally extended internal representation; we model this as a sequence of temporally extended internal "micro-stimuli" with a range of delays and proportional dispersions. Micro-stimuli turn a simple sensory event into a temporally extended and multi-component representation that enables precise timing of responses. We have shown that using this representation enables temporal-difference models of the dopamine system to make a qualitatively better match to observed data.

Gradient-descent Methods for Off-policy Learning

The problem of off-policy learning is a recurring challenge to the field of reinforcement learning, and is a particular concern to the RLAI project because of its implications for abstraction methods. Off-policy learning is learning about one way of behaving while actually behaving in some other way. The problem of off-policy learning is that classical bootstrapping RL methods, such as Q-learning, TD(λ), and dynamic programming, can become unstable during off-policy learning if function approximation is used. As both bootstrapping and function approximation are thought to be essential for large-scale applications, and off-policy learning is currently seen as necessary for learning temporally abstract system models, this instability is a key stumbling block to extending reinforcement learning abilities.

This year we have introduced two new classes of off-policy learning methods with appealing convergence properties. The *gradient temporal-difference* (GTD) methods are almost the perfect solution: they are guaranteed to be stable and asymptotically correct on any problem and for any function approximator, and their computational complexity is linear in the number of parameters in the approximator. GTD methods are the first and currently the only known methods to have these desirable properties. However, their rate of convergence may be significantly slower than comparable on-policy methods. The second new class of methods we are examining are the *correlation-change temporal-difference* (CCTD) methods. The computational complexity of these methods is greater than that of the GTD methods—quadratic rather than linear in the number of parameters—and is similar to that of existing stable off-policy methods. However,

compared to previous off-policy methods CCTD methods work with a more general class of function approximators and are more applicable to online control settings, where policies and value functions may change.

Convergence of Actor-Critic Algorithms

Actor-critic reinforcement learning methods are based on the simultaneous online estimation of the parameters of two structures, called the *actor* and the *critic*. The actor corresponds to a policy or control law, mapping states to actions. The critic corresponds to a value function, mapping states to expected cumulative future reward. Thus, the critic addresses a problem of prediction, whereas the actor is concerned with control. These problems are separable, but are solved simultaneously to find an optimal policy, as in policy iteration. One reason actor-critic methods are appealing is that, by explicitly representing the policy class, they can more easily be biased toward a particular kinds of policy, which can be important in applications where safety is a concern, or in providing prior knowledge. Another reason is that the mapping onto biological implementations seems more straightforward for actor-critic architectures. The intuitive appeal and other advantages of actor-critic methods make them the methods of choice in a variety of reinforcement learning contexts. This year we have extended the theory of actor-critic methods to include convergence proofs for several new methods using temporal-difference learning. Our results are the first to prove convergence when temporal-difference learning is used in both the actor and the critic.

OBJECTIVES FOR NEXT YEAR

The next year will include the five-year anniversary of the RLAI project and with that there will be some adjustment of its objectives. Overall, we will focus on three main interrelated areas. The first is extensions of core reinforcement learning theory and algorithms; there are many open problems in reinforcement learning, and the RLAI project will seek to solve them as opportunities arise. The second area is the extension of reinforcement learning ideas to address the more ambitious goals of artificial intelligence. There is a natural transition from the more advanced reinforcement learning methods to mechanisms for knowledge representation, search, and human-level reasoning. A major objective for the project is to explore, implement, and illustrate these relationships. The third main area of RLAI focus will be on designing algorithms and software particularly suited for applications, and on several specific application areas such as robotics and Computer Go.

In the area of core reinforcement learning algorithms, we will focus on function approximation, online learning, off-policy learning and gradient-descent methods. We seek to develop a new class of temporal-difference learning algorithms based on gradient descent that can be efficiently trained on-policy. In initial results we have identified the GTD and CCTD algorithms. We plan to test and develop these further. An online learning algorithm is one that improves its behavior during the normal operation of the system to be controlled or predicted. Current online learning solutions for MDPs are

limited to small finite domains or make strong assumptions (e.g., that the uncertainty is parametric) that seriously limit their applicability. We will seek to address the following issues: (1) understanding what makes efficient online learning possible; (2) characterizing the behavior of online learning algorithms; and (3) developing online learning algorithms that are efficient in terms of both data and computation. We propose to study online learning using a non-parametric approach.

