

Significant BITS

Newsletter of the
Department of Computer Science



Erik Learned-Miller

Learning on the Fly

Imagine a young girl on her first Easter egg hunt. As she starts looking around, she may have no specific ideas about where to find an egg—she’s just looking for brightly colored objects of a certain size and shape. When she spies her first egg behind a tree, there’s an “Aha!” moment. She realizes that looking behind other trees might be a good strategy. With each success, she refines her search strategy, getting faster and faster at finding the remaining eggs. “This type of rapid adaptation,” says Associate Professor Erik Learned-Miller, “is ubiquitous in human behavior. In any type of new environment, people adapt at an astounding rate, sometimes in a fraction of a second.”

Learned-Miller, whose work focuses on computer vision and machine learning, calls this type of adaptation “Learning on the Fly.” “The computer vision community,” he notes, “has largely embraced the idea that making vision algorithms that can learn from examples is essential to building general and robust systems. However,” he continues, “most of the learning methods used today do not exhibit the type of rapid adaptation one sees in people.” To help close the gap between the fantastic vision capabilities of people and the current capabilities of machines, Learned-Miller believes it will be essential to model this type of rapid adaptation.

One area in which Learned-Miller is applying these ideas is in the problem of face detection. Face detection is simply the problem of locating all of the faces (if any) that appear in an image. It is widely used in applications from digital cameras to FaceBook photo tagging. While for some images this is relatively easy, other images, such as the one on page 5 showing four hikers, can make this problem challenging. Issues like low resolution or strong shadows can cause problems for traditional face detection algorithms. In particular, the leading traditional face detection method, applied to the hiker image shown, successfully finds two of the faces, but misses the other two. Since the traditional detector treats each part of the image independently, it cannot adapt its strategy for detection of more faces given the faces it has already found.

continued on page 5

News

New computational social science initiative	4
Helping police nab Internet child pornographers	13

Awards

UMass President honors Adrion Osterweil receives Chancellor’s Award	6 14
---	---------

Alums

Alum Focus: Eric Brown	8
Recent Ph.D. Grads	10

Hanna Wallach joins CS faculty

Hanna Wallach began this fall as a tenure-track Assistant Professor in the department as part of UMass Amherst’s interdisciplinary research initiative in computational social science (see related article). Within this new initiative, she is collaborating with new faculty in the departments of Sociology, Political Science, and Mathematics & Statistics.

“Computational social science is an emerging research area at the intersection

continued on page 4





Writing this shortly after the Thanksgiving break, I was reminded of my first Thanksgiving break as a new undergraduate: it was like coming up for air; another day of classes and I might have drowned. Returning to my family—which I was so eager to leave three months before—was a most welcome intermission. Not too much has changed since then:

the family has changed (now the kids come to me), but this Thanksgiving break was a chance to unwind a bit from the intensity of the Fall term, to reflect on the difficult times that the campus and the department, indeed people everywhere, have been living through due to the financial downturn. This reflection led me to appreciate the extraordinary success our department has had even in these difficult times.

Because campus leaders recognize the importance of continuing to hire professors despite budget reductions, the department's faculty has been able to grow. The process was somewhat more centrally orchestrated than usual, driven by the administration's desire to strategically invest in clusters of faculty members from multiple disciplines who, by working together, can move us into exciting new areas of research and teaching. We are absolutely delighted that Hanna Wallach joined us this September. She teams up with other new professors in the departments of Political Science, Sociology, and Mathematics and Statistics to catalyze a campus focus on Computational Social Science. You can read more about Hanna and the Computational Social Science initiative elsewhere in this newsletter.

We are also delighted that Aria Haghighi will join us next Fall. Aria specializes in statistical natural language processing. He will join new professors in the departments of Linguistics

and Psychology, as well as many current faculty members, to focus on natural language (i.e., not computer language) from the combined perspectives of these three disciplines. You will be able to read more about Aria in this newsletter's next issue.

In addition, we are currently conducting searches to fill a faculty position in machine learning and a Five College Joint Faculty Position shared with Mount Holyoke College in Computational Biology/Bioinformatics. We hope to welcome new professors in these areas next Fall. All of these new faculty positions significantly contribute to the department's efforts to add additional strength to our collaborative relationships with other departments and other disciplines.

Our department continues to work toward increasing the diversity of its student body. I am pleased to report that we recently announced that we will award a \$1,000 scholarship to any high-school woman from across the nation who wins the NCWIT Award for Aspirations in Computing and is accepted and enrolls in our major at UMass Amherst. NCWIT, the National Center for Women & Information Technology, is a coalition of over 200 prominent corporations, academic institutions, government agencies, and non-profits working to increase women's participation in information technology. The more women who qualify, the better—we will find the funds somewhere!

I could list more that is worthy of celebration: for example, that the number of our undergraduate majors has steadily increased since 2007, or that our grant awards are up seven percent from last year to a record of \$17M, comprising nine and one half percent of the award dollars of the entire campus. But instead I invite you to see in more detail what we have been doing by reading the rest of this issue.

I am sure that the next holiday break will be another opportunity to come up for air and to celebrate more accomplishments. The energy and creativity of the department's faculty, students, and staff—undiminished in these difficult times—never ceases to amaze me.

