AITF ANNUAL REPORT 2013

DR. RICHARD SUTTON

REINFORCEMENT LEARNING AND ARTIFICIAL INTELLIGENCE

AITF ANNUAL REPORT MARCH 31, 2013

1. EXECUTIVE SUMMARY

The RLAI research program pursues an approach to artificial intelligence and engineering problems in which they are formulated as large optimal-control problems and approximately solved using reinforcement-learning methods. Reinforcement learning is a new body of theory and techniques for optimal control that has been developed in the last twenty-five years primarily within the machine learning and operations research communities, and which have separately become important in psychology and neuroscience. Reinforcement learning researchers have developed novel methods to approximate solutions to optimal-control problems 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 deployment, 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. 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 60 members, 39 of whom are graduate students and, of those, 11 of which are recipients of major scholarships. The output of the research program remained strong this year, with 38 papers published or accepted for publication in refereed archival venues during the reporting period. Four PhD and five 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 here 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 machine's 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 for doing this.

Highlights of the research program this year include: 1) learning to anticipate intentions of an amputee connected to a prosthetic robot arm, saving 14% of the time lost to switching between joints, 2) the first exact gradient descent algorithm for off-policy actor-critic reinforcement learning, and 3) proving that our latest super-human Poker program comes four times closer to perfect play than the previous best program.

2. RESEARCH PROGRAM OVERVIEW

In the project's proposal, research was divided into three main target areas. The first is extensions of existing reinforcement learning algorithms; there are many open problems in reinforcement learning, and we 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 (AI). There is a natural transition from the more advanced reinforcement learning methods to mechanisms for knowledge representation, search, and human-level reasoning. A major goal for the project is to explore, implement, and illustrate these relationships. The third main area of RLAI research is a focus on applications—on designing algorithms and software that are well suited for applied research, and on several specific applications. We discuss highlights of our research towards each of these target areas below.

In the area of extending core reinforcement-learning algorithms, our main focuses in the last two years have been on off-policy learning and actor-critic algorithms. Off-policy learning is key to our strategy for scaling reinforcement learning methods to address the more ambitious goals of AI. *Off-policy learning* means learning about a possible way of behaving (a *policy*) from following it for a period of time, perhaps only for a fraction of a second, without following it to completion. Moment to moment, the actions an AI agent selects can be seen as parts of many policies, but at most one will be followed to completion. If the agent can learn in parallel about all of the policies, then it can learn vastly more than if it is restricted to learning about only the one that is followed to completion, as it is with classical "on-policy" temporal-difference (TD) algorithms. In past years we have developed the world's first algorithms for off-policy learning, called gradient-TD methods, and we demonstrated that these algorithms can be used for real-time learning on robots with modest computational resources. In brief, in the last two years we have focused on extending this advance to actor-critic methods.

In actor-critic methods, a separate decision-making structure is learned in addition to the value functions ubiquitous in reinforcement learning. Our attention has been drawn to this older and, in recent decades, less popular category of algorithms because of our forays into robotics. In this and other real-time applications, the separate decision-making structure enables actions to be made with less computation and thus with a faster response time. In this year we have continued to work towards extending the theory of actor-critic algorithms to make them applicable to off-policy learning. Last year we reported significant progress toward this goal with the development of the Off-PAC algorithm, but further research has revealed that this method is only guaranteed convergent in the tabular case, without function approximation. Late in this year we have, in collaboration with international colleagues, developed a new strategy and a new class of algorithms, building on gradient-TD, which looks set to overcome these limitations.

In the area of applications, we have continued to make good progress in robotics, in adaptive prosthetics, and in computer Poker.

3. RESEARCH PROJECTS

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

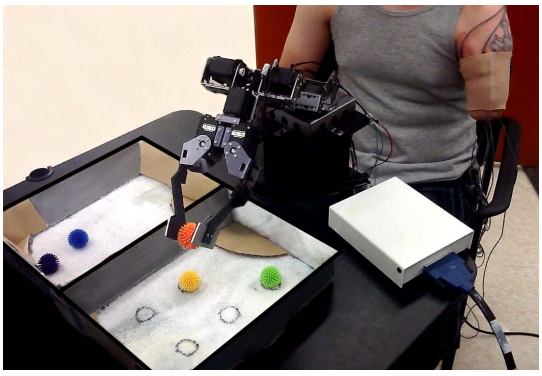
Off-policy Actor-Critic Reinforcement Learning

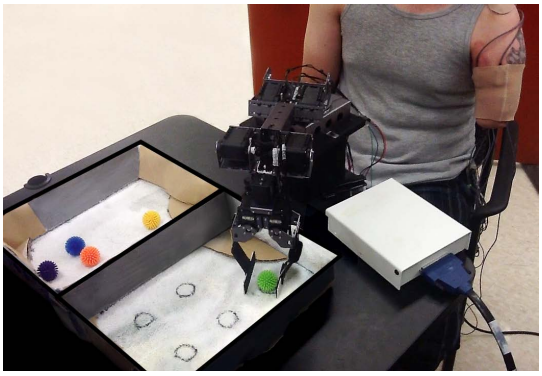
Core reinforcement-learning algorithms can be divided into two broad classes. The first class of methods, called *action-value methods*, including Q-learning, Sarsa, and our recently developed GQ algorithms, all work by estimating the value of state-action pairs, that is, by learning to predict the expected long-term reward obtainable starting from each action in each state. The second class of methods, called *actor-critic methods*, instead learn two objects called an *actor* and a *critic*. The critic's job is to estimate the expected long-term reward starting from each state, and the actor's job is to directly store a decision-making rule, a mapping from states to actions, also known as a *policy*. Because the policy is stored directly (rather than being computed from action values), actor-critic methods have the following advantages: 1) they can be easily initialized with human-designed policies, 2) they are more natural for large or continuous action spaces, 3) they can take actions stochastically with specific learned probabilities, and 4) they can act with minimal computation. The intuitive appeal and other advantages of actor-critic methods have made them a recent focus of research in reinforcement learning.