In the area of extending reinforcement learning to artificial intelligence, we will focus on extending model-based reinforcement learning to incorporate abstractions in the form of options, predictive state representations, and temporal-difference networks. The immediate extensions of our current work with Dyna-style planning will include comparisons with our own iLSTD and CCTD algorithms, the use of semi-linear models (linear followed by an exponential form nonlinearity), and more thorough experimental work.

One of the application focuses next year will be on a new robot project explicitly designed to drive research on experience-grounded knowledge representation. In this project we will custom design and build a series of small mobile robots outfitted with an unusually rich set of sensors, including sensors for touch, acceleration, motion, sound, vision, and several kinds of proximity. The initial objective will be for the robot to form an extended multi-level model of the relationships among its sensors and between its sensors and its actuators. We have proposed that higher-level knowledge can be grounded in raw data of sensations and actions; this robotic platform will challenge and inspire us to see if it can really be done. We also plan to use this platform as a test case for rapid learning and for the use of reinforcement learning by non-experts. We would like a person whose has no training to be able to teach the system new ways of behaving in an intuitive manner much as one might train a particularly cooperative dog.

RESEARCH TEAM MEMBERS AND CONTRIBUTIONS

Team Leader

Name	Role / Topic	Awards / Special Info
Richard Sutton	Principal investigator	NSERC, AICML

Faculty Team Members

Name	Role / Topic	Awards / Special Info
Michael Bowling	Principal investigator Robotics, games, reinforcement learning	NSERC, AIF New Faculty, AICML
Dale Schuurmans	Principal investigator Probabilistic methods in artificial intelligence, machine learning	CRC II Chair, NSERC, AICML, Director - AICML
Csaba Szepesvari	Principal investigator Nonparametric learning, statistical techniques	NSERC, AICML

Postdoctoral Fellows

Name	Topic	Awards / Special Info
Yaakov Engel	Gaussian process reinforcement learning	Alberta Ingenuity Fellowship, thru 9/07
Mohammad Ghavamzadeh	Bayesian reinforcement learning, actor-critic reinforcement learning	
Yuxi Li	Computational finance with machine learning	
Elliot Ludvig	Computational models of animal learning	
Barnabás Póczos	Machine learning, independent component analysis, unsupervised learning	From 9/07
Eric Wiewiora	Reinforcement learning theory	From 1/08
Linli Xu	Unsupervised learning, large margin learning	From 6/07
Martin Zinkevich	Minimal regret algorithms	Thru 9/07, now at Yahoo!

PhD Students

Name	Role / Topic	Scholarships / Awards / Special Info
Arash Afkanpour	Basis construction with high generalization in reinforcement learning	iCORE, AIF, Provost Doctoral Entrance Award
Nolan Bard	Using state estimation for dynamic agent modeling	NSERC, AIF
Gabor Bartok	Machine learning – representation learning	
Marc Gendron-Bellemare	Partially observable models	NSERC CGS, iCORE, Walter H. Johns
Amir massoud Farahmand	Regularities in reinforcement learning	
Alborz Geramifard	Least-squares reinforcement learning methods and planning	
Yuhong Guo	Learning bayesian networks	Graduated 9/07, awarded NSERC Postdoctoral Fellowship
Jiayuan Huang	Spectral clustering and semi-supervised learning	University of Waterloo, graduated 4/07
Feng Jiao	Probabilistic graphical models and algorithms for protein problems	University of Waterloo, graduated 3/08, now at Google

Anna Koop	Predictive representations in complex domains	AIF, iCORE
Daniel Lizotte	Global optimization using Gaussian processes	Killam, iCORE
Hamid Reza Maei	Reinforcement learning with function approximation from off-policy data	AIF
Adam Milstein	Improved particle filter based localization and mapping techniques	University of Waterloo, graduated 3/08
Yi Shi	Machine learning, bioinformatics	Provost Doctoral Entrance Award
David Silver	Reinforcement learning in computer GO	iCORE, AIF
Ivett Szabo	Bandit algorithms	Budapest University of Technology and Economics
Brian Tanner	Fast, incremental reinforcement learning	AIF, iCORE, Izaak Walton Killam Memorial, Dorothy J Killam Memorial
Qin Wang	Natural language processing and machine learning	Queen Elizabeth II graduate scholarship
Tao Wang	New representation and approximations for sequential decision making	SIGART/AAAI Doctoral Consortium Scholarship, graduated 7/07
Adam White	Using human guidance to speed up reinforcement learning methods	AIF, iCORE, NSERC
Dana Wilkinson	Subjective mapping	University of Waterloo, Ontario Graduate Scholarship, President's Scholarship, graduated
Linli Xu	Convex large margin training techniques	University of Waterloo, graduated 4/07, now a postdoc with the project

