AITF ANNUAL REPORT 2017

DR. RICHARD SUTTON
REINFORCEMENT LEARNING AND ARTIFICIAL INTELLIGENCE

AITF ANNUAL REPORT MARCH 31, 2017

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 by reinforcement-learning methods. Reinforcement learning is a new body of theory and techniques for optimal control that has been developed in the last thirty years primarily within the machine learning and operations research communities, and which has 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 mathematical analysis, through computational experiments, through the development of robotic systems, and through the development and testing of computational models of natural learning processes.

The overall research team consists of about 74 members, 46 of whom are graduate students and, of those, 13 of which are recipients of major scholarships. The output of the research program remained strong, with 9 journal articles and 15 highly-refereed archival conference papers published or accepted for publication during the reporting period. Three PhD students and three MSc students were graduated.

Highlights of the research program this year include: 1) the first computer program to outplay human professionals at Heads-up No Limit Poker, 2) hiring Dr. Martha White as an assistant professor who will become the sixth RLAI principal investigator, 3) contributing to the establishment of a $125 million national research and training program in artificial intelligence, and 4) new commercial relationships with Mitsubishi, Huawei, Kindred, and RBC research, and 5) a new algorithm, Q(σ), unifying and extending many existing reinforcement learning methods.

2. RESEARCH PROGRAM OVERVIEW

In the RLAI project's original 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.

The project has continued to make steady progress in the development of improved reinforcement learning algorithms. This year we introduced Q(σ), a new algorithm that unifies many classic reinforcement learning algorithms including Q-learning, Sarsa, Expected Sarsa, and Tree Backup. The unified algorithm simplifies software development and enables a spectrum of new intermediate algorithms that are empirically superior to all the classic existing algorithms. The project has also made progress toward the more ambitious goals of artificial intelligence this year by developing a new algorithm, called ABQ(zeta), that is more generally capable of efficient, reliable off-policy learning, which is viewed as key to learning powerful models of the world.

In the realm of applications, the project has reached another major milestone in computer Poker. In January of 2015, a team led by RLAI principal investigator Michael Bowling published a paper in *Science* announcing that two-person Limit Poker was essentially solved, meaning that the play of their program, *Cepheus*, could not be distinguished from perfect play in a lifetime of playing poker 24 hours a day. Since then, the project has turned to No Limit Poker, in which any numerical bid was possible, and for which entirely new methods are unfortunately required. In February of this year Bowling and another team announced, *Deepstack*, the first program capable of outplaying human professionals in two-person No Limit Poker. This paper also appeared in *Science* and was featured on the cover.

Artificial intelligence (AI) has risen to a new level of public prominence in Canada this year. Recent spectacular success of AI, such as Deepstack in Poker, have been widely discussed in the mainstream press. RLAI personnel have participated publically by giving many interviews and behind-the-scenes by being part of a new national program to build and retain Canada's strength in AI. In March the program was announced by the Government of Canada, allocating $125,000,000 over five years for research and training in AI, focused on deep learning and reinforcement learning. The funding will be used to develop AI research centers in Montreal, Toronto-Waterloo, and Edmonton. The Edmonton research center will be an expansion of the Alberta Machine Intelligence Institute (Amii), five of whose eleven principal investigators are the principal investigators of the RLAI project. This federal funding will significantly increase the funding available to the project's research and is a significant affirmation of our mission.

This year a new faculty member has been hired who will become the sixth principal investigator of the RLAI project. Martha White, previously a professor at Indiana University, will become an assistant professor of Computing Science at the University of Alberta, starting in July 2017. Professor White is a graduate of the RLAI research program in 2014 (supervised by Bowling and Schuurmans) and is uniquely well prepared to extend the research program.

3. RESEARCH PROJECTS

The overall RLAI project comprises many component research projects in three main areas: 1) designing improved reinforcement learning algorithms, 2) extending reinforcement learning to address the more ambitious goals of artificial intelligence, and 3) exploring applications of reinforcement learning algorithms. This section describes in more detail a few of these component research projects.

Unifying Classic Reinforcement-Learning Methods with the Q(σ) Algorithm

Seemingly disparate reinforcement learning algorithms can sometimes be unified by introducing a new parameter or other generalizations to produce a larger space of algorithms with the previously existent algorithms at the extreme corners. Surprisingly, when this is done, intermediate algorithms are sometimes found that perform better that any of the extremes and thereby improve the state of the art. As a primary example, the classical TD(λ) algorithm (from 1988) elegantly unifies one-step temporal-difference (TD) methods with Monte Carlo methods through the use of eligibility traces and the trace-decay parameter λ.

There are a multitude of one-step TD methods that can be used to learn action-value functions, including Sarsa, Q-learning, Tree Backup, and Expected Sarsa. These methods can also be extended across multiple time steps to achieve better performance. Each of these algorithms is seemingly distinct, and no one dominates the others for all problems. This year we have developed a new multi-step action-value algorithm, called *Q(σ)*, which unifies and generalizes these existing algorithms, subsuming them as special cases. A new parameter, σ, is introduced to allow the degree of sampling performed by the algorithm at each step to be continuously varied, with Sarsa existing at one extreme (full sampling), and Expected Sarsa existing at the other (pure expectation). Q(σ) is also general in that it includes arbitrary target and behavior policies. If the target policy is specialized to the greedy policy, then Q(σ) becomes classic Q-learning, whereas if it is left arbitrary, then multi-step Q(σ) becomes Tree Backup.

In computational experiments, we have shown that an intermediate value of σ, which results in a mixture of the existing algorithms, performs better than either extreme. The mixture can also be varied dynamically to achieve even greater performance.

Efficient Off-policy Learning without Importance Sampling (the ABQ Algorithm)

Off-policy learning constitutes an important class of reinforcement learning problems, where the goal is to learn about a designated target policy while behaving according to a different policy. In recent years, off-policy learning has garnered a substantial amount of attention in policy evaluation (prediction) tasks. When safety and data frugality is paramount, the ability to evaluate a policy without actually following it can be invaluable. Learning a large number of off-policy predictions is also important for model learning, option learning, scalable life-long learning, and knowledge representation, as has been discussed and highlighted in past years' reports.