High Performance Computing Center breaks ground

State and University officials gathered in Holyoke, MA on October 5th for the groundbreaking ceremony of the \$90 million Massachusetts Green High Performance Computing Center. The Center will be located in the Holyoke canal area to take advantage of the green and cost-effective hydroelectric energy and access to high-speed data transmission lines

At the groundbreaking ceremony, UMass President Jack Wilson celebrated “the largest and most significant collaboration of government and private industry and public and private universities in the history of the Commonwealth.” Governor Deval L. Patrick said later that the computing center is an example of the state working with higher education and industry to use innovation to help the region, noting “I think it's a real opportunity to serve as a magnet on a whole host of levels.”

The project will create a world-class, green, high performance computing center that will provide an infrastructure for research computing and a collaborative research agenda in advanced computing and applications such as life sciences, clean energy, and green computing. It will act as a hub that will facilitate collaboration in R&D and education among our higher education institutions and will serve as a catalyst for technology-based economic development.

The five university partners in this collaboration include the University of Massachusetts, Massachusetts Institute of Technology, Boston University, Harvard University and Northeastern University, who are working together with EMC, Cisco, Holyoke Community College, Springfield Technical Community College, the city of Holyoke, and others on this project.

iPhone™ app developed to rescue oiled Gulf Coast wildlife



Assistant Professor Deepak Ganesan is among the UMass Amherst researchers who launched an iPhone™ app in June to help in the Gulf oil spill crisis. Any iPhone™ user who comes upon oiled birds and other wildlife in the Gulf Coast region can immediately transmit the location and a photo to animal rescue networks using a free iPhone™ app, MoGO, for Mobile Gulf Observatory.

In addition to connecting users with the Wildlife Hotline so wildlife experts can find and rescue the oiled and injured animals, photos of oiled wildlife plus the GPS location are also uploaded to MoGO's comprehensive database for review by wildlife and fisheries experts using a Web browser.

Ganesan and his UMass Amherst colleagues, Andy Danylchuk, Curt Griffin, and Charlie Schweik, developed the app to enable 'citizen scientists' with the tools to help save wildlife along the 14,000 miles of northern Gulf coastline. As a result of the largest oil spill in U.S. history, over 400 wildlife species and 35 national wildlife refuges were at risk.

The app takes advantage of "mobile crowdsourcing," that is, the power of smart personal mobile devices to provide thousands of eyes and ears on the ground. Ganesan's research group has designed a software framework called "mCrowd," which simplifies the usual weeks- to months-long process of developing a new mobile crowdsourcing app. "It provides easy-to-use templates that can be tailored to a new application," Ganesan explains. His mCrowd technology allowed the UMass Amherst team to create the MoGO app and infrastructure in a little more than a week. Details on the project can be found at www.savegulfwildlife.org.

In a similar project, Ganesan and graduate student Mohamed Musthag worked with U.S. Fish and Wildlife Service biologist Verena Gill this fall to develop a free iPhone™ application that can be used to report beached marine mammals in Alaska.

International student programming competition finalists

A group of CS undergrads was named as one of twelve teams chosen from around the world as finalists in the 2010 IEEE Computer Society Student Competition. With over 80 teams registered worldwide for the competition, the UMass Amherst CS team was one of only three U.S. teams selected as finalists.

The department's team consisted of current CS undergraduates Ryan Hurley and Caleb Raitto, and CS alum Stevie Sellers (B.S. 2010). The team members, all undergrads in Commonwealth College during the spring semester, were classmates in the course "Inside the Box: How Computers Really Work" (Compsci 391B) taught by Associate Professor Chip Weems. The students joined together to complete a class project that was submitted for the IEEE competition. They designed a computer architecture from scratch, including the entire instruction set, and then implemented a software simulation of it using high quality software engineering techniques. For the competition, the team also submitted a report that explained their use of software engineering.

"The team decided to create a novel low-power design that packs two instructions in each word, with some clever use of the registers," says Weems. "Their Java-based simulator included an assembler, disassembler, and complete graphical user interface that allows users to observe the activity of the processor. In addition, they created a gate-level simulation of the hardware, using Logisim, that ran in parallel, providing a real-time display of dynamic power consumption. The team also wrote benchmarks to demonstrate its operation."

According to the IEEE Computer Society, "the purpose of the competition is to promote excellence in the design of a system by a team of students. The teams designed a simulator able to run on a typical PC using a Windows-based operating system. The system will be judged on the originality of its architecture, the simulator's functionality, quality, and versatility, and the use of software engineering in the simulator's design. Credit will be awarded for construction of the instruction set and the tradeoff between complexity and elegance."

"This team did phenomenally well to be named one of the twelve best in the world given that the team had the minimum number of participants (three students versus the max of five) and that two of the teammates were completing senior year Commonwealth College requirements while working on this project," adds Weems.

The UMass Amherst CS team joined the other finalists from universities in Colombia, Cuba, Egypt, Jordan, Peru, Russia, Thailand, and the United States. The winning teams were from Egypt, S. Korea, Colombia, and Russia.