MSc Students

Name	Role / Topic	Scholarships / Awards / Special Info
Yasin Abbasi-Yadkori	Continuum-armed bandit problem	
Andrew Albert	Machine learning and games	NSERC, Walter H. Johns
Nolan Bard	Using state estimation for dynamic agent modeling	NSERC, AIF, graduated 2/08, now a PhD student with the project
Kate Davison	Bidding in bridge	
Varun Grover	Active learning	
Michael Johanson	Robust strategies and counter-strategies: building a champion level computer poker player	Graduated 2/08
Brad Joyce	Continuous action reinforcement learning	
Anna Koop	Predictive representations in complex domains	NSERC, AIF, iCORE, graduated 12/07, now a PhD student with the project

Volodymyr Mnih	Efficient stopping rules	NSERC, iCORE, Walter H. Johns, Jeffrey R. Sampson Graduate Memorial award, AIF
Yavar Naddaf	Game theory applied to computer games	
Gergely Neu	Inverse reinforcement learning	Budapest University of Technology and Economics, Scholarship of the Hungarian Republic
James Neufeld	Autonomous robot navigation	NSERC PGSD-3, iCORE, Walter H. Johns
Dave Schnizlein	No-limit poker	
Masoud Shahamiri	Reinforcement learning in environments with independent delayed-sense dynamics	
Kevin Waugh	Poker	

Other team members

Name	Role
Martha Lednicky	NSERC summer student, AIF scholarship holder
Akiko Green	Administrative assistant
Mark Lee	Programmer
Mathew Radtke	Programmer
Leah Hackman	Undergraduate research assistant
Eriv Verbeek	Undergraduate research assistant
Jason Roberts	Segway programming
Mike Sokolsky	Robot engineer
Lori Troop	Program administrator
Stephen Walsh	Segway programmer

Visitors

Name	Institution
Li Cheng	NICTA, Australian National University
Greg Grudic	University of Colorado at Boulder
Matt Taylor	University of Austin Texas
Kevin Murphy	University of British Colombia
Jefferson Provost	University of Texas
Balazs Csanad Csaji	MTA SZTAKI
Pedro Domingos	University of Washington

Carlos Guestrin	Carnegie Mellon
Marcus Hutter	ANU College of Engineering and Computer Science
Dmitry Kamanetsky	NICTA, Australian National University
Sam Roweis	University of Toronto
Antonio de Padua Braga	Federal University of Minas Gerais, Brazil
Alexandru Niculescu-Mizil	Cornell University
Jiming Peng	University of Illinois at Urbana-Champaigne
Joelle Pineau	McGill University

COLLABORATIONS

Provincial	
Participants	Nature of Collaboration
Alberta Ingenuity Center for Machine Learning	R. Sutton, D. Schuurmans, Cs. Szespesvari and M. Bowling are among the eight principle investigators for this center at the U. of Alberta. Total annual funding for AICML is $2M.

National	
Participants	Nature of Collaboration
Doina Precup, McGill University Prakash Panangaden, McGill University Yoshua Bengio, U. Montreal Shie Mannor, McGill University	NSERC Collaborative Research and Development Grants-Project, "Learning and prediction in high-dimensional stochastic domains," with Nortel Networks, Bell Canada, to R. Sutton and the listed collaborators $186,523 total (Sept 1/06 – Aug 31/09).
Jonathan Schaeffer, Robert Holte, Duane Szafron, and Michael Buro University of Alberta	NSERC Strategic Grant, "Intelligent Agents for Interactive Entertainment" to M. Bowling, and D. Schuurmans, and the listed collaborators $480K total (Oct 1/04-Sept 30/07).
Yoshua Bengio, McGill Hugh Chipman, U. Waterloo Christian Leger, U. Montréal William Welch, U. British Columbia Jim Ramsay, McGill Mu Zhu, U. Waterloo	MITACS grant "Statistical Learning of Complex Data with Complex Distributions" to D. Schuurmans, and the listed collaborators $59K to D. Schuurmans (Apr 1/05-Mar 31/09).

