ICORE ANNUAL REPORT MARCH 31, 2005

EXECUTIVE SUMMARY

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

The objectives of the reinforcement learning and artificial intelligence (RLAI) research program are to create new methods for reinforcement learning that remove some of the limitations on their widespread application, and to develop reinforcement learning as a model of intelligence that could approach human abilities.

In this second year, the RLAI research team has grown to about 32 members, 21 of which are graduate students and, of those, 11 are recipients of major scholarships. The output of the research program has also grown, with 21 papers published or accepted for publication in archival venues during the reporting period.

The primary focus of the research program this year has been on how intelligent machines represent their knowledge of the world. The key issue here is expressing the knowledge in such a way that it can be verified, learned, and used without continual human input. This project has pursued an unusual research strategy in which the knowledge is expressed in terms the machines' sensors and actuators, which enables verification and learning. A new knowledge and learning technology has been developed that generalizes from low-level sensations and actions to higher-level concepts that could enable faster solution of much larger problems. Several proof-of-concept systems have been developed this year using small problems; larger problems will be used next year to extend and test the ideas. These results may be a conceptual advance toward a long-standing objective of informatics: a general technology of learning and reasoning about the optimal control of dynamical systems.

RESEARCH PROGRAM OVERVIEW

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

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

- One proposed focus was on *predictive state representations*, a new set of ideas for modeling sequential decision problems that are not Markov, that is, for which an appropriate state representation is not available a priori but must be constructed from the stream of raw sensor data.

  This year we have introduced a new formulation, *temporal-difference networks*, that extends prior work on predictive state representations to include temporal-difference learning, the core component of most RL algorithms. Temporal-difference networks appear to be a qualitative advance over previous work in machine learning in that they explicitly represent the *goals* of the learning process. This enables a wide variety of complex goals, and enables the choice of goals to be made by the machine rather than by people.

- A second proposed focus was on *temporal abstraction*, the extension of RL beyond the flat and low-level representations commonly used with MDPs to the more flexible, structured, and higher-level representations used by AI systems. The ability to represent knowledge about possible courses of action at a multiplicity of interrelated temporal scales would vastly increase the generality and range of application of RL methods. The existing theory of options was proposed to be extended in this direction.

  This year we have integrated the options framework with temporal-difference networks. Simple temporal-difference networks make predictions at each time step that are conditional on the primitive action taken on that time step. Our extension enables the predictions to be conditional on options—multi-step closed-loop ways of behaving—rather than only on primitive actions. The options framework itself had to be substantially generalized in several ways to permit this extension. Work to date has been proof-of-concept demonstrations on small problems.

- A third proposed focus was broadly concerned with approximation and generalization in RL. Approximation is required in all large-scale applications yet is incompletely understood in both theory and practice. It was proposed to create an RL toolkit to facilitate the use of approximation and other RL techniques in applications.

  In this year we have completed and released the first version of the RL toolkit. RL Toolkit 1.0 is currently used in this research program and by reinforcement learning researchers throughout the world.

- The fourth proposed area of investigation was the demonstration of advances in the first three areas in robotics applications.

  This year we have established a dedicated robotics laboratory, acquired a variety of new robots, and completed our first robotic demonstrations of reinforcement learning.

These achievements are discussed in more detail in the following section.

The primary focus of the research program this year has been on means by which an intelligent machine can represent knowledge of its environment or problem domain. Knowledge representation is widely recognized as critical to the performance of all AI systems, and critical to knowledge representation is expressing the knowledge in such a way that it can be verified, learned, and used without continual human monitoring. This project has pursued an unusual research strategy in which the knowledge is directly related to the machine's sensors and actuators. Temporal-difference networks with options are used to abstract from these low-level signals to the higher-level concepts required to solve very large problems, while maintaining verifiability and learnability. Tests on small problems have been encouraging. Over the next years we plan to extend these ideas and apply them to larger problems with the hope of addressing this long-standing challenge to all AI systems.

The department of computing science has dedicated room CSC-265 as RLAI lab space this year. This room has space for about 12 students to work at desks plus a large meeting and reading area. A second laboratory (CSC 140, 20.6m^2) is dedicated for RLAI research in robotics.

RESEARCH PROJECTS

Temporal Difference Networks

It has long been a goal for AI systems to be able to express their knowledge in terms of their interaction with the world. Connecting knowledge to experience provides a way of verifying and learning it autonomously. This ability may be essential to the practical development of large AI systems. This year we have taken a few steps further toward the goal of grounding knowledge in experience. We have introduced a new framework and learning algorithm called "temporal-difference networks." Temporal-difference networks translate world knowledge into explicit predictions of future experience that can be compared with actual experience. The major conceptual advance was having the predictive questions being asked about experience be in machine readable form (and not just their answers); prior work with prediction learning has taken the questions to be implicit, in the human designer's mind but not in the machine's. Representing them explicitly enables the set of questions being asked to be large and subject to autonomous elaboration.