With nearly 85,000 members, the IEEE Computer Society is the world's leading organization of computing professionals. Founded in 1946, and the largest of the 39 societies of the Institute of Electrical and Electronics Engineers (IEEE), the Computer Society is dedicated to advancing the theory and application of computer and information-processing technology.

WALLACH - - - - - continued from page 1

of computer science, statistics, and the social sciences, driven by new sources of data from the Internet, government databases, voting records, universities, and more,” says Wallach. “I develop new mathematical models and computational tools for analyzing vast quantities of structured and unstructured data in order to identify and answer social science questions regarding science and innovation policy, political science, government transparency, and free/open source software development.”

Wallach adds, “Collaboration is imperative to achieving my research goals: social scientists are uniquely positioned to identify the most pertinent and vital questions and problems, as well as to provide insight into data generation and acquisition, while computer scientists, such as myself, can contribute significant expertise in developing novel, quantitative methods and tools. Thanks to UMass Amherst’s interdisciplinary research, my colleagues and I are well-placed to make truly groundbreaking advances in computational social science.”

Wallach works on making reliable inferences about the semantic and social dimensions of communication and collaboration in the face of uncertain and incomplete information. Although her central methodological framework is situated in machine learning and Bayesian statistics, Wallach’s research touches upon important problems in many other areas of computer science: processing and storing massive quantities of data (algorithms and data structures); digitizing documents (computer vision); transcribing audio and video streams (speech processing and computer vision); representing and analyzing text (natural language processing); building powerful and usable software tools (software engineering, human-computer interaction, and information visualization).

Prior to joining the department’s faculty, she was a Senior Postdoctoral Research Associate in the department’s Information Extraction and Synthesis Laboratory. Wallach received a B.A. in Computer Science and a Ph.D. in Physics from the University of Cambridge, in 2001 and 2008, respectively. She received an M.S. in Informatics from the University of Edinburgh in 2002. She was awarded the University of Edinburgh’s 2001/2002 prize for Best M.Sc. Student in Cognitive Science, and her undergraduate project, “Visual Representation of Computer-Aided Design Constraints,” won the award for the best computer science student in the 2001 U.K. Science Engineering and Technology Awards.

Wallach, along with co-authors Ryan Adams and Zoubin Ghahramani, received the Best Paper Award at the 2010 Thirteenth International Conference on Artificial Intelligence and Statistics for their paper “Learning the Structure of Deep, Sparse Graphical Models.”

In addition to her research, Wallach works to promote and support women’s involvement in computing. In 2006, she co-founded an annual workshop for women in machine learning, in order to give female faculty, research scientists, postdoctoral researchers, and graduate students an opportunity to meet, exchange research ideas, and build mentoring and networking relationships.

“I’m truly delighted and honored to be joining the UMass Amherst faculty,” says Wallach. “UMass Amherst has demonstrated clear commitment to providing one of the best possible institutional contexts for my research. Additionally, having worked in the department for three years, I’m thrilled to be establishing my faculty career in a department that I know firsthand to be academically excellent, and also the friendliest department I’ve ever worked in. I’m very excited about continuing existing collaborations and establishing new collaborations with the faculty and students here at UMass.”

New multidisciplinary initiative in computational social science

Through the campus’s newly created multidisciplinary initiative in computational social science (CSS), four faculty were hired this fall at UMass Amherst to focus on this emerging and rapidly expanding field of research.

Hanna Wallach joined the Department of Computer Science (see accompanying article), Ryan Acton joined the Department of Sociology, Krista Gile joined the Department of Mathematics & Statistics, and Bruce Desmarais joined the Department of Political Science, all as Assistant Professors in the CSS initiative which provides a mechanism for interdisciplinary collaboration within these four departments at UMass Amherst.

“Computational social science is particularly important now, as the nation strives to reduce energy use, improve health care, revive the economy, and strengthen our educational system,” says CS Professor Andrew McCallum, Director of the new Computational Social Science Initiative (cssi.umass.edu). “Addressing each of these challenges will require significant improvements in our understanding of the interactions of people, institutions, markets, and the other components that make up these complex systems.”

With the establishment of this world-class cluster of excellence, the campus is poised to capture a leading position in the field of CSS, expand the reach and impact of computer science, and provide the ability to understand complex systems and social interactions using state-of-the-art methods involving novel ideas from statistical text processing and social network analysis.

With this new initiative, interdisciplinary research and teaching collaboration will span the campus, connecting departments that have not historically collaborated, but have a strong desire to work together now, adds McCallum.

LEARNING ON THE FLY – continued from page 1

Working with graduate student Vidit Jain, Learned-Miller has been looking for ways to adapt algorithms not just to specific types of environments, but to an individual photograph. “We wanted to find a way to tell the computer, ‘Hey, you found one face with a strong shadow in this image. You should raise your expectations for finding other faces with strong shadows.’ We wanted the algorithm to learn on the fly about the particular image it was working with.” Of course, the devil is in the details, but Jain and Learned-Miller eventually found a way to reprocess an image, using information gleaned from the first round of detections on that image. “We found a way to learn about what to expect in the image from initial detections, and we achieved a large accuracy increase over traditional methods.” For example, Jain and Learned-Miller’s algorithm finds three of the four faces in the hiker image without marking any non-face regions as faces. They were able to increase the sensitivity of detection to types of faces that were similar to faces they had already found, and lower the sensitivity to types of faces deemed unlikely by analysis of the initial detections. This algorithm currently outperforms all other face detection algorithms that have been tested on the UMass Amherst face detection database.