International	
Participants	Nature of Collaboration
Shalabh Bhatnagar, Indian Institute of Science, Bangalore	Joint research with R. Sutton, M. Ghavamzadeh, and M. Lee on actor-critic RL algorithms.
Sylvain Gelly, University of Paris South	Joint research with D. Silver on the use of RL in Computer Go.
Jean-Yves Audibert, CERTIS, France	Joint research with Cs. Szepesvari on variance estimation in bandit problems.
Remi Munos, INRIA, Lille, France	Joint research with Cs. Szepesvari on batch reinforcement learning
Peter Auer, Ronald Ortner, University of Leoben, Austria,	Joint research with Cs. Szepesvari on efficient exploration
Andras Antos, MTA SZTAKI, Hungary	Joint research with Cs. Szepesvari on statistical machine learning
E. James Kehoe, University of New South Wales, Australia	Joint research with R. Sutton, E. Ludvig, E. Verbeek and J. Neufeld on the relationship between reinforcement learning and learning in animals.

Industrial	
Participants	Nature of Collaboration
Michael James, Toyota Motor Corporation	With R. Sutton and M. Shahamiri exploring the use of reinforcement learning technology for improving the fuel efficiency of gas-electric hybrid cars.

GRADUATES

Name	Degree	Research topic	Current Position
Martin Zinkevich	Post Doctoral Fellow	Minimal regret algorithms	Yahoo! Research
Yaakov Engel	Post Doctoral Fellow	Gaussian process reinforcement learning	
Yuhong Guo	PhD	Learning Bayesian networks from data: structure optimization and parameter estimation	Research fellow at Research School of Sciences and Engineering, Australian National University. Awarded an NSERC Postdoctoral Fellowship

Jiayuan Huang	PhD	Learning from partially labeled data: unsupervised and semi-supervised learning on graphs and learning with distribution shifting	Technical lead at Telus, Business Intelligence, Toronto
Feng Jiao	PhD	Probabilistic graphical models and algorithms for protein problems	Research scientist at Yahoo! Inc., Sunnyvale, CA
Adam Milstein	PhD	Improved particle filter based localization and mapping techniques	Department of Computing Science, U of A
Tao Wang	PhD	New representations and approximations for sequential decision making under uncertainty	Research Fellow in the Research School of Information Science and Engineering at the Australian National University
Dana Wilkinson	PhD	Subjective mapping	Research Scientist at AiLive, Mountain View, CA
Linli Xu	PhD	Convex large margin training techniques: unsupervised, semi-supervised, and robust support vector machines	Postdoctoral fellow at University of Alberta, Departments of Physics and Computing Science
Nolan Bard	MSc	Using state estimation for dynamic agent modeling	PhD student – Department of Computing Science, U of A
Michael Johanson	MSc	Robust strategies and counter-strategies: building a champion level computer poker player	Research Assistant to Dr. Michael Bowling, U of A
Armita Kaboli	MSc	Bayesian calibration for Monte Carlo localization	PhD student, University of Massachusetts
Anna Koop	MSc	Investigating experience: temporal coherence and empirical knowledge representation	PhD student – Department of Computing Science, U of A

INTELLECTUAL PROPERTY

None.

PUBLICATIONS

REFEREED JOURNAL PUBLICATIONS

A. Antos, Cs. Szepesvari, R. Munos, "Learning Near-optimal Policies with Bellman-Residual Minimization Based Fitted Policy Iteration and a Single Sample Path," *Machine Learning Journal,* 71:89-129, 2008.

B. Bai, J. Harms, Y. Li, "Configurable Active Multicast Congestion Control," *Computer Networks*, Vol 52/7, pp. 1410-1432.

M. Ghavamzadeh, and S. Mahadevan, "Hierarchical Average Reward Reinforcement Learning," *Journal of Machine Learning Research,* 8:2629-2669, 2007.