Grounding knowledge in experience is challenging because knowledge is high-level and conceptual whereas experience is low-level and sensori-motor. Several levels and types of abstraction are required to bridge the gulf. Temporal-difference networks combine two recently developed abstraction technologies: *predictive state representations* for abstracting over state, and the *options* framework for abstracting over time. Temporal-difference networks are still a very new idea and their strengths and limitations have yet to be fully determined. This year we have introduced the idea and shown some of its potential in micro-worlds—small problems where the knowledge and abstractions that need to be learned can be completely understood by us.

Temporal-difference networks have been the primary focus of RLAI research this year. Four papers on this subject have been submitted and accepted for publication. The paper introducing the idea was accepted for oral presentation at the prestigious Conference on Neural Information Processing Systems. Only 2% of the more than 800 papers submitted were selected for oral presentation at that meeting. Two other papers were accepted for publication at the International Joint Conference on Artificial Intelligence. One concerned extending temporal-difference networks to include history information, as in k-order Markov models. The other was a test of the utility of the primary idea underlying predictive state representations. The fourth paper extended temporal-difference networks to incorporate eligibility traces and to provide a conceptual bridge from them to Monte Carlo methods. A fifth paper describing the extension to include options is in preparation.

Function Approximation, Off-policy Learning, and Recognizers

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

This year we have empirically tested and mathematically analyzed a variety of off-policy learning algorithms, including least-squares methods and importance-sampling methods. A severe problem with these methods is that they have very high variance, causing learning to be very slow. As part of the work with temporal-difference networks, we have developed a new concept, that of a "recognizer," which appears promising in this regard. A recognizer observes behavior and accepts it, or not, as something that it is learning about. It *recognizes* a portion of the behavior as, in effect, corresponding to the target policy. Recognizers are used to condition the predictions made by temporal-difference networks on the options taken. Recent experiments suggest that importance sampling using recognizers has lower variance and is much better behaved than previous importance-sampling methods for off-policy learning.

RL Toolkit

The RL Toolkit is a collection of software and guidelines to facilitate the development of RL research and applications. The first version of the toolkit was completed this year. RL Toolkit 1.0 includes 1) a standard interface for connecting reinforcement learning algorithms to reinforcement learning problems; 2) a variety of examples of using the interface; 3) an efficient implementation of tile coding, a popular function approximation method for domains with continuous state variables; 4) an implementation of eligibility traces for tile coding, 5) a graphics language "G" that is portable across Apple, Unix, Linux, and Windows machines; 6) utilities for easily generating two and three dimensional graphs using G, and 7) five programs demonstrating reinforcement learning and the use of the toolkit. Early in the year we switched our primary development language from Lisp to *Python*, a popular open-source programming language that is available for almost any kind computer and that is easy to learn and use. Almost all of RL Toolkit 1.0 is written in Python. Some of the lower-level tools are written in the language C++ to maximize efficiency.

The RL Toolkit is used within the RLAI group and by other researchers and students throughout the world. It has been used in two courses on reinforcement learning at the University of Alberta, leading to several of the examples that are now part of the toolkit proper. The RL Toolkit software has been placed in the public domain. Instructions for downloading and installing it are available on the RLAI web site discussed below.

Open Pages and the RLAI Web Site

An extensive new web site has been developed to coordinate and encourage RLAI research across the world. The web address for the site is http://rlai.net. The web site contains information about this project, pointers to other reinforcement learning research groups, substantial discussion of a variety of research topics, and information about the RL Toolkit and related RL software tools.

The RLAI website was created using a new set of tools for collaborative authoring of web pages which were developed as part of this research project. The *open pages* system has three key features. First, it encourages a community of researchers to create and author web pages jointly. Literally anyone can edit or add to any open page. The open pages software prevents two authors from accidentally writing over each other's work and provides an archiving mechanism such that previous versions of an open page can be restored if necessary. The second key feature of open pages is that they are edited "in place." An author or editor does not download and work with web pages locally and then transfer them to the website. Instead the web pages are presented as directly editable and saved out to the web site just as one might save a file locally. These first two features are available with existing software tools such as "wikis," but the open page system appears to be unique in its third feature, which is that its pages can be edited in the WYSIWYG (What You See Is What You Get) style that we are all familiar with from modern word processors. That is, as you edit a page it appears just as it will when published on the web. This is in contrast to wikis, which are edited only in a low-level text mode in which textual codes are used for formatting of headings, fonts, bolding and so forth. These three features work well together in encouraging authoring and frequent maintenance of web pages, which in turn encourages their timeliness and accuracy.