Another area in which Learned-Miller is applying the concept of learning on the fly is in the area of optical character recognition, or OCR. “While many researchers will try to tell you that OCR is a ‘solved problem,’ real users will tell you their frustrations with the errors that are still made by such systems on difficult-to-read documents,” says Learned-Miller. Like face detection algorithms, most modern OCR software suffers from the fact that it interprets the shape of each character without considering the shapes of other characters. In documents that have been heavily corrupted by stray marks, photocopying degradation, or poor print quality, there are often characters that simply cannot be classified correctly using the pre-trained models. “Our insight,” says Learned-Miller, “was to realize that characters that have already been recognized correctly in the same document could be used as more powerful models of the difficult-to-recognize characters. In other words, the easy-to-recognize characters in a document, recognized in a first-pass OCR system, can be used to build a document-specific model of the poorly formed characters in the same document. These new, document-specific character models could then be used to recognize the more difficult characters in a second pass of the OCR system.” This is another example of learning on the fly.

Working with graduate students Andrew Kae and Gary Huang, Learned-Miller developed a method for adapting character models to a specific document. The key difficulty was in deciding which of the characters that were recognized in the first pass were most likely to be correct. On a set of difficult documents, the team was able to reduce the average error rate by more than 30% over Google’s widely used Tesseract OCR system. “This was a great validation of the general idea of learning on the fly,” says Learned-Miller. “While it is not always obvious how to do it at first glance, learning on the fly



is proving to be a valuable tool in reducing the performance gap between human and machine recognition performance, and we’re excited to find new areas in which to apply these principles.”

Learned-Miller joined the faculty of the Department of Computer Science UMass Amherst in 2004. Prior to joining the department, he spent two years as a post-doctoral researcher at the University of California, Berkeley, in the Computer Science Division. In 1989, he co-founded CORI-Techs, Inc., where he and co-founder Rob Riker developed the second FDA cleared system for image-guided neurosurgery. Learned-Miller obtained a B.A. in Psychology from Yale University in 1988 and an M.S. and Ph.D. in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology in 1997 and 2002, respectively. In 2006, he received an NSF CAREER award for his work in computer vision and machine learning.

We need your support

Your gifts are invaluable in helping the department fulfill its goals of excellence in research and teaching. These gifts are important in augmenting our regular programs by promoting undergraduate and graduate research, supporting seminars by outstanding scientists, and helping new faculty establish their research programs. Those contributions that are designated for specific programs also fund activities that enrich our educational and research programs.

One of our goals is to enhance the undergraduate experience to more accurately reflect the excitement that the field of computer science offers. To meet this goal, we must provide awards and scholarships for our students, new equipment for our labs, improved classrooms, and more seminar series. In addition, gifts to our CS Endowment fund will have continuing benefits to the department’s graduate and undergraduate programs.

Adrion receives UMass President's Award

On November 29, Professor Rick Adrion received the UMass President's Award for Public Service from Jack M. Wilson in recognition of his leadership in developing, broadening, expanding, and improving computing and information technology activities for students in grades K-12 and at colleges and universities across the Commonwealth.

Adrion has organized many information technology workshops over the past 10 years and expanded the Commonwealth Information Technology Initiative (CITI). He has played a key role in expanding CITI from a college-level-only partnership to a program that touches the lives of hundreds of students in grades K through 12, and at community colleges and universities.

Prof. Adrion also leads the Commonwealth Alliance for Information Technology Education (CAITE), a \$4 million National Science Foundation-sponsored project that is active at nine community colleges, two state universities, and four UMass campuses. It focuses on bringing more women and minorities into IT education and the work force. Since it began in 2007, CAITE has involved approximately 15,000 Massachusetts students and educators at all levels. Activities

range from professional development sessions for teachers to more than 130 outreach events, exciting activities and fun competitions for students from middle school through college.

"This is a very well-deserved honor. Throughout his career at UMass Amherst, Rick has worked tirelessly to expand and strengthen computing and information technology opportunities across the Commonwealth and beyond," says Department Chair Andrew Barto. "Rick cares deeply about expanding opportunities for study and work in these fields, and through his role as director of the CAITE and co-director of CITI, he has been able to make crucial contributions."

"Through these awards, we celebrate the exemplary work of these faculty members who truly embody the University's ethos of academic excellence and service," says President Jack M. Wilson. Adrion is one of five Massachusetts educators who received the president's award.



Balasubramanian, Hay, and Zhu receive CIFellowships

Aruna Balasubramanian, Michael Hay, and Ting Zhu have been named 2010 Computing Innovation Fellows (CIFellows). Balasubramanian and Hay are recent UMass Amherst CS Ph.D. graduates, and Zhu joined the department this fall.

Balasubramanian will begin her fellowship work at the University of Washington under the mentorship of Professor David Wetherall. She will be working on making mobile computing truly ubiquitous by building protocols that sup-

port seamless connectivity as users move between different networks and by enabling energy efficient computing to allow users to run critical applications on their mobile devices.