A number of computationally scalable algorithms have been proposed that can learn off-policy without requiring a model of the environment, i.e., that are *model-free*. A core component of these algorithms is a classical Monte Carlo technique known as *importance sampling*, in which samples are scaled by the ratio of the two policies so that they appear to be drawn from the target policy. Although importance sampling plays a key role in correcting the discrepancy between the policies, the high variance of the importance sampling ratios often results in large estimation uncertainty. It is thus desirable to learn off-policy without importance sampling.

All previous methods that successfully avoided importance sampling were one-step temporal-difference (TD) methods. This is unfortunate because multi-step TD methods can be much superior, particularly when function approximation is used. This is doubly unfortunate because the problem of large variance with importance sampling is also the most severe in multi-step learning and, consequently, multi-step off-policy learning remains problematic and its promise largely unfulfilled.

Our key contribution this year was to develop an algorithmic technique based on modulating the degree of TD learning in an action-dependent manner. We introduced an action-dependent bootstrapping parameter, *zeta*. For action-value estimation, importance sampling ratios can be eliminated by varying zeta for different state-action pairs in a particular way. Using this technique, we introduced a new algorithm, called ABQ(zeta), with greatly reduced estimation variance compared to state-of-the-art off-policy algorithms. ABQ(zeta) is the first off-policy method to achieve multi-step function approximation solutions without explicitly using importance sampling ratios. The prior Tree-Backup algorithm can be shown to be a special case of ABQ(zeta). In addition, a recently proposed off-policy algorithm in the literature, Retrace, can be derived and extended to the case of function approximation with stability using the action-dependent bootstrapping technique.

No Limit Poker

Games have long served as benchmarks and milestones of progress in AI. In recent decades, computer programs have exceeded expert human players in many games, including backgammon, checkers, chess, Jeopardy!, Atari video games, and go. However, all these successes involve games with information symmetry, where all players have identical information about the current state of the game. This property of *perfect information* is also at the heart of the algorithms that enabled these successes, e.g., local search during play. Poker, on the other hand, is a game of *imperfect information*, where players' private cards give them asymmetric information about the state of the game, and therefore each player has to reason about the knowledge of the other players.

Heads-up no-limit Texas hold'em (HUNL) is a two-player version of poker in which two cards are initially dealt facedown to each player, and additional cards are dealt face-up in three subsequent rounds of betting. *No limit* is placed on the size of the bets. RLAI project research in 2015 essentially solved the simpler game of heads-up *limit* Texas hold'em, where all bets are of a fixed size, and which has just under 10^{14} decision points. The imperfect information game HUNL is vastly larger, with more than 10^{160} decision points.

Imperfect information games require more complex reasoning than similarly sized perfect information games. The correct decision at a particular moment depends upon the probability distribution over private information that the opponent holds, which is revealed through their past actions. However, how our opponent's actions reveal that information depends upon their knowledge of our private information and how our actions reveal it. Competitive AI approaches in imperfect information games (like our 2015 program) typically reason about the entire game and produce a complete strategy prior to play. HUNL is too large for that.

Our new program, *DeepStack*, takes a fundamentally different approach. It does not compute and store a complete strategy prior to play. Instead it considers each particular situation as it arises during play. It avoids reasoning about the entire remainder of the game by substituting the computation beyond a certain depth with a fast, approximate estimate. This estimate can be thought of as DeepStack's intuition: a gut feeling of the value of holding any possible private cards in any possible poker situation. DeepStack's intuition, much like human intuition, needs to be trained. We trained it with deep learning using examples generated from random poker situations. DeepStack is theoretically sound, produces strategies substantially more difficult to exploit than abstraction-based techniques, and defeats professional poker players at HUNL with statistical significance.

We evaluated DeepStack by playing it against a pool of professional poker players recruited by the International Federation of Poker. 44,852 games were played by 33 players from 17 countries. Eleven players completed the requested 3,000 games with DeepStack beating all but one by a statistically-significant margin. Over all games played, DeepStack outperformed players by over four standard deviations from zero. With this study and a publication in *Science* in March 2017, DeepStack became the first AI capable of beating professional poker players at heads-up no-limit Texas hold'em poker.

Pan-Canadian Artificial Intelligence Strategy

Canada has been the origin of some of the best research in AI. Notably, *deep learning* originated early in this decade in the work of Geoff Hinton at the University of Toronto, and then continued with that of Yoshua Bengio, now at the Université de Montréal. Canada remains home to some of the most dynamic deep learning research in the world. Canada is also known for its reinforcement learning research, with a strong program at McGill University, and the world's strongest program arguably here at the University of Alberta in the RLAI laboratory. More generally, the University of Alberta is ranked second in the world in AI and machine learning by csrankings.org.

Although Canada has produced some of the strongest AI research and researchers, many of them have had to leave Canada for other parts of the world to find jobs or to otherwise grow their AI research. The vast majority of our graduates go on to work outside of Canada at large companies such as Google, Microsoft, Twitter, or Facebook. Many Canadian faculty are also drawn away from Canada for lucrative opportunities in industry.

In the attempt to stem the exodus and develop the Canadian AI industry, the Government of Canada has allotted $125,000,000 over five years for the training and retention of Canadian AI researchers, particularly those working in deep learning and reinforcement learning. This funding will be concentrated in three AI research institutes located in Montreal, Toronto-Waterloo, and Edmonton. The Montreal institute will be an outgrowth of Bengio's MILA. The Toronto-Waterloo institute will be the newly announced Vector Institute directed by Geoff Hinton. The Edmonton institute will be a greatly enlarged Alberta Machine Intelligence Institute (Amii).

The RLAI project helped create and will be substantially strengthened by this new national research effort. The RLAI principal investigators comprise five of the eleven principal investigators of Amii. Some of the funding will be used to retain RLAI principal investigators and recruit new ones. The national effort will also support collaboration and exchanges across Canada. For example, RLAI students have already attended deep learning summer schools held in Montreal.