Robotics Research

The RLAI robotics effort has acquired $20.6m^2$ of dedicated laboratory space on the ground floor of the Computing Science Center. Several new robots have been acquired over the past year, including 8 Aibo robots, a Pioneer mobile robot with laser range-finder, and a Segway robot. The recently commercialized Segway Robot Mobility Platform (RMP) is a robust, fast, industrial-strength, indoor-outdoor robot platform. Based on the Segway Human Transporter (HT), it provides an excellent opportunity for exploring human-robot coordination. In the past six months an RMP and HT have been acquired and the basic work has been done to prepare the RMP as a research platform for the development of new machine learning techniques.

Action-Respecting Embedding

A map is a key component for a mobile robot. Maps at their core allow the robot to answer three questions: (1) "where have I been?" (2) "where am I now?" and (3) "how do I get where I want to go?" A huge body of robotics research assumes their existence, and another large body of research tries to build them. But building maps is time-consuming, manually intensive, and requires expert knowledge in the form of detailed models of the robot's motion and sensor apparatus. This project investigates how maps can be learned directly from the robot's own subjective experience of sensations and actions, without any models. A new algorithm has been developed, Action Respecting Embedding (ARE), inspired by kernel-based dimensionality-reduction techniques. ARE extracts a low dimensional representation of data that also respects the provided action labeling. The resulting subjective map explicitly encodes the robot's trajectory (answering question one), and can be used for both planning (question three) and localization (question two). Although originally conceived in the context of mobile robots, ARE is a general technique for extracting representations from a sequence of observations and actions.

Graphical Models for Optimal Point Matching

Point set matching is a fundamental problem of computer vision where one seeks an optimal match between a template point set and a larger image point set. This year we developed the first efficient algorithm that is guaranteed to solve this problem exactly, even though, for decades, computer vision researchers have been using approximate "relaxation" methods that are not guaranteed to produce an optimal matching. We have found that we can solve this problem by taking a new approach that combines graphical probability models and the theory of rigid graphs. By formulating rigid graph (point set) matching as a problem of finding a maximum probability configuration in a graphical model, we have developed the first provable, efficient, and exact point set matching algorithm for computer vision.

OBJECTIVES FOR NEXT YEAR

The research team will continue to grow next year as several new students are expected. We also expect to hire one or two new postdoctoral fellows; two offers have been made to top candidates and are outstanding. Arrangements have been made with the Faculty of Science to hire a research faculty member, but a suitable candidate has not yet been found.

A major focus in the upcoming year will continue to be on temporal-difference networks as a framework for grounding world knowledge in subjective experience. A paper on combining temporal-difference networks with abstraction will be submitted to the Conference on Neural Information Processing Systems in June. There are several further steps to be taken with regard to temporal-difference networks: the learning methods must be made more efficient, off-policy learning methods must be incorporated, methods for automatically discovering questions must be found, and generally more experience and intuition have to be obtained regarding their power and limitations. The near term objective is to develop a few implemented examples to the point where a strong case can be made for the possibility and strength of grounding knowledge in experience, and then to prepare a journal article laying out the fundamental ideas of this approach. Some of these implemented examples may be with robots.

Function approximation methods will also continue to be a focus of our research. A paper on the theory and practice of the off-policy learning of recognizers with linear function approximation will be prepared and submitted for publication, probably to the Conference on Neural Information Processing Systems. We will also be exploring methods whose performance is similar to least-squares methods, but are of lower complexity and thus can be applied to larger problems. Methods for learning efficiently by de-correlating input signals will also be explored.

Work on the RL Toolkit over the next year will focus on benchmarks and inter-language integration. The RL interface makes it possible to connect learning agents to environments very flexibly, and this in turn enables the development of benchmarks for comparing different algorithms. Such benchmarks are widely recognized as a pressing need in reinforcement learning, and we intend to take significant steps in this direction. The RL Toolkit consists of software in several languages, most notably in Python, which is preferable for rapid prototyping, and C++, which is more efficient in computer time. We will continue to develop tools facilitating communication between software modules written in different languages. A second major release, RL Toolkit 2.0, is planned for next year. Further development of open pages and of the G graphics system are also envisaged.