Hay joined Cornell University this fall for his fellowship under the mentorship of Professor Johannes Gehrke. He will be doing research on designing practical tools for privacy-preserving analysis of complex data.

Zhu received his Ph.D. in Computer Science and Engineering from the University of Minnesota. He joined the department as a CIFellow under the mentorship of Distinguished University Professor Don Towsley. Zhu's research interests are in wireless networks, communications, embedded systems, real-time systems, and security.

The competitive CIFellow program is sponsored by The Computing Community Consortium (CCC) and the Computing Research Association (CRA), with funding from the National Science Foundation. The program allows new Ph.D. graduates to obtain one-to-two year postdoctoral positions at host organizations including universities and industrial research laboratories that advance the field of computing and its positive impact on society. The goals of the CIFellows project are to retain new Ph.D.s in research and teaching and to support intellectual renewal and diversity in the computing fields at U.S. organizations.

In the 2009 competition, CS alums Jeffrey Johns and Victoria Manfredi received CIFellowships along with Leena Razzaq who joined the department as a CIFellow.

SAVE THE DATE: Alum Banquet 2011

The third Outstanding Achievement and Advocacy (OAA) Awards Banquet will be held on campus on the evening of Friday, May 6, 2011. The OAA awards program recognizes the achievement of our alums in such areas as entrepreneurship, scientific research, and education. Please join us to celebrate the accomplishments of our award recipients and to socialize with faculty and fellow alums. Details will be posted at www.cs.umass.edu/oaa2011.

Helping to lay 'clean slate' foundation for the future Internet

Because today's Internet was designed for "tethered" computers such as desktop models, its behind-the-scenes operations are often fragile, inefficient, and difficult to manage for connecting via smart phones, laptops, or sensors. Plus, with an estimated 4 billion wireless, mobile devices in use worldwide, more than the total of tethered devices and still rising, it's important to start addressing the problem, says Assistant Professor Arun Venkataramani.

He is part of a three-year Future Internet Architecture collaboration launched recently by the National Science Foundation (NSF) and known as "MobilityFirst: A Robust and Trustworthy-Centric Architecture for the Future Internet." Its goal is to develop a "clean-slate" candidate platform for future Internet design, optimized for mobile networking and communication. Distinguished University Professors Jim Kurose and Don Towsley are working with Venkataramani on the project, one of four recently funded by NSF. Other participating institutions are project leader Rutgers, Duke, Massachusetts Institute of Technology, UMass Lowell, Univ. of Michigan, Univ. of Nebraska-Lincoln, and the Univ. of North Carolina at Chapel Hill.

"As the name suggests, the compelling motivations for this effort are 'mobility' and 'trustworthiness,' two aspects in which today's Internet woefully falls short," says Venkataramani, the lead architect for MobilityFirst. "Mobility means the Internet should seamlessly support mobile devices like smart phones and laptops, the way most of us access the Internet now." Instead, today's Internet remains fixated on stationary computers.

A compelling aspect of the MobilityFirst architecture is that mobility and trust are synergistic and complementary goals, Venkataramani says. Thus, many of the underlying mechanisms used to enhance mobility also improve trust, and vice versa.

A good way to think of desirable changes in Internet mobility which can be provided by new architecture is our current postal system, he adds. It assumes that people stay in one place most of the time and, if you move, you must manually inform all your contacts of your new address. If you forgot to update someone, mail could get lost. If you travel frequently, you can't expect to receive your mail at all, or it is delayed by forwarding. "Wouldn't it be nice if there were a simple and seamless way of receiving mail no matter where you moved, for however short a time? That is analogous to one of the goals of the MobilityFirst architecture," says Venkataramani.

The other issue, trustworthiness, means that the Internet should be reasonably secure against malicious entities. Security can never be perfect, but today's Internet, designed with benign users in mind, is "far from acceptable for what has become a global communication backbone of immense importance," he adds. "Even a benign error by a network operator in some remote corner of the world can make most of the Internet unavailable for many hours. A coordinated, large-scale targeted attack on a company or a nation would be disastrous." Another primary goal of the MobilityFirst project is to make the Internet resilient to resist such an attack.

Further, Venkataramani says, "As things stand today, your Internet account can easily be hijacked or spoofed, allowing

a malicious entity to receive all of your communications, and to impersonate you. The MobilityFirst architecture will make hijacking and spoofing difficult by using a 'self-certifying' addressing scheme based on public-key cryptography. Although public-key cryptography is used in e-commerce applications today, these applications run on top of an insecure network with unverified addresses. MobilityFirst aims to fix these and other significant security weaknesses."

CIIR continues to shine in TREC competition



The Text REtrieval Conference (TREC), sponsored and run by the National Institute of Standards and Technology (NIST), has provided a framework for the evaluation of Information Retrieval and Processing technologies for the last 19 years. This year, the CIIR team took top honors in the TREC Web Track.

At TREC, top research labs, academic and industrial, come from all over the world to pit their systems, techniques, and theories against each other across a wide variety of tasks, such as web page retrieval, question answering, and other information processing tasks. This year, 75 organizations from 18 different countries participated in seven tracks.