A second broad categorization of reinforcement learning methods has to do with what policies they can learn about. *On-policy* methods can learn only about the policy currently being used by the agent to select actions (e.g., about its value function), whereas the more general *off-policy* methods can learn about policies without following them exactly. Off-policy learning is key to our strategy for learning knowledge about the world from sensorimotor data, because it enables learning in parallel and thus on a large scale. In the last five years we have developed gradient-TD methods, the first scalable methods for off-policy learning with function approximation. But these are *action-value* methods; for the last two years we have attempted to extend our advance to *actor-critic* methods. Last year we obtained our first result in this area, a convergence result which applies to the tabular case and only approximately to function approximation, the real case of interest. At the end of this year, in collaboration with Dr. Hamid Maei (Postdoc at Stanford University, recent RLAI grad) and Prof. Susan Murphy (University of Michigan, visiting RLAI on her sabbatical) we made what appears to be a significant advance: the first off-policy actor-critic algorithm with function approximation that performs exact stochastic gradient descent. Moreover, the algorithm applies to both the average-reward and discounted-reward settings. This breakthrough should lead in short order to convergence proofs. We expect as a result of this breakthrough to fully solve the off-policy learning problem for actor-critic methods, providing one of the key technologies for achieving the long-term goals of the RLAI project.

Prediction in Adaptive Prosthetics

Assistive biomedical devices augment the abilities of amputees and other patients with lost physical or cognitive function. Specifically, these devices replace abilities lost due to traumatic injury, disease, aging, or congenital complications. This project focuses on one representative class of assistive robots: powered artificial limbs. Powered prostheses monitor electromyographic (EMG) signals produced by muscle tissue in a patient's body, and use these signals to control the movement of a robotic appendage with one or more controllable dimensions. Such devices are tightly coupled to a human user, with control processes that operate at high frequency and over extended periods of time. Despite the potential for improved function with myoelectric control, many patients reject the use of powered artificial limbs. Recent needs-assessment surveys of prosthesis users point out three key principal reasons for patient rejection: lack of intuitive control, insufficient functionality, and insufficient feedback from the myoelectric device. RLAI research, in collaboration with the Glenrose Rehabilitation Hospital (GRH), the UofA Faculty of Rehabilitation Medicine, and the UofA Dept. of Mechanical Engineering, aims to remove these barriers and increase the independence and ability of amputees through the use of adaptive machine learning techniques.





This year we have shown in pre-clinical trials that RLAI prediction-learning algorithms can significantly improve the amputee-prosthesis interface. In one common EMG setup, the amputee uses a myoelectric signal from a stump muscle to control joint at a time of the prosthetic arm, switching between joints using a separate toggle switch. The time spent switching through joints often consumes half the total time to perform a box-and-blocks task such as that shown in the figure. This wasted time could be significantly reduced if the joints were presented in an intelligent, user- and task-specific order. This year we have shown that RLAI prediction-learning algorithms can be applied to this problem to anticipate which joint the amputee will want next. In a controlled preliminary study with an able-bodied subject, we were able to increase by 23% the number of times the correct joint was presented to the subject as the first choice, Reducing the total time spent switching by 14.3%. We have also anticipated the intentions of amputees in a laboratory setting without yet changing the switching order.

Super-human 2-player Poker

The application of modern artificial intelligence methods to Poker has become a very active, competitive area of academic research in the last decade, and also one of great interest to the public as the computers have begun to challenge the best human players. RLAI team members, led by PI Michael Bowling, have played a key role in these developments. Bowling's group is one of the original founders in 2006 of the Annual Computer Poker Competition, a common testbed shared by researchers in artificial intelligence and computational game theory. The competition includes teams from universities from around the world, including Carnegie Mellon University (USA), Queens University (Canada), University of Auckland (New Zealand), and Technical University of Darmstadt (Germany). Over 12 events in the 2011 and 2012 competitions, the U of A team took first place six times, second place three times, and third place twice.

The single most important achievement in computer Poker was in 2008 when *Polaris*, a program developed by Bowling's group, defeated the world's best human players in the game of 2-player limit Texas hold'em Poker at the Man-Machine Poker Competition in Las Vegas. This success demonstrated super-human abilities, leaving just the question of how close computer programs could come to having a perfect, unbeatable Poker strategy, measured by the maximum amount the program could lose, on average, against a perfect adversary. If a program was developed whose maximum loss was at or very near zero, then the game could be considered solved. This would be a landmark result in the field of computational game theory, since to date, no human-scale game of imperfect information has been solved.

Last year we developed the first feasible algorithm, requiring just one day to run, for measuring how close a computer Poker player is to optimal play. Although Polaris was stronger than the world's best human players in 2008, it was still far from perfect: in fact, it was possible for a perfect opponent to beat Polaris for nearly five times the rate that a human professional aims to win in order to make a consistent profit. The best program in the Annual Computer Poker Competition in 2011 was developed by the University of Alberta and was stronger than Polaris, yet could still be defeated for twice the rate that human experts hope to achieve.

This year, we have developed a new Poker program, called the CFR-BR algorithm, which shattered all previous records for programs being close to perfect Poker play. The development of this algorithm reduced the worst-case loss of the world's strongest Poker program to just one quarter of what had previously been possible, meaning that it is no longer profitable (by human standards) for even a perfect adversary to play Poker against the program.

4. OBJECTIVES FOR NEXT YEAR

With regard to core reinforcement learning algorithms, we plan to formally establish the convergence properties of the new off-policy actor-critic algorithms and test them empirically. We also expect to further explore a new class of off-policy methods that combine the best features of gradient-TD and conventional-TD methods. The preliminary results suggest that these *hybrid-TD* methods may be as data efficient as conventional-TD methods without losing the stability of gradient-TD methods. We also still plan to combine actor-critic algorithms with *Autostep*, an automatic step-size adaptation algorithm, and make other refinements, to reduce or eliminate the number of parameters that must be tuned manually.