E.J. Kehoe, E.A. Ludvig, J.E. Dudeney, J. Neufeld, and R.S. Sutton, "Magnitude and Timing of Nictitating Membrane Movements During Classical Conditioning of the Rabbit *(Oryctolagus cuniculus)," Behavioral Neuroscience*. (to appear)

E. A. Ludvig, R. S. Sutton, and E. J. Kehoe, "Stimulus Representation and the Timing of Reward-Prediction Errors in Models of the Dopamine System," *Neural Computation*. (to appear)

E. A. Ludvig, F. Balci, and K. M. Longpre, "Timescale Dependence in a Conditional Temporal Discrimination Procedure," *Behavioural Processes,* Mar. 2008, 77, pp. 357-363.

A. Milstein and T. Wang, "Dynamic Motion Models in Monte Carlo Localization," *Integrated Computer-Aided Engineering*, Volume 14, Number 3, 3007, pp. 243-262.

R. Munos and Cs. Szepesvári, "Finite Time Bounds for Sampling Based Fitted Value Iteration," *Journal of Machine Learning Research*, 2008.

Z. Szabó, B. Póczos, and A. Lorincz, "Independent Process Analysis without Combinatorial Efforts," *Pattern Analysis & Applications Journal*. (to appear)

HIGHLY REFEREED ARCHIVAL CONFERENCE PROCEEDINGS

A. Antos, R. Munos, and Cs. Szepesvári, "Fitted Q-Iteration in Continuous Action-Space MDPs," *Proc. Neural Information Processing Systems 21* (NIPS-07). (22% acceptance)

J. Y. Audibert, R. Munos, and C. Szepesvári, "Tuning Bandit Algorithms in Stochastic Environments," *Proc. of Algorithmic Learning Theory* (ALT-07). (20% acceptance)

S. Bahatnagar, R. Sutton, M. Ghavamzadeh, and M. Lee, "Incremental Natural Actor-Critic Algorithms," *Proc. Neural Information Processing Systems 21* (NIPS-07). (22% acceptance)

M. Bowling, A. Geramifard, and D. Wingate, "Sigma Point Policy Iteration," *Proc. of the 7th Int'l Joint Conf. on Autonomous Agents and Multi-Agent Systems* (AAMAS-08). (22% acceptance).

M. Bowling, M. Johanson, N. Burch, and D. Szafron, "Strategy Evaluation in Extensive Games with Importance Sampling," *Proceedings of the 25th International Conference on Machine Learning* (ICML-2008). (27% acceptance)

A. M. Farahmand, Cs. Szepesvári, and J. Y. Audibert, "Manifold-Adaptive Dimension Estimation," *Proc. Int'l. Conf. on Machine Learning* (ICML-2007). (29% acceptance)

S. Gelly and D. Silver, "Achieving Master Level Play in 9x9 Computer Go", *Proc. of the 23rd Conf. on Artificial Intelligence* (AAAI-08 Nectar). (21% acceptance)

Y. Guo, and D. Schuurmans, "Convex Relaxations of Latent Variable Training," *Proc. Neural Information Processing Systems 21* (NIPS-07). (22% acceptance)

Y. Guo, and D. Schuurmans, "Discriminative Batch Mode Active Learning," *Proc. Neural Information Processing Systems 21* (NIPS-07). (22% acceptance)

A. Isaza, Cs. Szepesvari, V. Bulitko, and R. Greiner, "Speeding Up Planning in Markov Decision Processes via Automatically Constructed Abstractions," *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence* (UAI-08). (28% acceptance)

M. Johanson, M. Zinkevich, and M. Bowling, "Computing Robust Counter-Strategies," *Proc. Neural Information Processing Systems 21* (NIPS-07). (22% acceptance)

G. Neu and C. Szepesvári, "Apprenticeship Learning using Inverse Reinforcement Learning and Gradient Methods," *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence* (UAI 2007). (32% acceptance)

J. Neufeld and M. Bowling, "Autonomous Geocaching: Navigation and Goal Finding in Outdoor Domains," *Proc. of the 7th Int'l. Joint Conf. on Autonomous Agents and Multi-Agent Systems* (AAMAS-08). (22% acceptance)