4. OBJECTIVES FOR NEXT YEAR

With regard to core reinforcement learning algorithms, we envision taking several further steps over the next year. Much of this work is organized around producing a second edition of the textbook on reinforcement learning that Sutton authored together with Prof. Andrew Barto of the University of Massachusetts. The first edition has been extremely successful; at one time last year it was listed as the most influential publication in computer science by SemanticScholar.org. The second edition will be completed next year and will be more than 50% larger, with new chapters on psychology, neuroscience, function approximation, and policy-gradient methods. The new edition will be published by MIT Press and will be available for free on the Internet.

With regard to the larger ambitions of artificial intelligence, we plan another summer of empirical system building with robot microworlds. In a simplified robot environment, we intend to further explore how RLAI innovations like general value functions, nexting, and the Horde architecture can be used to learn substantial knowledge of the world that is both grounded in sensorimotor data and significantly abstract. Such knowledge can then form a model of the world suitable for reinforcement-learning based planning methods. This coming year we will graduate to a new generation of robots based on the Kobuki mobile-robot base, replacing the iRobot Create base.

5. RESEARCH TEAM MEMBERS AND CONTRIBUTIONS

a. Team Leader

Name	Role	% of salary support paid by Alberta Innovates Chair Grant	Awards / Special Info
Richard Sutton	Principal Investigator	5 %	AICML, NSERC, Google DeepMind

b. Faculty Team Members

Michael Bowling	Faculty team member	0	
Dale Schuurmans	Faculty team member	0	AICML, NSERC, MITACS
Csaba Szepesvári	Faculty team member	0	AICML, NSERC
Patrick Pilarski	Faculty team member	0	AICML, NSERC
Martin Müller	Associated faculty member	0	NSERC

c. Post Doctoral Fellows & Research Associates

Name	Role	% of salary support paid by Alberta Innovates Chair Grant	Awards / Special Info
Huizhen Yu	Research Associate	100%	
Harm van Seijen	Post doctoral fellow	100%	Research Manager at Microsoft
Tor Lattimore	Post doctoral fellow	0	Now faculty at Indiana University, moving to DeepMind, London, UK
Karim Abou-Moustafa	Post doctoral fellow	0	
Jae Young Lee	Post doctoral fellow	50%	
Villiam Lisy	Post doctoral fellow	100%	Assistant professor, Department of Computer Science, FEE, Czech Technical University in Prague
Nolan Bard	Post doctoral fellow	100%	
Omid Namaki Shoushtari	Post doctoral fellow	0	NSERC Engage
Chandra Shekar Lakshminarayanan	Post doctoral fellow	100%	

d. PhD Candidates

Name	Role	% of salary support paid by Alberta Innovates Chair Grant	Scholarships / Awards / Special Info
Gabor Balazs	PhD candidate	34%	Graduated. Now at Quantitative Analyst, Causality Group, Madrid, Spain
Neil Burch	PhD candidate	0 %	
Pooria Joulani	PhD candidate	100%	
Ruitong Huang	PhD candidate	0	
Anna Koop	PhD candidate	0%	Sessional Instructor at Augustana Campus
Ashique Mahmood	PhD candidate	0	
Kory Mathewson	PhD candidate	0	NSERC, AITF, President's Doctoral Prize of Distinction
Ozlem Aslan	PhD candidate	0%	Graduated
Bernardo Ávila Pires	PhD candidate	0%	Graduated. Now a research scientist at DeepMind, London, UK.
Roshan Shariff	PhD candidate	33%	AITF Graduate Student Scholarship, University of Alberta Doctoral Recruitment Scholarship
Craig Sherstan	PhD candidate	0 %	Vanier Scholarship, AITF, Walter Johns Scholarship
Sina Ghiassian	PhD candidate	33%	
Eric Graves	PhD candidate	31%	QE II Scholarship
Jincheng Mei	PhD candidate	0 %	
Farzaneh Mirzazadeh	PhD candidate	0 %	
Mohammad Ajallooeian	PhD candidate	0 %	
Junfeng Wen	PhD candidate	0 %	AITF Graduate Student Scholarship
Marlos Machado	PhD candidate	0 %	AITF Graduate Student Scholarship
Leah Hackman	PhD candidate	0 %	
Trevor Davis	PhD candidate	0 %	
Zaheen Farrez Ahmad	PhD candidate	0 %	
Dustin Morrill	PhD candidate	0 %	AITF Graduate Student Scholarship, NSERC PGSD, Science Graduate Scholarship

Omar Rivasplata	PhD candidate	64%	QE II Scholarship
Kiarash Shaloudegi	PhD candidate	0	Now a PhD student at Imperial College, London, UK
Nadia Ady	PhD candidate	24%	NSERC CGS-M, Walter H. Johns Graduate Fellowship, Science Graduate Scholarship

e. MSc Candidates

Name	**Role**	**% of salary support paid by Alberta Innovates Chair Grant**	**Scholarships / Awards / Special Info**
Robert Post	MSc candidate	0	
Shun Jie Lau	MSc candidate	0	
Ann Edwards	MSc candidate	0	Graduated. Employed by General Dynamics Mission Systems-Canada
Brendan Bennett	MSc candidate	31%	
Tian Tian	MSc candidate	31%	
Xiaowei Hu	MSc candidate	100%	Graduated. Now at Microsoft
Yifan Wu	MSc candidate	31%	Graduated. Now a PhD student at Carnegie Mellon University
Nadia Ady	MSc candidate	0	Now a PhD student at the University of Alberta
Tim Yee	MSc candidate	0	
Vivek Veeriah	MSc candidate	31%	
Banafshe Rafiee	MSc candidate	39%	
Gautham Vasan	MSc candidate	0	
Alexandra Kearney	MSc candidate	0	
Jaden Travnik	MSc candidate	67%	QE II Scholarship, Alberta Graduate Student Scholarship
Dustin Morrill	MSc candidate	0 %	AITF Graduate Student Scholarship, NSERC PGSD, Science Graduate Scholarship
Gordon Holland	MSc candidate	0	NSERC CGS-M, Walter H. Johns Graduate Fellowship, Faculty of Science Graduate Scholarship
Dylan Brenneis	MSc candidate	0	