With regard to the larger ambitions of AI, work will proceed on several fronts. One of the advantages of off-policy learning is that it frees behavior from having to match the target policies, enabling it to be sculpted to other purposes. For example, the psychological notion of curiosity is that we sculpt our behavior so as to maximize our learning. Using off-policy learning it is possible to implement a computational version of curiosity, using an overall measure of learning progress in a parallel network of learning agents (as in our *Horde* architecture). A second front is the automated construction of appropriate representations from and for sensorimotor interaction. Progress in this area has been slow in AI, but we hope to be more successful by leveraging the rich stream of data provided by our robots' sensorimotor experience. This data provides many subtasks in the form of sensory signals to predict and control, all of which come from to the same physical system. Our robots, we believe, will make it easy to automatically create large numbers of naturally interrelated prediction and control tasks. It may still be too early to directly take on this goal, but we will be reaching for it as soon as possible.

With regard to the adaptive prosthetics project, we plan to test our anticipatory switching algorithm with amputees rather than able-bodied subjects. In addition, we will explore not just changing the order of the proffered joints, but actually, if the algorithm is sufficiently confident in its prediction of which joint the amputee will want, simply giving it to him without him having to explicitly ask for a new joint. We want to explore whether this will be disturbing and annoying, or if it can be done in a way that seems natural and saves additional time.

With regard to Poker, we expect to completely solve two-person Limit Poker by computing the optimal minimax Poker strategy by the end of next year. Once that is done, attention will switch to removing the restrictions on the game, considering first No-Limit Poker, and then Poker with three or more players. The extension to No Limit is expected to be fairly straightforward, but multi-player Poker involves a host of new issues and is expected to take longer.

With regard to our mobile robot fielded in the Courage Centre at the Glenrose Hospital (a

subproject not highlighted in this year's report), our plan for next year is to add a learning component to the robot. Currently, people respond to the robot, primarily by watching it, and the robot responds to the people's behavior (sensed visually through a Kinect sensor), but the robot's response is built in and not learned. Next year we plan to have the robot vary its behavior to discover what behavior maximizes specific reactions by people, perhaps simply spending more time near the robot.

5. RESEARCH TEAM MEMBERS AND CONTRIBUTIONS

a. Team Leader

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

b. Faculty Team Members

Michael Bowling	Faculty team member	AICML, NSERC
Dale Schuurmans	Faculty team member	AICML, CRC II Chair, NSERC, MITACS
Csaba Szepesvári	Faculty team member	AICML, NSERC
Martin Mueller	Associated faculty member	DARPA, NSERC
András György	Adjunct faculty member; Research Associate	Computer and Research Institute of the Hungarian Academy of Sciences

c. Post Doctoral Fellows & Research Associates

Name	Role	Awards / Special Info
Joseph Modayil	Research Associate	
Patrick Pilarski	Post doctoral fellow	
Harm van Seijen	Post doctoral fellow	
Joel Veness	Post doctoral fellow	
Xinhua Zhang	Post doctoral fellow	Researcher at the Machine Learning Group of National ICT Australia
Karim Abou-Moustafa	Post doctoral fellow	

Mohamed Elgendi	Post doctoral fellow	
Christopher Archibald	Post doctoral fellow	

d. PhD Candidates

Name	**Role**	**Scholarships / Awards / Special Info**
Yasin Abbasi-Yadkori	PhD candidate	
Arash Afkanpour	PhD candidate	
Gabor Balazs	PhD candidate	
Nolan Bard	PhD candidate	AITF ICT
Neil Birch	PhD candidate	AIF, NSERC PGS
Katherine Chen	PhD candidate	
Kenneth Dwyer	PhD candidate	
Alireza Farhangfar	PhD candidate	AICT ICT
Marc Gendron-Bellemare	PhD candidate	
Ruitong Huang	PhD candidate	
Michael Johanson	PhD candidate	
Anna Koop	PhD candidate	
Marc Lanctot	PhD candidate	
Ashique Mahmood	PhD candidate	Provost Doctoral Entrance Award
Aslan Ozlem	PhD candidate	Provost Doctoral Entrance Award
Bernardo Ávila Pires	PhD candidate	University of Alberta Doctoral Recruitment Scholarship
James Neufeld	PhD candidate	
Chris Rayner	PhD candidate	AITF ICT, NSERC PGS, Queen Elizabeth II Scholarship
Yi Shi	PhD candidate	
Adam White	PhD candidate	

Martha White	PhD candidate	NSERC CGS D, Honorary Izaak Walton Killam Memorial Scholarship
Min Yang	PhD candidate	
Hengshuai Yao	PhD candidate	
Yaoliang Yu	PhD candidate	
Mohammad Ajallooeian	PhD candidate	
Farzaneh Mirzazadeh	PhD candidate	
Pooria Joulani	PhD candidate	

e. MSc Candidates

Name	**Role**	**Scholarships / Awards / Special Info**
Mohammad Ajallooeian	MSc candidate	
Hao Cheng	MSc candidate	
Leah Hackman	MSc candidate	
Pooria Joulani	MSc candidate	University of Alberta Doctoral Recruitment Scholarship,
Michael Joya	MSc candidate	
Parisa Mazrooei	MSc candidate	
Roshan Shariff	MSc candidate	Alberta Innovates Graduate Student Scholarship
Mostafa Vafadost	MSc candidate	
Travis Dick	MSc candidate	
Joshua Davidson	MSc candidate	
Shun Jie Lau	MSc candidate	
Ryan Kiros	MSc candidate	Queen Elizabeth II Graduate Scholarship, Alberta Enterprise and Advanced Education Graduate Student Scholarship

f. Other Members

Name	Role
Beverly Balaski	Administrative assistant
Lori Troop	Program administrator
Alexandra Kearney	High school summer student, July-August 2012
Andrea Astle	High school summer student, July-August 2012