D. Silver, R. S. Sutton and M. Müller, "Sample-Based Learning and Search with Permanent and Transient Memories," *Proceedings of the 25th International Conference on Machine Learning* (ICML-2008). (27% acceptance)

V. Mnih, Cs. Szepesvari, and J.-Y. Audibert, "Empirical Bernstein Stopping," *Proceedings of the 25th International Conference on Machine Learning* (ICML-2008). (27% acceptance)

N. Sturtevant, H. J. Hoover, J. Schaeffer, S. Gouglas, M. Bowling, F. Southey, M. Bouchard, and G. Zabaneh, "Multidisciplinary Students and Instructors: A Second-year Games Course," *Proc. of the 39th ACM Technical Symposium on Computer Science Education*, 2008. (31% acceptance)

R. S. Sutton, Cs. Szepesvari, A. Geramifard, M. Bowling, "Dyna-Style Planning with Linear Function Approximation and Prioritized Sweeping," *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence* (UAI-08). (28% acceptance)

U. Syed, R. Schapire, and M. Bowling, "Apprenticeship Learning Using Linear Programming," *Proceedings of the 25th International Conference on Machine Learning* (ICML-2008). (27% acceptance)

Z. Szabó, B. Póczos, and A. Lörincz, "Undercomplete Blind Subspace Deconvolution via Linear Prediction," *Proc. European Conf. on Machine Learning 18* (ECML-07). (24% acceptance)

T. Wang, D. Lizotte, M. Bowling, and D. Schuurmans, "Stable Dynamic Programming," *Proc. Neural Information Processing Systems 21* (NIPS-07). (22% acceptance)

Q. Wang, D. Schuurmans, D. Lin, "Semi-supervised Convex Training for Dependency Parsing," *Proc. Of Association of Computational Linguistic Human Language Technologies Conference* (ACL-08). (to appear) (25% acceptance)

M. Zinkevich, M. Johanson, M. Bowling, and C. Piccione, "Regret Minimization in Games with Incomplete Information," *Proc. Neural Information Processing Systems 21 (NIPS-07)*. (22% acceptance)

BOOKS AND CHAPTERS

Y. Shi, Z. Cai and Guohui Lin. Classification Accuracy Based Microarray Missing Value Imputation. Chapter 14 in *Bioinformatics Algorithms: Techniques and Applications*. pp. 303-328, 2008.

SPECIAL INVITED PRESENTATIONS

R. Sutton, "Stimulus Representation in Temporal-difference Models of the Dopamine System." Seminar Series of the Interdisciplinary Program in Computation and Neural Systems, California Institute of Technology, Pasadena, California. June 4, 2007.

R. Sutton, "How Simple Can Mind Be?" International Workshop on Natural and Artificial Cognition, Pembroke College, Oxford University, June 26, 2007.

Cs. Szepesvari, "Reinforcement Learning," Machine Learning Summer School, 2008, Kioloa, Australia, March 6-7, 2008.

M. Bowling, "State of the Art in Computer Poker," NIPS Workshop on Machine Learning and Games, December 8, 2007.

M. Bowling, "AI After Dark: Computers Playing Poker," Yahoo! Research, February 25, 2008.

M. Ghavamzadeh, “Bayesian Reinforcement Learning,” Microsoft Research, Cambridge, UK, May 9, 2007.

E. Ludvig, “Magnitude and Timing of Nictitating Membrane Movements During Classical Conditioning of the Rabbit: A Fine-grain Analysis,” Australian Learning Group Christmas Workshop, Sydney, Australia, November 30, 2007.

E. Ludvig, “Response Timing in Temporal-difference (TD) Models of Conditioning,” Associative Learning Symposium, Gregynog, Wales, UK, March 18, 2008.

AWARDS

Michael Bowling was awarded the 2005 PE Publishing Award for the *Journal of Systems and Control Engineering* for the article entitled “STP: Skills, tactics and plays for multi-robot control in adversarial environments,” co-authored with Brett Browning, James Bruce, and Manuela Veloso.

Csaba Szepesvari was awarded an Alberta Ingenuity New Faculty grant for $300K.

OUTREACH

Anna Koop coordinated and participated in demonstrations for grade seven students at the Celebrating Ingenuity Event and grade nine students at the 2008 CIPS Women in Technology Program.