One of the robotics projects planned for next year is in the new area of "any-team" cooperative robotics. The state of the art in building a team of robots (or software agents) that can cooperate to achieve a task has vastly improved in recent years. The RoboCup Initiative is one example of this, where the top teams can play a very effective game of soccer, given the mobility and sensor capabilities of the robots. A very similar task, but vastly more complex, is the problem of building a single robot to play "pick-up soccer". Pick-up soccer is an example of an impromptu multi-agent system, where many independently developed agents are combined into an impromptu team to carry out a common task. The independence of the agents make carefully tuned coordination strategies ineffective, and make adaptation a critical component to effective behavior on a variety of teams. We are planning on building an "any-team" robot for the RoboCup small-size league. The goal is a scaled-back version of pick-up soccer, where one robot will be added to an existing team of robots, replacing one of their teammates. The vast number of existing capable teams provide an excellent testbed. The robot will not only seek to determine its role within the team, it will also adapt its attempts to coordinate through the use of a previously developed play-based team coordination system. Evaluation will initially be modest, seeking for the team performance to be better than if that robot had been replaced with a stationary obstacle.

RESEARCH TEAM MEMBERS AND CONTRIBUTIONS

The project holds weekly group meetings attended by all faculty team members.

Faculty Team Members

Name	Role
Richard Sutton	Team leader and Principal investigator Reinforcement learning and artificial intelligence
Michael Bowling	Principal investigator Robotics, multi-agent systems, reinforcement learning
Dale Schuurmans	Principal investigator, CRC II Chair Probabilistic methods in artificial intelligence, machine learning
Vadim Bulitko	Associated faculty Decision-making in artificial intelligence
Mark Ring	Visiting professor, Chapman University Continual learning, neural networks, sequence learning. Dr. Ring participates half-time in RLAI research by special arrangement as part of a collaboration with researchers at the Universities of Rutgers, Michigan, and Massachusetts
Katsunari Shibata	Visiting professor, Oita University, Feb-August 2005 Reinforcement learning, neural networks

Postdoctoral Fellows

Name	Topic/notes
Finnegan Southey	Machine learning, opponent modeling
Li Cheng	Bayesian image modeling
Tiberio Caetano	Structural pattern recognition
Yaakov Engel	Gaussian process reinforcement learning Hired, to begin in April 2005 Alberta Ingenuity postdoctoral fellowship
Martin Zinkevich	Minimal regret algorithms Hired, to begin Sept 2005

PhD Students

Name	Supervisor	Scholarship/Notes
Tao Wang	Schuurmans/Bowling	
Dana Wilkinson	Schuurmans	University of Waterloo
Adam Milstein	Schuurmans	NSERC PGS University of Waterloo
Qin Wang	Schuurmans/Dekang Lin	iCORE, Provost Doctoral Entrance Award
David Silver	Sutton/Martin Mueller	FS Chia, AIF upcoming
Yuhong Guo	Schuurmans/Russ Greiner	
Feng Jiao	Schuurmans	
Linli Xu	Schuurmans	
Daniel Lizotte	Schuurmans/Russ Greiner	NSERC, Killam/Steinhauer upcoming
Ali Ghodsi	Schuurmans/Brendan Frey	Ontario graduate scholarship University of Waterloo
Jiayuan Huang	Schuurmans	Summer internship, Max Planck Institute

MSc Students

Name	Supervisor	Scholarship/notes
Brian Tanner	Sutton	NSERC PGS-M, Killam, AIF
Peter McCracken	Bowling	NSERC, Jeffery R. Sampson Graduate Memorial Prize
Nolan Bard	Bowling	
Lihong Li	Bulitko	Graduated, now a PhD student at Rutgers University
Cosmin Paduraru	Sutton/Bulitko	iCORE
Eddie Rafols	Sutton	NSERC upcoming
Adam White	Sutton	NSERC upcoming
Alborz Geramifard	Sutton/Bowling	
Armita Kaboli	Bowling/Petr Musilek	ECE student
Anna Koop	Undergraduate/MSc	Will be MSc with AIF scholarship in fall
James Neufeld	Undergraduate/MSc	Will be MSc in fall

Other team members

Name	Role	Notes
Matt Stephure	Team leader, Segway and robot simulation projects	
Jason Roberts	Segway programming	

Adrien Melanson	Segway and robot simulation programming	
Stephanie Schaeffer	RL Toolkit programming	Part-time
Samantha Vincent	Aibo programming	
Wesley Lo	Programming, Action-Respecting Embedding	
Mark Lee	RL Toolkit and open pages programming	
Lori Troop	Administrative assistant	Part-time
Priyanka Pareek	Open pages programmer	Part-time