Kristopher DeAsis	MSc candidate	0	NSERC CGS-M
Michael Disyak	MSc candidate	0	
Eric Graves	MSc candidate	31%	QE II Scholarship. Now a PhD student at the University of Alberta
Juan Hernandez Garcia	MSc candidate	0	
David Quail	MSc candidate	0	
Kenneth Young	MSc candidate	0	QE II Scholarship, NSERC CGS-M
Shangtong Zhang	MSc candidate	0	

f. Other Members

Name	**Role**	**% of salary support paid by Alberta Innovates Chair Grant**
Beverly Balaski	Program administrator	100%
Adam Parker	Undergrad researcher	0
Michael (Rory) Dawson	Research engineer	0
Nikolaus Yasui	Undergrad researcher	100%
Zach Goldthorpe	Undergrad researcher	0
Tyler Lazar	Undergrad researcher	100%
Kenny Young	Undergrad researcher	0
Dylan Ashley	Undergrad researcher	0
Anshil Gandhi	Undergrad researcher	0
Flavius Poenaru	Undergrad researcher	0
Alex Rutar	Undergrad researcher	0
Jenny Lee	High school summer student July-August 2016	0
Claire Duong	High school summer student July-August 2016	0
Abdullah Ali Alghamd	High school summer student July-August 2016	0

g. Visitors

Name	Institution
Mark Ring	Cogitai
Dr. Ozaki	FHUR, Nagoya, Japan, 1 year
Dr. Ohtsuka	FHUR, Nagoya, Japan, 1 year
Johannes Gunther	Technical University of Munich, Germany
Geordie Rose	Kindred
Suzanne Gilder	Kindred
Olivia Norton	Kindred
Cam O'Neill	Kindred
Shotaro Miwa	MELCO
Tomoharu Takeuchi	MELCO
Mariyama Toshisada	MELCO
Song Zhang	Huawei Technologies Canada
Sebastien Bubeck	Microsoft Research
Elias Barenboim	Purdue University
Alan Malek	University of California, Berkeley
Shai Ben-David	University of Waterloo
Bodhisattva Sen	Columbia University, NY
Bo Pang	Shanghai Jiao Tong University, China (three months)
M. Schmid	Charles University, Prague (full Year)
M. Morvacik	Charles University, Prague (full Year)
Jefferson Huang	Stony Brook University
Shalabh Bhatnagar	Indian Institute of Science
Elliot Ludvig	University of Warwick
Clayton Connors	New York State University
Daniel Filan	University of California, Berkeley
Julian Zimmert	Humboldt University of Berlin (three months)
Shuai Li	Chinese University of Hong Kong (three months)
Michael Przystupa	University of British Columbia

Geoff Sutcliffe	University of Miami
Ishan Durugkar	University of Massachusetts
Thommen Karimpanal George	Singapore University of Technology and Design (six months)
Tom Everitt	Australian National University

6. COLLABORATIONS

Provincial	
Participants	Nature of Collaboration
Alberta Machine Intelligence Institute (Amii)	R. Sutton, D. Schuurmans, Cs. Szepesvári, Patrick Pilarski, and M. Bowling are among the ten principal investigators of this research centre at the University of Alberta. Total annual funding for Amii is $2M/year
Dr. Kim Adams, University of Alberta Rehab Med and Dr. Mahdi Tavakoli, Department of Electrical and Computer Engineering University of Alberta	Collaborative work with Patrick Pilarski and Richard Sutton 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. (CHRP $333K over three years)
Martin Ferguson-Pell, Faculty of Rehabilitation Medicine, University of Alberta	Collaboration between Patrick Pilarski 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.
Vivian Mushawar and others of the CFI SMART Neuroprosthetics Team, University of Alberta	Work with Richard Sutton and Patrick Pilarski on the use of machine learning to adapt and optimize neural interfaces and assistive robotic devices for people with motor disabilities.

L.M. Pilarski (Dept. Oncology, University of Alberta), L.M. McMullen (Agriculture, University of Alberta), M. Gaenzle (Agriculture, University of Alberta), X. Yang (Lacombe Research Centre)	With Patrick Pilarski. The meat processing industry needs a portable and rapid platform to detect pathogens during meat processing. We work with end users and colleagues to develop inexpensive computer systems that enable the automation of the testing system. This work includes software and hardware automation, biomedical pattern analysis, and machine intelligence.
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
Doina Precup, Pierre-luc Bacon, McGill University	Joint research with Richard Sutton on reinforcement learning architectures for planning, learning, and acting. Precup student Pierre-luc Bacon has been an extended visitor.
Harm van Seijen, Maluuba Inc (now owned by Microsoft)	Richard Sutton provided advice and guidance on reinforcement learning to Maluuba. Ongoing joint work with van Seijen
BLINC Lab. Jason P. Carey (Department of Mechanical Engineering, University of Alberta), Michael R. Dawson, Jacqueline S. Hebert, K. Ming Chan (Glenrose Rehabilitation Hospital).	With Patrick Pilarski and Richard Sutton (RLAI). "Bionic limbs for improved natural control" 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.
Foteini Agrafioti, Gabriel Woo, RBC Research	Richard Sutton is head advisor to RBC Research, which is making a machine-learning research laboratory in Edmonton
Geordie Rose, Suzanne Gilbert, Rupam Mahmood, and others at Kindred, a startup company located in Vancouver and Toronto	Richard Sutton is an official advisor to Kindred