Geneva Giang	High school summer student, July-August 2012
Goutam Venkatramanan	High school summer student, July-August 2012
Rachel Han	High school summer student, July-August 2012

g. Visitors

Name	Institution
Mark Crowley	Oregon State University
András Antos	MTA SZTAKI
Shiva Kasiviswanathan	General Electric Research
Barnabas Poczos	Carnegie Mellon University
Mathukumalli Vidyasagar	The University of Texas at Dallas
Mohammad Ghavamzadeh	INRIA Lille
Lev Reyzin	University of Illinois at Chicago
Sandra Zilles	University of Regina
Andreas Artemiou	Michigan Technological University
Andrew McCallum	University of Massachusetts Amherst
Susan Murphy	University of Michigan
Hado van Hasselt	Centrum Wiskunde & Informatica
Joelle Pineau	McGill University
Elliot A. Ludvig	Princeton University
Michael Wellman	University of Michigan
Thomas Degris	INRIA Bordeaux Sud-Ouest
Gerald Tesauro	IBM's TJ Watson Research Center
Eric Laber	North Carolina State University
Simon Wulff	Technische Universität München
Soeren Laue	University of Jena, Germany
Dan Lizotte	University of Waterloo
Carolina Elias	Instituto Politechnico Nacional, Mexico

6. GRADUATES

Name	Degree	Research topic	Current Position
Leah Hackman	MSc	Faster Gradient-TD Algorithms	Junior Software Engineer at Granify in Edmonton
Pooria Joulani	MSc	Multi-Armed Bandit Problems under Delayed Feedback	PhD candidate in RLAI team
Mohammad M. Ajallooeian	MSc	Optimal Mechanisms for Machine Learning: A Game-Theoretic Approach to Designing Machine Learning	PhD candidate in RLAI team

Michael Joya	MSc	Methods for Automatic Heart Sound Identification	Unknown
Yasin Abbasi-Yadkori	PhD	Online Learning for Linearly Parametrized Control Problems	Postdoctoral researcher at Queensland University of Technology
Alireza Farhangfar	PhD	Novel Machine Learning Algorithms	Unknown
Arash Afkanpour	PhD	Multiple Kernel Learning with Many Kernels	Google Waterloo
Navid Zolghadr	MSc	Probe-Efficient Learning	BlackBerry
Yi Shi	PhD	Bio-relation Discovery and Sparse Learning	Postdoc at USC, Dept. Computer Science

7. COLLABORATIONS

Provincial	
Participants	Nature of Collaboration
Alberta Ingenuity Center for Machine Learning	R. Sutton, D. Schuurmans, C. Szepesvári, and M. Bowling are among the ten principle investigators for this center at the University of Alberta. Total annual funding for AICML is $2M.
Patrick M. Pilarski and Richard Sutton (RLAI), Linda Pilarski (U of A, Faculty of Medicine), and Carina Debes Marun (Cross Cancer Institute).	Collaboration with researchers from the AHFMR Team Microfluidics, Alberta Innovates Health Solutions, University of Alberta, and Cross Cancer Institute to explore intelligent biomedical image analysis methods to facilitate rapid lab-on-a-chip diagnostics. Part of the technology transfer focused "nanoBridge Research Grant RES-NAN-07-G10: Development of Fluorescence In Situ Hybridization (FISH) Platform, Chip, and Analysis Software" ($145K)
Patrick Pilarski and Richard Sutton (RLAI) and Dr. Kim Adams (University of Alberta, Division of Rehabilitative Medicine)	Collaborative work on the use of assistive robots to facilitate the development, skill acquisition, and cognitive assessment of young children and infants with severe motor impairments (e.g., children with cerebral palsy); work involves the use of machine intelligence to enable the automatic adaptation of a robot control system to match the needs and abilities of individual children. Pilarski and Adams co-supervise visiting graduate students from a partner institution in Mexico city.

Patrick Pilarski and Richard Sutton (RLAI), Martin Ferguson-Pell (Dean, Faculty of Rehabilitation Medicine, University of Alberta), Liping Qi (University of Alberta, Rehabilitation Medicine), Simon Grange (University of Calgary / University of Alberta / Alberta Health Services).	Collaboration between the RLAI and the Rehabilitation Robotics Sandbox Laboratory (Rehabilitation Medicine, U. Alberta); this collaboration focuses on using new machine-learning methods to predict fatigue in wheelchair users and enable novel muscle-stimulation paradigms.
Patrick Pilarski and Richard Sutton (RLAI), and AHFMR SMART Neuroprosthetics Team, University of Alberta	Preliminary work on the use of machine learning to adapt and optimize neural interfaces and assistive robotic devices for people with motor disabilities.
Joseph Modayil and Gary Faulkner, Jim Raso, Michael Cimolini from the Glenrose Rehabilitation Hospital	Collaboration with the "Robotics in the Glenrose Courage Centre" project is studying the interaction between people and autonomous robots in a rehabilitative medicine setting.
Dr. Ian Adatia, Director of the Pediatric Cardiac Critical and Intermediate Care Program, Stollery Children's Hospital and Roger Zemp, Department of Electrical and Computer Engineering University of Alberta	Joint research with Dale Schuurmans on adaptive signal processing techniques for heart disease diagnosis, focusing on predicting events with ECG and PCG signals.