COLLABORATIONS

Provincial	
Participants	Nature of Collaboration
Alberta Ingenuity Center for Machine Learning	Sutton, Schuurmans, and Bowling are among the seven principle investigators for this center at the U. of Alberta. AICML contributes substantially to the RLAI project, totaling approximately $100K for the reporting period

National	
Participants	Nature of Collaboration
Doina Precup, McGill University	Joint research on algorithms for off policy learning, temporal abstraction. A grant proposal is being prepared for submission to NSERC
Yoshua Bengio, McGill Hugh Chipman, U. Waterloo Christian Leger, U. Montréal William Welch, U. British Columbia Jim Ramsay, McGill Mu Zhu, U. Waterloo	MITACS grant "Statistical Learning of Complex Data with Complex Distributions" to Schuurmans, Sutton, and the listed collaborators $170K/year, RLAI portion ~$25K
Jonathan Schaeffer, Robert Holte, Duane Szafron, and Michael Buro University of Alberta	NSERC Strategic Grant, "Intelligent Agents for Interactive Entertainment" to Bowling, Schuurmans, and the listed collaborators $195K/year, RLAI portion ~$30K

International	
Participants	Nature of Collaboration
Satinder Singh, University of Michigan	Joint research with Sutton on predictive state representations and temporal-difference networks, conducted primarily by video-conference
Peter Stone, University of Texas at Austin	Joint research with Sutton on applying reinforcement learning to RoboCup soccer
Michael Littman, Rutgers University Satinder Singh, University of Michigan Andrew Barto, University of Massachusetts	DARPA grant "Intrinsically Motivated Reinforcement Learning" to Sutton with the listed participants. This grant funds Mark Ring's participation in RLAI $480K/year, RLAI portion ~$36K
Jim Kehoe, University of New South Wales, Australia	Joint research with Sutton and Neufeld on the relationship between reinforcement learning and learning in animals. Visits were exchanged in the spring and summer; work proceeded remotely through the reporting period.
Michael Littman, Rutgers University	Prof. Littman visited the RLAI group for six and a half weeks in summer 2004 for joint research on reinforcement learning and predictive state representations
Frederic Maire, Queensland University of Technology, Australia	Prof. Maire visited Bulitko for the month of November 2004 for joint research
Eric Wiewiora, University of California, San Diego	Wiewiora is a PhD student who visited the RLAI group for three weeks for joint research on predictive state representations

Industrial	
Participants	Nature of Collaboration
Electronic Arts	Bowling and Rob Holte presented a machine learning tutorial at Tiburon studios in Florida
Syncrude Research Centre Ian Parsons et alia	Sutton (with Robert Holte and Bill Bridger) discussed possible applications of reinforcement learning with Syncrude scientists.
IBM, the Minister of Alberta for Innovation, the University of Alberta	After extensive negotiation, an IBM Centre for Advanced Studies was agreed to be established at the University of Alberta. RLAI team members played a role in this coup for Alberta by identifying research projects of joint interest between them and IBM researchers
Syncrude Research. Ltd, Centre for Intelligent Data-Mining at the U of A	Bulitko on machine learning for vision systems
Automated Vision Systems, Inc., Los Gatos, California	Bulitko discussed possible technology transfer for the machine learning for vision systems project.

RLAI team members hosted many prominent short-term visitors during the reporting period, including John Lafferty (Carnegie Mellon University), Steven Boyd (Stanford University), Douglas Aberdeen (Australian National University), Ronald Parr (Duke University), Brett Browning (Carnegie Mellon University), Jacob Crandall (Brigham Young University), Mohammad Ghavamzadeh (University of Massachusetts), Elliott Ludvig (Rutgers University), Cosma Shalizi (University of Michigan), Achim Rettinger (University of Koblenz), John Langford (University of Chicago), and Sebastian Thrun (Stanford University).

The following table lists some of the collaborative activities of RLAI team members within the research community.