International	
Participants	Nature of Collaboration
Song Zhang, Jun Luo, etc., Huawei	Joint research with Richard Sutton and Patrick Pilarski on reinforcement learning approaches to perception
Xinhua 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 computer vision and dimensionality reduction.
Yaoliang Yu, University of Waterloo	Joint research with Dale Schuurmans on efficient optimization methods for machine learning, convex reformulations of representation learning problems, and algorithmic techniques for exploiting structured sparsity.
Craig Boutilier, Google	Joint research with Dale Schuurmans on scalable methods for solving large scale approximate dynamic programming problems.
Martin Zinkevich, Google	Joint research with Dale Schuurmans on game theoretic and equilibrium methods for solving deep learning problems.
Mohammad Norouzi, Google	Joint research with Dale Schuurmans on entropy regularization and Bregman divergence minimization methods for reinforcement learning.
Branislav Kveton, Adobe Research, San Jose	Joint research with Csaba Szepesvári on bandit combinatorial optimization.
Andras György, Imperial College, UK	Joint research with Csaba Szepesvári on online learning.
Mohammad Ghavamzadeh, Adobe Research, San Jose	Joint research with Csaba Szepesvári on classification calibration.
David Pal, Yahoo Research, New York	Joint research with Csaba Szepesvári on online learning.
Yasin Abbasi-Yadkori, Adobe Research, San Jose	Joint research with Csaba Szepesvári on online learning in control.
Tor Lattimore, DeepMind London	Joint research with Csaba Szepesvári on online learning in bandit algorithms.
Venkatesh Saligrama, Boston University	Joint research with Csaba Szepesvári on online learning in unsupervised learning.
Viliam Lisy, Czech Technical University, Milind Tambe, USC	Joint research with Michael Bowling on exploring the application of extensive form game solvers to security game problems.
Erik Talvitie, Franklin and Marshall College	Joint research and supervision of graduate students with Michael Bowling, developing new algorithms for intelligent exploration in domains with sparse rewards.

Joel Veness and Marc Bellemare, Google DeepMind	Joint research with Michael Bowling on developing the Arcade Learning Environment, and algorithms for using compression in game theory.
Martin Schmid, Charles University, Matej Morvacik,Charles University, Viliam Lisy, Czech Technical University	Joint research with Michael Bowling on developing new algorithms for combining local search in imperfect information games.
J. Hebert, Craig Chapman (Phys Rec), Albert Vette (Mec ENG), and Cleveland Clinic (USA) and UNB (Canada)	With Patrick Pilarski. DARPA HAPTIX program to measure effectiveness of bi-directional neuro-prostheses. (data science and multimodal sensorymodal data streams)
Mitsubishi Electric Co. (J.P.)	Richard Sutton, Patrick M. Pilarski, Osmar Zaiane, Cameron Schuler, and others in Amii, New reinforcement learning techniques for dealing with industrial problems with uncertain data. This has become a substantial collaboration through Amii supported by Mitsubishi funds.
Sophia Adamia, Dana Farber Cancer Institute, Harvard Medical School, USA	With Patrick M. Pilarski. Investigation of cancer-related genetic markers in patients. Collaboration involving machine learning and biomedical data mining methods.
David Silver, DeepMind, London, UK	Joint research with Richard Sutton on reinforcement learning. This collaboration received funding of CAD$250,000 in support of RLAI activities in 2016.
Derek Pridmore and other principals at Osaro, a startup company in San Francisco, USA	Through skype meetings and visits in each direction, Sutton has provided advice and guidance on research directions for Osaro. Sutton is an official advisor to the company.
Cogitai, a new startup company founded by Peter Stone (University of Texas), Satinder Singh (University of Michigan), and Mark Ring	Sutton and Bowling are associated with Cogitai as advisors and as members of their "brain trust," researchers who can be drawn upon or can contribute as needed. Bowling took his sabbatical at Cogitai
Andrew Barto, University of Massachusetts at Amherst	Sutton and Barto are working together to produce a second edition of their successful textbook *Reinforcement Learning: An Introduction*.
Samsung SDS, San Jose, USA	Samsung SDS has made a major effort toward delivering reinforcement learning as a service, and Sutton is advising them as a consultant
CIFAR research project on Learning in Machines and Brains	Sutton is a member of this long-standing CIFAR research program, with members around the world. The program is led by Yann LeCun (New York University and Facebook) and Yoshua Bengio (Universite de Montreal). It is now largely focused on deep learning and unsupervised learning.
Doina Precup (McGill) and Elliot Ludvig (Warwick University)	Organization of the 10th Barbados Workshop on Reinforcement Learning, held April 23-28, 2017, in Holetown, Barbados. With Richard Sutton.

Satinder Singh (University of Michigan), Nathanial Daw (Princeton), Emma Brunskill (CMU), Joelle Pineau (McGill), Susan Murphy (University of Michigan), Rick Lewis (University of Michigan), Nick Roy (MIT)	Richard Sutton is general chair of the third Multi-disciplinary Conference on Reinforcement Learning and Decision Making, to be held on June 11-14, 2017, in Ann Arbor, Michigan, USA
Elliot Ludvig, Warwick University, and E. James Kehoe, University of New South Wales (retired)	Joint work with Richard Sutton on replay models of associative learning

7. GRADUATES

Name	**Degree**	**Research topic**	**% of salary support paid by Alberta Innovates Chair Grant**	**Current Position**
Yifan Wu	MSc	Online Learning under Partial Feedback	31%	PhD student at Carnegie Mellon University
Ann Edwards	MSc	Adaptive and Autonomous Switching: Shared Control of Powered Prosthetic Arms Using Reinforcement	0	Graduated, employed by General Dynamics Mission Systems-Canada
Xiaowei Hu	MSc	Bandit convex optimization with biased, noisy gradient oracles	100%	Employed at Mircosoft
Gabor Balazs	PhD	Convex Regression: Theory, Practice, and Applications	34%	Quantitative Analyst, Causality Group, Madrid, Spain
Bernardo Avila Pires	PhD	Model-Based Reinforcement Learning, Classification-based Policy Iteration,	0	Research scientist at DeepMind, London, UK
Ozlem Aslan	PhD	Convex Latent Modeling	0	

8. INTELLECTUAL PROPERTY

Intellectual Property	Status	Short Description
PATENTS	none	
LICENSES		
Spinoff Companies	none	

9. PUBLICATIONS

REFEREED JOURNAL PUBLICATIONS

K. Abou-Moustafa and F. Ferrie, "*Local Generalized Quadratic Distance Metrics. Application to the k Nearest Neighbour Classifier,*" Journal of Advanced in Data Analysis and Classification, 2017.