National	
Participants	Nature of Collaboration
D. Precup, McGill University and Elliot Ludvig, Princeton University	Richard Sutton and Joseph Modayil co-organized the 8th Bellairs Workshop on Reinforcement Learning
Daniel Lizotte, University of Waterloo	Joint research with Michael Bowling on reinforcement learning for medical applications

International	
Participants	Nature of Collaboration
Zinhua Zhang, NICTA	Joint research with Dale Schuurmans on representation learning and optimization.
Yuhong Guo, Temple University	Joint research with Dale Schuurmans on machine learning in bioinformatics.
Andras Antos, MTA SZTAKI, Hungary	Joint research with Csaba Szepesvári on active learning.
Rong Zheng, University of Houston	Joint research with Csaba Szepesvári on application of online learning techniques in networking
Mohammad Ghavamzadeh, INRIA Lille	Joint research with Csaba Szepesvári on classification calibration.
David Pal, Google	Joint research with Csaba Szepesvári on online learning.
Yasin Abbasi-Yadkori, Queensland University of Technology, Brisbane, Australia	Joint research with Csaba Szepesvári on online learning in control.
Susan Murphy, University of Michigan	Joint research with Richard Sutton on off-policy actor-critic algorithms for clinical trials
Kevin Waugh, Carnegie Mellon University	Joint research with Michael Bowling on massive-scale computational game theory.
Nathan Sturtevant, University of Denver	Joint research with Michael Bowling on multiplayer game theory and automating heuristic construction for search.
Marc Lanctot, Maastricht University	Joint research with Michael Bowling on computational game theory.
Dr. Xuanto Su, Shandong University, China	Ongoing collaboration with Patrick Pilarski to explore the use of machine intelligence, pattern analysis, and optical lab-on-chip technology for the assessment of disease-related nanostructures in human cells.
Sophia Adamia, Dana Farber Cancer Institute, Harvard Medical School, USA	Investigation with Patrick Pilarski of cancer-related genetic markers in patients. Collaboration involving machine learning and biomedical data mining methods.
Jason P. Carey (Department of Mechanical Engineering, University of Alberta), Farbod Fahimi (University of Alabma at Huntsville, USA), Michael R. Dawson, Jacqueline S. Hebert, K. Ming Chan (Glenrose Rehabilitation Hospital)	Joint research with Patrick Pilarski and Richard Sutton. "Reinforcement Learning for Adaptive Prosthetics;" this collaboration investigates the use of reinforcement learning and real-time machine learning to enable adaptive, intuitive control of myoelectric prostheses and other assistive robotic devices.
Tamas Linder, Queen's University	Joint research with Andras Gyorgy on prediction in changing environments and

	structural problems of quantization.
Gabor Lugosi, Pompeu Fabra University	Joint research with Andras Gyorgy on prediction in changing environments and developing a family of efficient algorithms
Levente Kocsis, Computer and Automation Research Institute of the Hungarian Academy of Sciences	Joint research with Andras Gyorgy on efficient boosting algorithms for ranking
Gergely Neu and Andras Antos of Budapest University of Technology and Economics	Joint research with Andras Gyorgy on Markov Decision Processes with adversarial reward models
Hamid Reza Maei, Stanford University	Joint research with Richard Sutton on gradient-TD learning algorithms

8. INTELLECTUAL PROPERTY

Intellectual Property	Status	Short Description
PATENTS	none	
LICENSES	none	
Spinoff Companies	none	

9. PUBLICATIONS

REFEREED JOURNAL PUBLICATIONS

S. Adamia, P.M. Pilarski, A.R. Belch, and L.M. Pilarski, "Aberrant Splicing, Hyaluronan Synthases and Intracellular Hyaluronan as Drivers of Oncogenesis and Potential Drug Targets," *Current Cancer Drug Targets*, to appear.

S. Adamia, M. Bar-Natan, R.M. Stone, J.D. Griffin, and P.M. Pilarski, "Alternative splicing in Chronic Myeloid Leukemia (CML): a novel therapeutic target?" *Current Cancer Drug Targets*, to appear.

M. Arestu, P. Smitham, M. Loizidou, G. Jell, S. Grange, K. Chettiar, and P.M. Pilarski, "Nanotechnology & Medical Devices: Risk, Regulation and 'Meta' Registration," *World Journal of Engineering*, to appear.

A. Gyorgy, T. Linder, and G. Lugosi, "Efficient Tracking of Large Classes of Experts," *IEEE Transactions on Information Theory*, Nov. 2012, pp. 6709-6725.

L. Kocsis, A. Gyorgy, and A.N. Ban, "BoostingTree: Parallel Selection of Weak Learners in Boosting, with Application to Ranking," *Machine Learning*, to appear.

D. Lizotte, M. Bowling, and S. Murphy, "Linear Fitted-Q Iteration with Multiple Reward Functions," *Journal of Machine Learning Research 13*, Nov. 2012, pp. 3253–95.

Y. Naddaf, J. Veness, M. Bowling, and M. Gendron-Bellemare, "The Arcade Learning Environment: An Evaluation Platform for General Agents," *Journal of Artificial Intelligence Research*, Feb. 2013.

G. Neu, A. Gyorgy, Cs. Szepesvari, and A. Antos, "Online Markov Decision Processes under Bandit Feedback," *IEEE Transactions on Automatic Control*, to appear.

Y. Shi, B. Yuan, G. Lin, and D. Schuurmans, "Protein phosphorylation site prediction via feature discovery support vector machine," *Tsinghua Science and Technology*, Dec. 2012, pp. 638-644.

REFEREED ARCHIVAL CONFERENCE PROCEEDINGS

A. Afkanpour, Cs. Szepesvari, M. Bowling, and A. Gyorgy, "A Randomized Mirror Descent Algorithm for Large Scale Multiple Kernel Learning," *Proc. 30th Int'l Conf. on Machine Learning* (ICML 2013), to appear, 24% acceptance.

B. Ávila Pires, and Cs. Szepesvari, "Statistical linear estimation with penalized estimators: an application to reinforcement learning," *Proc. 29th Int'l Conf. on Machine Learning* (ICML 2012), Jun. 2012, pp. 1535-1542, 27% acceptance.

B. Ávila Pires, Cs. Szepesvari, and M. Ghavamzadeh, "Cost-sensitive Multiclass Classification Risk Bounds," *Proc. 30th Int'l Conf. on Machine Learning* (ICML 2013), to appear, 24% acceptance.