Collaborative activities in the research community		
Name	conference/journal/activity	role
Schuurmans	International Conference on Machine Learning	Program co-chair
Sutton	NIPS workshop on "Reinforcement Learning: Benchmarks and Bake-offs"	Co-organizer
Sutton	ICML workshop on "Predictive Representations of World Knowledge"	Co-organizer
Sutton	International Conference on Machine Learning	Area chair
Sutton	International Joint Conference on Autonomous Agents and Multi-Agent Systems	Area chair
Sutton	Neural Information Processing Systems Conference	Program committee
Sutton	Conference on Uncertainty in Artificial Intelligence	Program committee
Sutton	Int'l Joint Conference on Artificial Intelligence	Program committee

Sutton	IJCAI 2005 Workshop on "Planning and Learning in A Priori Unknown or Dynamic Domains"	Program committee
Sutton	AAAI Fall Symposium on "Real Life Reinforcement Learning"	Organizing committee
Schuurmans	Machine Learning Journal	Associate editor
Schuurmans	Journal of Machine Learning Research	Associate editor
Schuurmans	International Conference on Machine Learning	Area chair
Schuurmans	Conference on Uncertainty in Artificial Intelligence	Area chair
Bowling	Machine Learning Journal	Guest co-editor
Bowling	Int'l Conference on Intelligent Robots and Systems	Exhibition co-chair
Bowling	AAAI Fall Symposium on "Artificial Multi-Agent Learning"	Organizing committee
Bowling	AAAI workshop on "Multiagent Learning"	Organizing committee

Public outreach	
event	Participant
public lectures and demonstrations at the Odyssium's exhibit on robotics	Bowling, Sutton
mentoring of Edmonton high school students	Bowling
presentations at Edmonton elementary and high schools (4)	Bowling
public lecture at the Edmonton centennial "Philosopher's Cafe"	Sutton
participation in iCORE public debate on robot rights	Sutton
Press: Edmonton Journal, Folio, Vancouver Sun, A Channel TV	Bowling

GRADUATES

Student	Supervisor	Research topic	Currently
Lihong Li, MSc	Bulitko	Reinforcement learning, computer vision	PhD student, Rutgers University
Relu Petrascu	Schuurmans	Factored Markov Decision Processes	Postdoctoral Fellow at U. Toronto
Finnegan Southey	Schuurmans		Postdoctoral Fellow at U. Alberta
Ali Ghodsi	Schuurmans		Assistant professor, Univ. of Waterloo Statistics dept

INTELLECTUAL PROPERTY

None.

FINANCIAL REPORT

Financial Report
March 31, 2005

REVENUES

iCORE	600,000
Interest	546.60
(see other funding sources below)	
BALANCE at April 1, 2004	386,962.76
EXPENSES	
Faculty salaries and benefits	179,880.57
Research staff salaries and benefits	12,093.45
Post Docs and Graduate students	21,973.85
Support staff salaries and benefits	91,404.74
Communications, Outreach, Travel	34,314.73
Equipment	17,046.72
General Operations	16,277.34
TOTAL EXPENSES	372,991.59
BALANCE at March 31, 2005	614,517.77

FUNDING

Summary of funding from other sources:

Canada research chair to Schuurmans	100,000
NSERC discovery grant to Sutton	50,000
NSERC discovery grant to Schuurmans	47,000
NSERC discovery grant to Bowling	26,000
NSERC discovery + RTI grant to Bulitko	15,000 + 8,700
CFI to Schuurmans + matching from U. of Alberta	224,100
IRIS grant to Schuurmans	22,500
DARPA grant to Universities of Rutgers, Michigan, and Massachusetts Supports Mark Ring	~480,000/yr, RLAI portion 36,000
MITACS grant "Statistical Learning of Complex Data with Complex Distributions" to Schuurmans, Sutton, with Yoshua Bengio, Hugh Chipman, Christian Leger, William Welch, Jim Ramsay, and Mu Zhu	170,000/yr, RLAI portion ~25,000
AICML provided support for programming staff, software management, visitors, robot lab renovation, robot hardware and software support	~100,000
Scholarships: AIF, NSERC, iCORE, University, Killam, Steinhauer, Sampson, Ontario, FS Chia Reporting period only	~285,000
Advanced International Education and Research Support Program - Oita University Supports Katsunari Shibata	46,600
NSERC Strategic Grant, "Intelligent Agents for Interactive Entertainment" to Bowling, Schuurmans, with Jonathan Schaeffer, Robert Holte, Duane Szafron, and Michael Buro	195,000/yr, RLAI portion ~30,000

PUBLICATIONS

All the publications listed as appeared, accepted, or submitted are to journals or to highly selective conferences that are rightly regarded as equivalent to journals. Papers in these conferences are rigorously reviewed and of the highest quality and visibility. Contributions to less stringently reviewed conferences are included as other contributions. In all cases, the categories are mutually exclusive and correspond to the reporting period.

APPEARED ARCHIVAL PUBLICATIONS

B. Browling, J. Bruce, M. Bowling, and M. Veloso, "STP: Skills, Tactics and Plays for Multi-robot Control in Adversarial Environments," *IEEE Journal of Systems and Control Engineering*, pages 33-52, 2005.