T.Y. Chun, J.Y. Lee, J.B. Park, and Y.H. Choi, "Integral temporal difference learning for continuous-time linear quadratic regulations," *Int'l. Journal of Control, Automation and Systems,* Feb. 2017, pp. 226-238.

T.Y. Chun, J.Y. Lee, J.B. Park, and Y.H. Choi, "Adaptive Dynamic Programming for Discrete-Time Linear Quadratic Regulation Based on Multirate Generalized Policy Iteration," *Int'l Journal of Control*, Mar. 2017.

A.L Edwards, M.R. Dawson, J.S. Hebert, C. Sherstan, R.S. Sutton, K.M. Chan, and P.M. Pilarski, "Application of Real-time Machine Learning to Myoelectric Prosthesis Control: A Case Series in Adaptive Switching," *Prosthetics & Orthotics International*, Vol. 40, no. 5, 573–581, 2016.

A. Farahmand, M. Ghavamzadeh, Cs. Szepesvári, and S. Mannor, "Regularized Policy Iteration with Nonparametric Function Spaces," *Journal of Machine Learning Research,* pp. 1−66, 2016.

M. Moravčík, M. Schmid, N. Burch, V. Lisý, D. Morrill, N. Bard, T. Davis, K. Waugh, M. Johanson, and M. Bowling, "DeepStack: Expert-level artificial intelligence in heads-up no-limit poker," *Science,* Mar. 2017.

P.M. Pilarski, and J.S. Hebert, "Upper and Lower Limb Robotic Prostheses," in *Robotic Assistive Technologies: Principles and Practice, Eds. P. Encarnacao and A. M. Cook*, pp. 99–144, 2017.

R. Vega, T. Sajed, K.W. Mathewson, K. Khare, P.M. Pilarski, R. Greiner, G. Sanchez-Ante, and J.M. Antelis, "Assessment of feature selection and classification methods for recognizing motor imagery tasks from electroencephalographic signals," *Artificial Intelligence Research*, Vol. 6, no. 1, pp. 37–51, 2017.

H. Yu, "Weak Convergence Properties of Constrained Emphatic Temporal-Difference Learning with Constant and Slowly Diminishing Stepsize," *Journal of Machine Learning Research,* vol. 17(220) pp.1–58, 2016.

HIGHLY REFEREED ARCHIVAL CONFERENCE PROCEEDINGS

B. Avila Pires, and Cs. Szepesvari, "Policy Error Bounds for Model-Based Reinforcement Learning with Factored Linear Models," *Proc. 29th Annual Conf. on Learning Theory* (COLT 2016), Jun. 2016, 32% acceptance.

F. Lattimore, T. Lattimore and M. Reid, "Causal Bandits: Learning Good Interventions via Causal Inference," *Proc. 30th Annual Conf. on Neural Information Processing Systems* (NIPS 2016), Dec. 2016.

A. Gariviar, T. Lattimore and E. Kaufmann, "On Explore-Then-Commit Strategies," *Proc. 30th Annual Conf. on Neural Information Processing Systems* (NIPS 2016), Dec. 2016.

S. Gerchinovitz and T. Lattimore, "Refined Lower Bounds for Adversarial Bandits," *Proc. 30th Annual Conf. on Neural Information Processing Systems* (NIPS 2016), Dec. 2016.

R. Huang, T. Lattimore, A. Gyorgy and Cs. Szepesvari, "Following the Leader and Fast Rates in Linear Prediction," *Proc.30th Annual Conf. on Neural Information Processing Systems* (NIPS 2016), Dec. 2016.

T. Lattimore, "Regret analysis of the finite-horizon Gittins index strategy for multi-armed bandits," *Proc. 29th Annual Conf. on Learning Theory* (COLT 2016), Jun. 2016.

J. Leike, T. Lattimore, L. Orseau and M. Hutter, "Thompson sampling is asymptotically optimal in general environments," *Proc. Int'l. Conf. on Uncertainty in Artificial Intelligence* (UAI 2016), Jun. 2016.

K. Shaloudegi, A. György and Cs. Szepesvári, "SDP Relaxation with Randomized Rounding for Energy Disaggregation," *Proc. 30th Annual Conf. on Neural Information Processing Systems* (NIPS 2016), Dec. 2016.

S. Katariya, B. Kveton, Cs. Szepesvári and Z. Wen, "DCM Bandits: Learning to Rank with Multiple Clicks," *Proc. 33rd Int'l Conf. on Machine Learning* (ICML 2016), Jun. 2016.

L.A. Prashanth, C. Jie, M. Fu, S. Marcus and Cs. Szepesvári, "Cumulative Prospect Theory Meets Reinforcement Learning: Prediction and Control," *Proc. 33rd Int'l Conf. on Machine Learning* (ICML 2016), Jun. 2016.

A. György and Cs. Szepesvári, "Shifting Regret, Mirror Descent, and Matrices," *Proc. 33rd Int'l Conf. on Machine Learning* (ICML 2016), Jun. 2016.

S. Wang, A. Mohamed, R. Caruana, J. Bilmes, M. Plilipose, M. Richardson, K. Geras, G. Urban, and O. Aslan, "Analysis of Deep Neural Networks with Extended Data Jacobian Matrix," *Proc. 33rd Int'l Conf. on Machine Learning* (ICML 2016), Jun. 2016.

V. Ganapathiraman, X. Zhang, Y. Yu and J. Wen, "Convex Two-Layer Modeling with Latent Structure," *Proc. 30th Annual Conf. on Neural Information Processing Systems* (NIPS 2016), Dec. 2016.

T. Yee, V Lisy, and M. Bowling, "Monte Carlo Tree Search in Continuous Action Spaces with Execution Uncertainty," *Proc. 25th Int'l Joint Conf. on Artificial Intelligence* (IJCAI 2016), Jul. 2016.