N. Bard, M. Johanson, N. Burch, and M. Bowling, "Online Implicit Agent Modelling," *Proc. 12th Int'l Conf.* on *Autonomous Agents and Multiagent Systems* (AAMAS 2013), to appear, 23% acceptance.

G. Bartok, N. Zolghadr, and Cs. Szepesvari, "An adaptive algorithm for finite stochastic partial monitoring," *Proc. 29th Int'l Conf. on Machine Learning* (ICML 2012), 27% acceptance.

M. Bowling, C. Archibald, and J. Davidson, "Baseline: Practical Control Variates for Agent Evaluation in Zero-Sum Domains," *Proc. 12th Int'l Conf. on Autonomous Agents and Multiagent Systems* (AAMAS 2013), to appear, 23% acceptance.

K. Chen, and M. Bowling, "Tractable Objectives for Robust Policy Optimization," *Proc. 26th Annual Conf. on Neural Information Processing Systems,* Dec. 2012, 25% acceptance.

R. Gibson, N. Burch, M. Lanctot, and D. Szafron, "Efficient Monte Carlo Counterfactual Regret Minimization in Games with Many Player Actions," *Proc. 26th Annual Conf. on Neural Information Processing Systems* (NIPS 2012), to appear, 25% acceptance.

Y. Guo, and D. Schuurmans, "Semi-supervised multi-label classification," *Proc. European Conf. on Machine Learning* (ECML 2012), Sep. 2012, pp. 355-370, 24% acceptance.

M. Johanson, N. Bard, N. Burch, and M. Bowling, "Finding Optimal Abstract Strategies in Extensive-Form Games," *Proc. 26th Conf. on Artificial Intelligence* (AAAI 2012), Jul. 2012, 26% acceptance.

M. Johanson, N. Burch, R. Valenzano, and M. Bowling, "Evaluating State-Space Abstractions in Extensive-Form Games," *Proc. 12th Int'l Conf. on Autonomous Agents and Multiagent Systems* (AAMAS 2013), to appear, 23% acceptance.

P. Joulani, A. Gyorgy, and Cs. Szepesvari, "Online Learning under Delayed Feedback," *Proc. 30th Int'l Conf. on Machine Learning* (ICML 2013), to appear, 24% acceptance.

R. Kiros, and Cs. Szepesvari, "Deep Representations and Codes for Image Auto-Annotation," *Proc. 26th Annual Conf. on Neural Information Processing Systems* (NIPS 2012), Dec. 2012, 25% acceptance.

M. Lanctot, R. Gibson, N. Burch, M. Zinkevich, and M. Bowling, "No-Regret Learning in Extensive-Form Games with Imperfect Recall," *Proc. 29th Int'l Conf. on Machine Learning* (ICML 2012), Jun. 2012, 27% acceptance.

M. Lanctot, A. Saffidine, J. Veness, C. Archibald, and M.H.M. Winands, "Monte Carlo *-Minimax Search," *Proc. Int'l Joint Conf. on Artificial Intelligence* (IJCAI 2013), to appear, 28% acceptance.

J. Modayil, A. White, P.M. Pilarski, and R.S. Sutton, "Acquiring a broad range of empirical knowledge in real time by temporal-difference learning," *Proc. IEEE Int'l Conf. on Systems, Man, and Cybernetics*, Oct. 2012.

G. Neu, A. Gyorgy, and Cs. Szepesvari, "The Adversarial Stochastic Shortest Path Problem with Unknown Transition Probabilities," *Proc. Int'l Conf. on Artificial Intelligence and Statistics* (AISTATS 2012), Apr. 2012, pp. 805-813, 33% acceptance.

P.M. Pilarski, T. Dick, and R.S. Sutton, "Real-time Prediction Learning for the Simultaneous Actuation of Multiple Prosthetic Joints," *Proc. IEEE Int'l Conf. on Rehabilitation Robotics* (ICORR), to appear.

C. Rayner, N. Sturtevant, and M. Bowling, "Subset Selection of Search Heuristics," *Proc. Int'l Joint Conf. on Artificial Intelligence* (IJCAI 2012), to appear, 28% acceptance.

H. van Seijen, and R.S. Sutton, "Efficient Planning in MDPs by Small Backups," *Proc. 30th Int'l Conf. in Machine Learning* (ICML 2013), to appear, 24% acceptance.

J. Veness, M. Bowling, and M. Gendron-Bellemare, "Sketch-Based Linear Value Function Approximation," *Proc. 26th Annual Conf. on Neural Information Processing Systems* (NIPS 2012), Dec. 2012. 25% acceptance.

J. Veness, K.S. Ng, M. Hutter, and M. Bowling, "Context Tree Switching," *Proc. IEEE Data Compression Conf.*, Apr. 2012, pp. 327-336, 30% acceptance.

J. Veness, M. White, M. Bowling, and A. Gyorgy, "Partition Tree Weighting," *Proc. IEEE Data Compression Conf.*, Mar. 2013.

M. White, Y. Yu, X. Zhang, and D. Schuurmans, "Convex Multi-view Subspace Learning," *Proc. 26th Annual Conf. on Neural Information Processing Systems* (NIPS 2012), Dec. 2012, 25% acceptance.

H. Yao, and Cs. Szepesvari, "Approximate Policy Iteration with Linear Action models," *Proc. 26th Conf. on Artificial Intelligence*, (AAAI 2012), Jul. 2012, 26% acceptance.

Y. Yu, O. Aslan, and D. Schuurmans, "A Polynomial-time Form of Robust Regression," *Proc. 26th Annual Conf. on Neural Information Processing Systems* (NIPS 2012), Dec. 2012, 25% acceptance.

Y. Yu, H. Cheng, D. Schuurmans, and Cs. Szepesvari, "Characterizing the representer theorem," *Proc. 30th Int'l Conf. on Machine Learning* (ICML 2013), Jun. 2013, to appear, 24% acceptance.