S. Wang, D. Schuurmans, F. Peng, and Y. Zhao, "Learning Mixture Models with the Regularized Maximum Entropy Principle," *IEEE Transactions on Neural Networks 15* (4): 903-916, July 2004.

J. Huang, F. Peng, A. An, and D. Schuurmans, "Dynamic Web Log Session Identification with Statistical Language Models," *Journal of the American Society for Information Science and Technology*, 2004, Vol. 55, No. 14, Pages 1290 –1303.

L. Xu, J. Neufeld, B. Larson, and D. Schuurmans, "Maximum Margin Clustering," in *Advances in Neural Information Processing Systems 17* (NIPS), 2005, pages 1537 – 1544.

M. Bowling, "Convergence and No-regret in Multiagent Learning," in *Advances in Neural Information Processing Systems 17* (NIPS), 2005, pages 209 – 216.

R.S. Sutton and B. Tanner, "Temporal-difference Networks," in *Advances in Neural Information Processing Systems 17* (NIPS), 2005, pages 1377 – 1384.

I. Levner and V. Bulitko, "Machine Learning for Adaptive Image Interpretation," in *Proceedings of the National Conference on Artificial Intelligence*, 2004, pages 870 – 876.

ACCEPTED ARCHIVAL PUBLICATIONS

S. Wang, D. Schuurmans, F. Peng, and Y. Zhao, "Combining Statistical Language Models via the Latent Maximum Entropy Principle," to appear in *Machine Learning*.

P. Stone, R.S. Sutton, and G. Kuhlmann, "Reinforcement Learning for RoboCup-Soccer Keepaway," accepted for publication in *Adaptive Behavior*.

M. Bowling, A. Ghodsi, and D. Wilkinson, "Action Respecting Embedding," to appear in the *Proceedings of the 22nd International Conference on Machine Learning* (ICML), Bonn, Germany, August 7 – 11, 2005.

T. Wang, D. Lizotte, M. Bowling, and D.Schuurmans, "Bayesian Sparse Sampling for On-Line Reward Optimization," to appear in the *Proceedings of the 22nd International Conference on Machine Learning* (ICML), Bonn, Germany, August 7 – 11, 2005.

B. Tanner and R.S. Sutton, "Temporal-Difference Networks with Eligibility Traces," to appear in the *Proceedings of the 22nd International Conference on Machine Learning* (ICML), Bonn, Germany, August 7 – 11, 2005.

L. Cheng, F. Jiao, and D. Schuurmans, "Variational Bayesian Image Modeling," to appear in the *Proceedings of the 22nd International Conference on Machine Learning* (ICML), Bonn, Germany, August 7 – 11, 2005.

L. Xu, and D. Schuurmans, "Unsupervised and Semi-supervised Multi-class Support Vector Machines," to appear in the *Proceedings of the Twentieth National Conference on Artificial Intelligence* (AAAI), Pittsburgh, Pennsylvania, July 9 – 13, 2005.

M. Bowling and P. McCracken, "Coordination and Adaptation in Impromptu Teams," to appear in the *Proceedings of the Twentieth National Conference on Artificial Intelligence* (AAAI), Pittsburgh, Pennsylvania, July 9 – 13, 2005.

B. Tanner and R.S. Sutton, "Temporal-Difference Networks with History," to appear in the *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence* (IJCAI), Edinburgh, Scotland, July 30^{th} – August 5^{th}, 2005.

D. Wilkinson, M. Bowling, and A. Ghodsi, "Learning Subjective Representations for Planning," to appear in the *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence* (IJCAI), Edinburgh, Scotland, July 30th – August 5th, 2005.

C. Boutilier, R. Patrascu, P. Poupart, and D. Schuurmans, "Regret-based Utility Elicitation in Constraint-based Decision Problems," to appear in the *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence* (IJCAI), Edinburgh, Scotland, July 30^{th} – August 5^{th}, 2005.

D. Schuurmans, Y. Guo, and R. Greiner, "Learning Coordination Classifiers," to appear in the *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence* (IJCAI), Edinburgh, Scotland, July 30^{th} – August 5^{th}, 2005.

E.J. Rafols, M.B. Ring, R.S. Sutton, B. Tanner, "Using Predictive Representations to Improve Generalization in Reinforcement Learning," to appear in the *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence* (IJCAI), Edinburgh, Scotland, July 30^{th} – August 5^{th}, 2005.

A. Ghodsi, J. Huang, F. Southey, D. Schuurmans, "Tangent-Corrected Embedding," to appear in the *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition* (CVPR), June 20 – 25, 2005.