In this year's evaluation, CIIR graduate student Michael Bendersky set out to model the quality of web pages using a set of simple and easily computable document features, enabling improved retrieval performance. Combining that document quality model with the CIIR-developed Indri search engine produced a set of experimental runs that were top performers in the TREC 2010 Web Track. Michael was aided by his advisor, Distinguished Professor Bruce Croft, and senior software engineer David Fisher. The team took first place in three of the four evaluation metrics, and third place on the fourth metric.

The CIIR has participated in evaluation at TREC and been involved in the organization of TREC since it was established in 1992 as part of the TIPSTER Text program. Professor James Allan has been on the TREC Program Committee for the last 10 years and alumnus and past faculty member Jamie Callan (Ph.D. '93) also served on that committee. Faculty and students within the CIIR have organized and participated in numerous evaluations, typically submitting top-ranking results across a wide range of tasks related to search. The technologies and theories applied and evaluated at TREC by CIIR students over the years have found their way into numerous publications and theses. TREC has given CIIR researchers the opportunity to validate their original ideas in a rigorous evaluation setting, with feedback from top members of the field. Michael Bendersky's successful work continues that tradition.

Eric Brown and *Watson* take on Jeopardy! challenge

Question Answering (QA) has been an active area of research for several decades. Instead of retrieving whole web pages in response to keyword queries, as is typical for web search engines, a QA system retrieves answers to questions. Eric Brown (Ph.D. '96) has been involved in QA since 1999, when he developed a custom search engine at the IBM T. J. Watson Research Center for one of the first QA research projects at IBM. Leveraging his experience as a student of Distinguished Professor Bruce Croft and a member of the Center for Intelligent Information Retrieval, Brown implemented a semantic search engine to support an approach called Predictive Annotation, where named entities are indexed according to their semantic type, then retrieved as candidate answers if they match the answer type detected in the question.

After conducting research in this area for several years with a relatively small team, IBM Research identified QA as an important technology of the future, with the potential for transforming how humans interact with computers. To highlight this technology and accelerate its advancement, IBM embarked on the Jeopardy! Grand Challenge, where the goal is to build a QA system that can play the popular television quiz show Jeopardy! and beat a human champion. In late 2006, Brown joined a dozen other researchers at the Watson Research Center and, under the lead of principal investigator David Ferrucci, took on this challenge in earnest.

At first glance, building a system that can play Jeopardy! may seem like a strange project for IBM, says Brown, who is currently an IBM Research Staff Member and Manager, Unstructured Information Management Systems. Closer inspection, however, reveals that playing Jeopardy! requires solving several hard problems that make it an ideal benchmark for advancing QA technology. First, Jeopardy! covers a huge variety of topics, including history, geography, science, sports, the arts, entertainment, current events, and more. Nearly any topic you can imagine is fair game. Second, the questions, or "clues" as they're called in Jeopardy!, are expressed using complex natural language, and may include tricky phrasing, puns, or even puzzles. Third, to answer a Jeopardy! clue the player must respond with a precise answer – not a ranked list of web pages, but rather the precise word or phrase that answers the clue. Moreover, the player must determine a confidence in their answer and decide whether or not to ring in and even attempt to answer the clue. The reason for this is simple; if the player answers correctly, the value of the clue is added to the player's score, but if the player answers incorrectly, the value of the clue is subtracted from the player's score. And to

top it all off, speed is essential. A player typically has just three or four seconds to come up with his/her answer and decide if he/she is confident enough to ring in and attempt the clue.

"When the idea of building a computer system to play Jeopardy! was first proposed, many people thought it was impossible," says Brown. "But we knew that if we were successful, we would accomplish something truly amazing." To tackle this challenge, the IBM team had to revisit the state-of-the-art in QA and develop new approaches to solving the problem. The core architecture that emerged, called DeepQA, is built on a fundamental philosophy of combining a large number of dif-



ferent natural language processing analytics that operate on existing unstructured information (e.g., documents) to generate candidate answers and evaluate evidence. Underpinning this approach is a probabilistic machine learning framework that has enabled the team (now nearly two dozen researchers) to independently develop a large number of analytics for analyzing a wide variety of information and evidence sources, and then combine those analytics and their results into a single, integrated system. The application of DeepQA that actually plays Jeopardy! is a system called *Watson* (in honor of IBM's founder Thomas J. Watson, see www.ibm.com/watson for more).

Brown's role on the project has spanned a number of areas, including responsibility for the initial architecture and data model, specifying and acquiring hardware for the team's development environment, leading the systems architecture and engineering team, leading several "task force" teams to address specific classes of questions, and coordinating many of the team's academic collaborations and relationships. In particular, the IBM team has collaborated with Professor James Allan at UMass Amherst to leverage Indri as one of the search engines used in DeepQA.

Addressing all of the aspects that make Jeopardy! so hard (and so compelling) will have clear and direct implications for business applications of the technology. "We view this technology as going well beyond QA and into the realm of information systems that support intelligent decision making," says Brown. There are a number of application areas that could benefit from this technology, including medical diagnosis, business intelligence, customer support call centers, compliance, legal research, etc. "We've made tremendous progress in a few short years to build a competitive Jeopardy! system. Effectively adapting and applying this technology to even just a few of the many relevant business applications should keep us busy for many more years."