K. Milan, J. Veness, J. Kirkpatrick, M. Bowling, A. Koop, and D. Hassabis, "The forget-me-not process," *Proc. 30th Annual Conf. on Neural Information Processing Systems* (NIPS 2016), Dec. 2016, 23% acceptance.

OTHER CONFERENCE AND WORKSHOP PROCEEDINGS

C. Sherstan, A. White, M. C. Machado, and P.M. Pilarski, "Introspective Agents: Confidence Measures for General Value Functions*," Proc. Conf. on Artificial General Intelligence* (AGI 2016), Jul. 2016, 39% acceptance.

Q. Boser, A. Valevicius, E. Lavoie, C.S. Chapman, P.M. Pilarski, J.S. Hebert, and A.H. Vette, "Comparison of Anatomical and Cluster-Based Upper Body Marker Models," *40th Annual Meeting of the American Society of Biomechanics*. Aug. 2016.

A. Valevicius, Q. Boser, E. Lavoie, A.H. Vette, C.S. Chapman, P.M. Pilarski, and J.S. Hebert, "Normative Kinematic Data for Two Functional Upper Limb Tasks," *40th Annual Meeting of the American Society of Biomechanics.* Aug. 2016.

D.P. Manage, J. Lauzon, P. Ward, P.M. Pilarski, L.M. Pilarski, and L.M. McMullen, "Cassette PCR for Rapid Detection of pathogenic Escherichia coli in Meat," *Biodefense World Summit*, Jun. 2016.

P.M. Pilarski, and C. Sherstan, "Steps Toward Knowledgeable Neuroprostheses," *Proc. 6th IEEE RAS/EMBS Int'l Conf. on Biomedical Robotics and Biomechatronics* (BioRob2016), Jun. 2016.

A.L. Edwards, J.S. Hebert, and P.M. Pilarski, "Machine Learning and Unlearning to Autonomously Switch Between the Functions of a Myoelectric Arm," *Proc 6th IEEE RAS/EMBS Int'l Conf. on Biomedical Robotics and Biomechatronics* (BioRob2016), Jun. 2016.

P.M. Pilarski, R.S. Sutton, A.L. Edwards, C. Sherstan, K.W. Mathewson, A.S.R. Parker, and J.S. Hebert, "Towards strong prosthetic machine intelligence," *Proc. of the Cybathlon Symposium.* Oct. 2016.

N.M. Ady, and P.M. Pilarski, "Domains for Investigating Curious Behaviour in Reinforcement Learning Agents*," Proc. 11th Women in Machine Learning Workshop* (WiML 2016), Dec. 2016.

N. Burch, M. Schmid, M. Moravcik, and M. Bowling, "AIVAT: A New Variance Reduction Technique for Agent Evaluation in Imperfect Information Games," *Proc. Workshop on Computer Poker and Imperfect Information Games* (AAAI 2017), Feb. 2017.

S. Ghiassian, B. Rafiee, and R.S. Sutton, "A first Empirical Study of Emphatic Temporal Difference Learning," *Proc. NIPS Workshop on Deep Reinforcement Learning Workshop* (CLDL 2016), Nov. 2016.

P. Joulani, Cs. Szepesvári, and A. György, "Generalized Online and Stochastic Optimization: Unification, Adaptivity, Optimism, and Non-Convexity. Neural Information Processing Systems," *Proc. Workshop on Optimization* (OPT-16), Dec. 2016.

V. Lisý, and M. Bowling, "Equilibrium Approximation Quality of Current No-Limit Poker Bots," *Proc. Workshop on Computer Poker and Imperfect Information Games* (AAAI 2017), Feb. 2017.

M. C. Machado, and M. Bowling, "Learning Purposeful Behaviour in the Absence of Rewards," *Proc. 33rd Int'l Conf. on Machine Learning Workshop on Abstraction in Reinforcement Learning* (ICML 2016), Jun. 2016.

K.W. Mathewson, and P.M. Pilarski, "Reinforcement Learning based Embodied Agents Modelling Human Users Through Interaction and Multi-Sensory Perception," *Proc. AAAI Spring Symposium on Interactive Multi-Sensory Object Perception for Embodied Agents* (AAAI 2017), Mar. 2017.

K.W. Mathewson, and P.M. Pilarski, "Simultaneous Control and Human Feedback in the Training of a Robotic Agent with Actor-Critic Reinforcement Learning," *Proc. 25th Int'l Joint Conf. on Artificial Intelligence, Interactive Machine Learning Workshop* (IJCAI-2016), Jul. 2016.

V. Veeriah, P.M. Pilarski, and R.S. Sutton, "Face valuing: Training user interfaces with facial expressions and reinforcement learning," *Proc. 25th Int'l Joint Conf. on Artificial Intelligence, Interactive Machine Learning Workshop* (IJCAI-2016), Jul. 2016.

J. Wen, N. Hassanpour, and R. Greiner, "Weighted Gaussian Process for Estimating Treatment Effect," *Proc. NIPS Workshop on Inference and Learning of Hypothetical and Counterfactual Interventions in Complex Systems* (NIPS 2016), Dec. 2016.

B. Xu, R. Huang, and M. Li, "Revise Saturated Activation Functions," Workshop track in Int'l Conference on Learning Representations (ICLR) May. 2016.