SUBMITTED ARCHIVAL PUBLICATIONS

D. Schuurmans, "Constraint-based Optimization and Utility Elicitation using the Minimax Decision Criterion," submitted to *Decision Analysis Journal.*

B. Price and J. Roberts, "Exploration Envelopes: A new approach to Transfer from Simulator to the Real-World," submitted to the 22nd International Conference on Machine Learning (ICML), 2005.

F. Southey, M. Bowling, B. Larson, C. Piccione, N. Burch, and D. Billings, "Bayes' Bluff: Opponent Modeling in Poker," submitted to the Conference on Uncertainty in Artificial Intelligence (UAI), 2005.

M. Bowling, D. Wilkinson, A. Ghodsi, and A. Milstein, "Subjective Localization with Action Respecting Embedding," submitted to Robotics: Science and Systems, 2005.

Q. Wang, D. Lin, and D. Schuurmans, "Improved Estimation for Unsupervised Part-of-Speech Tagging," submitted to the Conference of the Association for Computational Linguistics (ACL), 2005.

S. Wang, R. Greiner, D. Schuurmans, and S. Wang, "Consistency and Generalization Bounds for Maximum Entropy Density Estimation," submitted to the Conference on Computational Learning Theory (COLT), 2005.

Y. Guo, D. Wilkinson, and D. Schuurmans, "Maximum Margin Bayesian Networks," submitted to the Conference on Uncertainty in Artificial Intelligence (UAI), 2005.

OTHER CONTRIBUTIONS

A. Ghodsi, J. Huang, and D. Schuurmans, "Transformation-invariant embedding," in *Proceedings of the 8th European Conference on Computer Vision* (ECCV), pages 520 – 527, May 11 – 14, 2004, Prague, Czech Republic.

K. Shibata, "Discretization of Series of Communication Signals in Noisy Environment by Reinforcement Learning," in *Proceedings of the Seventh International Conference in Adaptive and Natural Computing Algorithms* (ICANNGA), 2005, pages 486-489, held in Coimbra, Portugal, Mar. 23, 2005

I. Levner, V. Bulitko and G. Lin, "Feature Extraction for Classification of Proteomic Mass Spectra: A Comparative Study," in *Feature Extraction, Foundations and Applications*. Springer Verlag, to appear.

G. Lee and V. Bulitko, "GAMM: Genetic Algorithms with Meta-Models for Vision," in *Proceedings of Genetic and Evolutionary Computation Conference* (GECCO). Washington, D.C.

G. Lee, V. Bulitko and I. Levner, "Automated Selection of Vision Operator Libraries with Evolutionary Algorithms," in *Proceedings of the Congress on Evolutionary Computing* (CEC), pages 1127 – 1134. Portland, Oregon.

S. Ishii, M. Shidara, and K. Shibata, "An Explanation of Emergence of Reward Expectancy Neurons Using Reinforcement Learning and Neural Net," to appear in the *Proceedings of the Fourteenth Annual Computational Neuroscience Meeting* (CNS), 2005.

D. Silver, "Cooperative Pathfinding," to appear in the *Proceedings of the 1st Conference on Artificial Intelligence and Interactive Digital Entertainment* (AIIDE), 2005.

I. Levner and V. Bulitko, "Comparison of Machine Learned Image Interpretation Systems in the Domain of Forestry," in *Proceedings of the IEEE Workshop on Applications of Computer Vision* (WACV), 2005, pages 421 – 426. Breckenridge, Colorado.

P. McCracken, and M. Bowling, "Safe Strategies for Agent Modeling in Games," AAAI 2004 Fall Symposium on Artificial Multiagent Learning, pages 103-110.

L. Li, V. Bulitko, and R. Greiner, "Focus of Attention in Sequential Decision Making," National Conference on Artificial Intelligence (AAAI) Workshop on Learning and Planning in Markov Processes – Advances and Challenges, Poster Session, San Jose, California, July 26, 2004.

L. Li, V. Bulitko, and R. Greiner, "Batch Reinforcement Learning with State Importance," in *Proceedings of the European Conference on Machine Learning, Poster Section,* Pisa, Italy, September 20 – 24, 2004, pages 566 – 568.

V. Bulitko, "Learning for Adaptive Real-time Search," Computer Science Research Repository (CoRR), 2004.

R.S. Sutton, "Knowledge Representation in TD Networks," invited talk at the workshop on *Learning and Planning in Markov Processes – Advances and Challenges* at AAAI 2004, the annual conference of the American Association for Artificial Intelligence (San Jose, CA).