David Miller, MasterChef finalist



We had a chance to talk with Dave Miller (B.S. '03, M.S. '06) after he returned from his successful stint on the first season of Fox TV's MasterChef competition. Chosen from thousands who auditioned, Miller was one of the top 14 contestants on the show and was named as the runner-up in the final competition.

"Aside from getting married in Las Vegas with a bridal party donned in tuxedo t-shirts, the MasterChef experience has been the greatest of my life. Yes, 'I'm a software engineer – it's a

fantastic career and I love it, but to toot my own horn: I'm a hell of a chef, too," says Miller. "To have the opportunity to take your biggest passion to the world stage for all to admire and poke fun at with no consequences to anything other than pride, well, at least it was my dream come true. In no small part, my extended stay at UMass Amherst (truthfully both in and out of the Computer Science building) shaped me into the questionably-entertaining, overconfident, eyebrow raising gourmand you know today."

Miller adds, "Now, not every software engineer wants to try out for MasterChef or be on TV at all for that matter, but it does hearken the age-old question: 'What is it, you say, you do here?' Here's something from MasterChef that I wish had made it past the cutting room floor:

Gordon Ramsay (MasterChef host): 'David Miller, what is it again that you do for a living?'

Miller: 'I'm a software engineer.'

Ramsay: 'Really? I could have sworn you were a professional face-maker.'

Miller works at Vistaprint in Lexington, MA: a predominately .NET shop, where he develops internal CRM applications to support their three international customer care centers. He has a flexible schedule, so he can be home to work in the kitchen and make dinner before his wife arrives home. "I'm sure if I worked in a restaurant all day, I'd come home and write iPhone apps for fun," says Miller. "Fortunately, things are the other way around - I can pick up a Chateaubriand and some chanterelles after work, but I don't code for fun anymore. Does that mean I've given up my passion? In all honesty, I hope not. Some people cook to live, but I live to cook. If I did it for a living, I don't know if I'd love it like I do now."

Miller adds, "Often people approach me skeptical that someone with such a technical mind can do what I do in the kitchen, and, to the surprise of many, I tell them that the two really aren't that different. During my graduate research with LASER in process modeling and simulation, I adopted the

This fall, **Jody Daniels**, CS alum (Ph.D. '97) and 2010 OAA award recipient, was nominated for a promotion to the rank of brigadier general. According to U.S. DoD news release, Daniels will be assigned "as commander, (troop program unit), U.S. Army Reserve Support Command, First Army/deputy commanding general, First Army (East), Fort Meade, Md." Daniels is also the Director of Advanced Programs at Lockheed Martin.

At the ACM/IEEE 32nd International Conference on Software Engineering (ICSE 2010), **Jay Corbett** (Ph.D. '92), **Matthew Dwyer** (Ph.D. '95), and co-authors received the ISCE Most Influential Paper Award for their paper "Bandera: Extracting Finite-State Models from Java Source Code." The award is given to the author(s) of the paper from the ISCE conference 10 years ago that is judged to have had the most influence on the theory or practice of software engineering during the decade since its original publication. Corbett is currently a Senior Software Engineer at Google and Dwyer is currently the Henson Professor of Software Engineering in the Department of Computer Science and Engineering at the University of Nebraska.

The Massachusetts Innovation & Technology Exchange (MITX), New England's premier association for Internet business and marketing, named **Steve Vinter** (Ph.D. '85) as one of the five new members to its Board of Directors. Vinter is the Director of Google's Cambridge, MA facility.

Intronis Online Backup and Recovery appointed **Jay Bolgatz** (B.S. '85) as its Vice President of Engineering and Delivery.

Massachusetts-based GenomeQuest, Inc. announced this fall that **Richard Resnick** (B.S. '94) was appointed acting CEO of the company.

Dr. **Tom Wagner** (Ph.D. '00) was recently appointed to the position of Senior Vice President and Chief Technological Officer at iRobot.

concept introduced to me by Professor Lee Osterweil that (pardon my gross oversimplification) software is process and process is software; essentially any formalized process, too, is itself software - something that has stuck with me to this day. Ironically, Lee most frequently referenced recipes as a real-world example of a formal process definition. Combine the mind of a software engineer with that of a passionate chef, and you've got one extremely process-driven individual. Thus, I don't find it that surprising that concepts like resource allocation, scale, exception handling, performance, and optimization apply in the kitchen - I'd even argue that a little creativity and flair are necessary in software as well."

If there's anything that computer scientists in general can take from his adventures on MasterChef, Miller hopes it would inspire confidence in the fact that talented people really can excel in more than one talent - no matter how seemingly dissimilar they may be. "In your own lives, please don't fall subject to the pigeonhole principle - make more holes and fill them with something tasty," says Miller.

Recent Computer Science Ph.D. graduates (AY 2009-2010)



Nilanjan Banerjee: *Improved Network Consistency and Connectivity in Mobile and Sensor Systems*; (Mark Corner, Advisor); Sept. 2009; Assistant Professor, Department of Computer Science and Computer Engineering, University of Arkansas, Fayetteville.