Y. Yu, J. Neufeld, R. Kiros, X. Zhang, and D. Schuurmans, "Regularizers versus Losses for Nonlinear Dimensionality Reduction," *Proc. 29th Int'l Conf. on Machine Learning* (ICML 2012), Jun. 2012, 27% acceptance.

X. Zhang, Y. Yu, and D. Schuurmans, "Accelerated Training for Matrix-norm Regularization: A Boosting Approach," *Proc. 26th Annual Conf. on Neural Information Processing Systems* (NIPS 2012), Dec. 2012, 25% acceptance.

OTHER CONFERENCE AND WORKSHOP PROCEEDINGS

K. Abou-Moustafa, and F. Ferrie, "A Note on Metric Properties for Some Divergence Measures. The Gaussian Case," *Proc. 4th Asian Conf. on Machine Learning* (ACML 2012), Nov. 2012, pp. 15.

K. Abou-Moustafa, and F. Ferrie, "Modified Divergence Measures for Gaussian Densities," *Proc. 9th Int. Workshop on Statistical Techniques in Pattern Recognition*, Nov. 2012, pp. 11.

C. Archibald, N. Burch, M. Rutherford, and M. Bowling, "Rating Players in Games with Real-Valued Outcomes," *Proc. 12th Int'l Conf. on Autonomous Agents and Multiagent Systems* (AAMAS 2013), 2 page extended abstract, to appear.

V. Bulitko, C. Rayner, and R. Lawrence, "On Case Base Formation in Real-Time Heuristic Search," *Proc. 8th Conf. Artificial Intelligence and Interactive Digital Entertainment* (AIIDE 2012), Oct. 2012, pp. 106-111, 54% acceptance.

N. Burch, and R.C. Holte, "Automatic Move Pruning Revisited," *Proc. 5th Annual Symposium on Combinatorial Search* (SOCS 2012), Jul. 2012.

T. Degris, P.M. Pilarski, and R.S. Sutton, "Apprentissage par Renforcement sans Modèle et avec Action Continue," *7èmes Journées Francophones Planification, Décision, et Apprentissage pour la conduite de systèmes* (JFPDA 2012), May 2012.

A. Gyorgy, T. Linder, and G. Lugosi, " Efficient Tracking of Large Classes of Experts," *Proc. IEEE Int'l Symposium on Information Theory*, Jul. 2012, pp. 885-889.

R. Kiros, "Training Neural Networks with Dropout Stochastic Hessian-Free Optimization," *Proc. Int'l Conf. on Learning Representations* (ICLR), to appear.

R. Kiros, and Cs. Szepesvari, "On Linear Embeddings and Unsupervised Feature Learning," *29th Int'l Conf. on Machine Learning* (ICML 2012) Workshop Paper, Jul. 2012.

M. Lanctot, A. Saffidine, J. Veness, and C. Archibald, "Sparse Sampling for Adversarial Games," *Proc. Computer Games Workshop at European Conf. on AI* (ECAI 2012), Workshop paper, Aug. 2012.

A.R. Mahmood and R.S. Sutton, "Online Representation Search and Its Interactions with Unsupervised Learning, " *Neural Information Processing Systems* (NIPS 2012) Workshop Paper, Dec. 2012

J. Modayil, A. White, P.M. Pilarski, and R.S. Sutton, "Acquiring Diverse Predictive Knowledge in Real Time by Temporal-difference Learning," *Proc. Int'l Workshop on Evolutionary and Reinforcement Learning for Autonomous Robot Systems* (ERLARS 2012), Aug. 2012

J. Modayil, A. White, and R.S. Sutton, "Multi-timescale Nexting in a Reinforcement Learning Robot," *Proc. 12th Int'l Conf. on Adaptive Behaviour*, (SAB 2012), Aug. 2012.

P.M. Pilarski, and R.S. Sutton, "Between Instruction and Reward: Human-Prompted Switching," *Proc. AAAI 2012 Fall Symposium on Robots Learning Interactively from Human Teachers* (RLIHT 2012), Nov. 2012.

P.M. Pilarski, T. Degris, M.R. Dawson, K.M. Chan, J.S. Hebert, J.P. Carey, and R.S. Sutton, "Towards Prediction-Based Prosthetic Control," *Proc. 17th Int'l Conf. Functional Electrical Stimulation Society, 2012*, Sep. 2012, pp. 26–29.

L. Qi, P.M. Pilarski, M. Ferguson-Pell, and S. Grange, "Pushrim kinetics and coordination patterns of shoulder muscles during surface and incline wheelchair propulsion," *Proc. 17th Int'l Conf. Functional Electrical Stimulation Society*, (IFESS 2012), Sep. 2012, pp. 151-153.

Y. Shi, X. Zhang, X. Liao, G. Lin, and D. Schuurmans, "Protein-chemical interaction prediction via a kernelized sparse learning SVM," *Proc. 18th Pacific Symposium on Biocomputing* (PSB 2013), Jan. 2013, pp. 41-52.

J. Veness, P. Sunehag, and M. Hutter, "On Ensemble Techniques for AIXI Approximation," *Proc. 5th Artificial General Intelligence* (AGI-2012), Dec. 2012.

A. White, J. Modayil, and R.S. Sutton, "Scaling Life-long Off-policy Learning," *Proc 2nd Joint IEEE Int'l Conf. on Development and Learning and on Epigenetic Robotics* (ICDL-Epirob), Nov. 2012.

MAGAZINE ARTICLES

P.M. Pilarski, T. Degris, M.R. Dawson, K.M. Chan, J.S. Hebert, J.P. Carey, and R.S. Sutton, "Adaptive Artificial Limbs: a Real-time Approach to Prediction and Anticipation," *IEEE Robotics and Automation Magazine, Special Issue on Assistive Robotics*, Mar. 2013, Vol. 20(1): pp. 53-64.