SPECIAL/INVITED PRESENTATIONS

Person	Title	Venue
R. Sutton	Toward a New View of Action Selection: The Subgoal Keyboard	10th Barbados Workshop on Reinforcement Learning
R. Sutton	The Future of Artificial Intelligence Belongs to Search and Learning	University of Toronto Computer Science seminar
R. Sutton	Learning Representations by Stochastic Gradient Descent in Cross-Validation Error	Deep Reinforcement Learning Workshop at Neural Information Processing Systems Conference, Barcelona, Spain
R. Sutton	An Outline of an AI Architecture for Continual Learning	Continual Learning and Deep Learning Workshop at Neural Information Processing Systems Conference, Barcelona, Spain
R. Sutton	An Outline of an AI Architecture for Continual Learning and Neurorobotics	Neurorobotics Workshop at Neural Information Processing Systems Conference, Barcelona, Spain
D. Schuurmans	Deep Learning Games	Google DeepMind, London UK
D. Schuurmans	Relating Forward and Inverse Reinforcement Learning	Google DeepMind, London UK
D. Schuurmans	Deep Learning Games	Google Brain
D. Schuurmans	Relating Forward and Inverse Reinforcement Learning	Google Brain
Cs. Szepesvári	Explore-exploit tradeoffs	Samsung Research
Cs. Szepesvári	Explore-exploit tradeoff	Budapest University of Technology and Economics (BME)
Cs. Szepesvári	Factored linear models in reinforcement learning	Google DeepMind, London UK
M. Bowling	Deep Learning Games	HPCS
M. Bowling	Games and Intelligence of the Artificial Kind	CAIMS Public Lecture
M. Bowling	Computer Curling: AI in Sports Analytics	NIPS Workshop on Deep Reinforcement Learning
M. Bowling	DeepStack	AAAI Workshop
M. Bowling	DeepStack	AAAI Special Session

M. Bowling	Games and Intelligence of the Artificial Kind	Charles University
M. Bowling	AI Goes All-In	IBM Prague
M. Bowling	AI Goes All-In	Czech Technical University
M. Bowling	AI Goes All-In	DeepMind. London, UK
P. M. Pilarski	Communicative Capital: Bridging the Gap between Users and their Assistive Technology	Banff, Alberta
P. M. Pilarski	Reinforcement Learning Algorithms for Acquiring and Using Predictive Knowledge	IBM Research – Zurich

AWARDS

Michael Bowling's "The DeepStack" paper is #59 out of all research stories of similar age: https://www.altmetric.com/details/16939457#score

Richard Sutton has been named one of Alberta Venture's 50 Most Influential Albertans of 2016 for his research in reinforcement learning [http://bit.ly/29bRTJd]

Richard Sutton is among the top 20 most influential computer(ing) scientists in this world.
http://www.sciencemag.org/news/2016/04/who-s-michael-jordan-computer-science-new-tool-ranks-researchers-influence

Richard Sutton has been elected a Fellow of the Royal Society of Canada. RSC Fellows have been elected by their peers in recognition of outstanding scholarly, scientific and artistic achievement. Election to the academies of the Royal Society of Canada is the highest honor a scholar can achieve in the Arts, Humanities and Sciences.

Richard Sutton (with Geoff Hinton and Yoshua Bengio) won the University of Toronto's Creative Destruction Lab's *Ideas* award "for outstanding contributions to Canada's competitiveness in the world economy."

THESES

Yifan Wu, MSc, "Online Learning under Partial Feedback," July 2016.

Xiaowei Hu, MSc, "Bandit convex optimization with biased, noisy gradient oracles," January 2017.

Ann Edwards, MSc, "Adaptive and Autonomous Switching: Shared Control of Powered Prosthetic Arms Using Reinforcement Learning," November 2016.

Gabor Balazs, PhD, "Convex Regression: Theory, Practice, and Applications," November 2016.

Bernardo Avila Pires, PhD, "Model-Based Reinforcement Learning, Classification-based Policy Iteration, Classification" June 2017.

Ozlem Aslan, PhD, "Convex Latent Modeling," March 2017

10. OUTREACH

Michael Bowling conducted approximately 40 interviews and media appearances related to DeepStack. See: https://docs.google.com/a/ualberta.ca/spreadsheets/d/1I4Uk9meIn4O5nfx_h2v4jkFhZ7hAW17eNcUv-IfgQfY

Anna Koop volunteered as a "Human book" speaking about depression for the Augustana campus library "Human library" session. About 12 students attended over two sessions.

Anna Koop developed and assisted with the Enigma Machine demonstration and Q&A at the Telus World of Science in Edmonton. August 2016.

Anna Koop organized and moderated a panel on "Careers in Computing Science" for undergraduate students at Augustana campus (25 attended). April 2016.

Anna Koop assisted with an "Introduction to Programming Workshop" for WISEST summer students at the Computing Science department at the University of Alberta.

Anna Koop presented on "Evaluating Information: The Internet is your Frenemy" at the Critical Thinking Workshop for undergraduates at Augustana Campus (20 attended).

Leah Hackman gave an overview of programming in Python and helpers assisted one-on-one for WISEST summer students at the Computing Science department at the University of Alberta (30 attended).

Patrick M Pilarski, Adam Parker, Dylan Brenneis, Jaden Travnik, and Nadia Ady, presented at "Dark Matters", Telus World of Science. Approximately 1000 attended.

Patrick M. Pilarski organized and coordinated a field trip for Memorial Composite High School, Stony Plain, Alberta. 55 students and two teachers from their computing and sport medicine classes (Grades 10-12) visited the BLINC Lab, the Rehab Robotics Lab, and the Department of Computing Science at the University of Alberta for a day-long session of hands-on demonstrations, activities, and learning sessions with faculty and graduate students.

Patrick M. Pilarski organized and coordinated a field Trip to the Pilarski lab for approximately 20 students from the Two Hills Menonite School.

Patrick M. Pilarski gave a demonstration and open-source launch of our Bento Arm platform at Maker Faire 2016, Bay Area, May 2016

Michael (Rory) Dawson, Demonstrated the Bento Arm and new software at the University of Alberta FGSR showcase at the Telus World of Science, April 2017.

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

Csaba Szepesvári gave a presentation, "We ain't ever scared of no bandits," at the "Lunch'n'learn" seminar of the High-School Internship Program, the University of Alberta Department of Computing Science (50 students attended).

Leah Hackman presented at the WISEST Introduction to Python Session, August 2016 (18 students attended).

Brendan Bennett gave a presentation to the Calgary NetEnergy Brokerage "Machine Learning" (20 attended).

Omar Rivasplata was a volunteer instructor at CMS Summer Camp for Junior High students. He ran a problem solving session on probability, July 2016.

Richard Sutton gave numerous interviews to the press on AI in general and on the RBC Research efforts in AI and in their Edmonton lab in particular.