SPECIAL/INVITED PRESENTATIONS

Person	**Title**	**Venue**
R. Sutton	Scaling Life-long Learning	European Workshop on Reinforcement Learning (Keynote), Edinburgh, Scotland
R. Sutton	Learning About Sensorimotor Data	Thinking Conference, Brooklyn College, NY
R. Sutton	Scaling Life-long Learning	University of Delft, Netherlands

R. Sutton	Scaling Life-long Learning	UniversitÃ della Svizzera Italiana, Switzerland
D. Schuurmans	Convex methods for representation learning (Keynote Address, ACML 2012)	Singapore Management University
D. Schuurmans	Convex Methods for Latent Representation Learning	Ecole Normale Superieure, France
D. Schuurmans	Introduction to Machine Learning	Machine Learning Summer School, University of California Santa Cruz
D. Schuurmans	Generalized Domain Representations and Regularization	Machine Learning Summer School, University of California Santa Cruz
D. Schuurmans	Generalized Output Representations and Structure	Machine Learning Summer School, University of California Santa Cruz
D. Schuurmans	Latent Representations and Unsupervised Training	Machine Learning Summer School, University of California Santa Cruz
D. Schuurmans	Convex Methods for Representation Learning	Machine Learning Summer School, University of California Santa Cruz
Cs. Szepesvari	Sampling Approaches in Lookahead Policies	NSF Workshop: A Conversation between CS and OR on Stochastic Optimization, Rutgers University
Cs. Szepesvari	Partial Monitoring Games	Rutgers University
Cs. Szepesvari	Tradeoffs in online learning under partial information feedback	Multi-Trade-offs in Machine Learning, NIPS Workshop
M. Bowling	Abstraction with an Adversary	Invited Talk at IEEE CIG
M. Bowling	Abstraction with an Adversary	University of Waterloo
M. Bowling	Abstraction with an Adversary	University of Texas-Austin
M. Bowling	General Atari 2600 Game Playing	Invited Talk at IEEE CIG
J. Modayil	Multi-timescale Nexting in a Reinforcement Learning Robot	INRIA- Bordeaux, France
J. Veness	Context Tree Switching	Data Compression Conference – Snowbird
P. M. Pilarski	A Novel Targeted Sensory Reinnervation Method to Improve Function of Myoelectric Prostheses after Arm Amputation	The 2012 ACRM-ASNR Annual Conference: Progress in Rehabilitation Research
P. M. Pilarski	Towards Intelligent Artificial Limbs and other Miraculous Thinking Machines	TEDx Edmonton

P. M. Pilarski	Prediction and Control with Real-time Machine Learning	DBR Technical Series, Schlumberger Canada Ltd.
P. M. Pilarski	The Adaptive Prosthetics Project	DemoCamp 19, Startup Edmonton
C. Archibald	Skill and Billiards: Game Theory in Complex Domains	Brigham Young University
C. Archibald	Skill and Billiards: Game Theory in Complex Domains	Mississippi State University

AWARDS

A. White, R.S. Sutton, J. Modayil: "Paper of Excellence" award by the *Second Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics*, San Diego, USA.

J. Modayil, A. White, P.M. Pilarski, and R.S. Sutton: "Best Paper Award" at the *International Workshop on Evolutionary and Reinforcement Learning for Autonomous Robot Systems* (ERLARS), Montpellier, France.

THESES

Leah Hackman, MSc, "Faster Gradient-TD Algorithms," September 26, 2012.

Pooria Joulani, MSc, "Multi-Armed Bandit Problems under Delayed Feedback," September 26, 2012.

Mohammad M. Ajallooeian, MSc, "Optimal Mechanisms for Machine Learning: A Game-Theoretic Approach to Designing Machine Learning," January 31, 2013.

Michael Joya, MSc, "Methods for Automatic Heart Sound Identification," April 10, 2012

Yasin Abbasi-Yadkori, PhD, "Online Learning for Linearly Parametrized Control Problems," September 28, 2012.

Alireza Farhangfar, PhD, "Novel Machine Learning Methods," March 28, 2013.

Yi Shi, PhD, "Bio-relation Discovery and Sparse Learning," November 15, 2012.

Arash Afkanpour, PhD, "Multiple Kernel Learning with Many Kernels," March 28, 2013.

Navid Zolghadr, MSc, "Probe-Efficient Learning," January 31, 2013.

10. OUTREACH

Csaba Szepesvari supervised two high school students and Richard Sutton three high school students in July and August as part of the High School Internship Program at the Department of Computing Science.

Oslan volunteered as activity leader for Lego Robotics activity in Choices conference. The conference was organized for 600 grade 6 girls to introduce science related topics by enrolling them to different activities.

Chris Rayner volunteered as a high school students' guide for the Iverson Exam -- two hours of leading a small group of high school students around to different information sessions on campus.

Joseph Modayil demonstrated aspects of reinforcement learning and artificial intelligence on robots to students enrolled in a summer math camp (grades 8-12).

Patrick Pilarski presentation on pursuing post-doctoral studies in the sciences, delivered to Alberta graduate students; presentation invited by FGSr Outreach, U Alberta.

Patrick Pilarski presented robotics and machine learning demos to groups of high school students participating the Iverson Exam.

Patrick Pilarski presented a hands-on robotics and artificial intelligence demonstration and lecture for students in the Department of Computing Science Summer Camps (Riveting Robotics) for students, ages 14-16.

Patrick Pilarski gave a tour and demonstration of the AICML Adaptive Prosthetics Project to senior government ministers and deputy ministers, university administration, and members of the media. Part of the Edmonton Health Clinic Academy grand opening event, and Rehabilitation Robotics Laboratory opening tour.

Patrick Pilarski presented one of five demos at DemoCamp 19. Demonstration of AICML Technology to the Edmonton business, start-up, and entrepreneur community.

Patrick Pilarski provided a presentation and panel session for the UofA Undergraduate Research Initiative (URI) Discovery Panel on Games, Machines, and Artificial Intelligence; 50+ undergraduate students.

Katherine Chen hosted an object-oriented programming exercise using Alice for grade six and grade nine girls at the Women in Technology program.