Edge networks such as sensor, mobile, and disruption tolerant networks suffer from topological uncertainty and disconnections due to a myriad of factors including mobility and limited battery capacity on client devices. Hence, providing reliable, always-on consistency for network applications in such mobile and sensor systems is non-trivial and challenging. However, the problem is of paramount importance given the proliferation of mobile phones, PDAs, laptops, and music players. This thesis identifies two fundamental deterrents to addressing the above problem. First, limited energy on client mobile and sensor devices makes high levels of consistency and availability impossible. Second, unreliable support from the network infrastructure, such as coverage holes in WiFi degrades network performance. We address these two issues through client- and infrastructure-end modifications. The first part of this thesis proposes a novel energy management architecture called Hierarchical Power Management (HPM). HPM combines platforms with diverse energy needs and capabilities into a single integrated system to provide high levels of consistency and availability at minimal energy consumption. We present two systems, Triage and Turducken, which are instantiations of HPM for sensor net microservers and laptops, respectively. The second part of the thesis proposes and analyzes the use of additional infrastructure in the form of relays, mesh nodes, and base stations to enhance sparse and dense mobile networks. We present the design, implementation, and deployment of Throwboxes—a relay system to enhance sparse mobile networks and an associated system for enhancing WiFi based mobile networks.



Patrick Deegan: *Whole-Body Strategies for Mobility and Manipulation*; (Roderic Grupen, Advisor); May 2010; Senior Robotics Engineer, Heartland Robotics.

The robotics community has succeeded in creating remarkable machines and task-level programming tools, but arguably has failed to apply sophisticated autonomous machines to sophisticated tasks. The dissertation introduces the uBot-5—a mobile manipulator concept to support new robotic applications in our culture that require fully integrated dexterous robots in unstructured environments. The integrated system provides dexterous modes for mobility and manipulation and control firmware that organizes these behavioral modes logically for use by application code.

The approach chosen in this study centers around a hardware and software co-development. The platform successfully pairs motor flexibility and performance with a hierarchical embedded control framework for constructing dexterous machines. In particular, postural control underlies the uniform treatment of several mobility modes that engage different combinations of sensor and motor resources. The result is a platform for studying “whole-body” control strategies that can be applied jointly to simultaneous mobility and manipulation objectives. Furthermore, dexterous machines can express the “aptitudes” implicit in the design of the robot in the embedded firmware and hierarchically organize the behavior of the system for programming. This is a win-win situation where the quality of the embedded firmware determines how efficiently programmers (autonomous learning algorithms or human programmers) can construct control programs that are robust, flexible, and respond gracefully to unanticipated circumstances.



Andrew Fast: *Learning the Structure of Bayesian Networks with Constraint Satisfaction*; (David Jensen, Advisor); Feb. 2010; Research Scientist, Elder Research Inc.

A Bayesian network is a graphical representation of the probabilistic relationships among a set of variables and can be used to encode expert knowledge about uncertain domains. The structure of this model represents the set of conditional independencies among the variables in the data. In this thesis, I focus on learning the structure of Bayesian networks from data with constraint-based algorithms. These algorithms use a series of conditional hypothesis tests to learn independence constraints on the structure of the model.

I show that new algorithms inspired by constraint satisfaction are able to produce significant improvements in structural accuracy. These constraint satisfaction algorithms exploit the interaction among the constraints to reduce error. First, I introduce an algorithm based on constraint optimization that is sound in the sample limit, like existing algorithms, but is guaranteed to produce a DAG. This new algorithm learns models with structural accuracy equivalent or better to existing algorithms. Second, I introduce an algorithm based constraint relaxation. Constraint relaxation combines different statistical techniques to identify constraints that are likely to be incorrect, and remove those constraints from consideration. I show that an algorithm combining constraint relaxation with constraint optimization produces Bayesian networks with significantly better structural accuracy when compared to existing structure learning algorithms, demonstrating the effectiveness of constraint satisfaction approaches for learning accurate structure of Bayesian networks.



Stephen Hart: *The Development of Hierarchical Knowledge in Robot Systems*; (Roderic Grupen, Advisor); Sept. 2009; Postdoctoral Researcher, Italian Institute of Technology, Genova, Italy.

I investigate two complementary ideas in the literature on machine learning and robotics—to address a unified framework for skill learning and knowledge acquisition. “Embodied” systems make use of structure derived directly from sensory and motor configurations for learning behavior. Intrinsically motivated systems learn by searching for native, hedonic value through interaction with the world. Psychological theories of intrinsic motivation suggest that there exist internal drives favoring open-ended cognitive development and exploration. I argue that intrinsically motivated, embodied systems can learn generalizable skills, acquire control knowledge, and form an epistemological understanding of the world in terms of behavioral affordances.

I propose that the development of behavior results from the assembly of an agent’s sensory and motor resources into state and action spaces that can be explored autonomously. I introduce an intrinsic reward function that can lead to the open-ended learning of hierarchical behavior. This behavior is factored into declarative “recipes” for patterned activity and common sense procedural strategies for implementing them in a variety of run-time contexts. These skills form a categorical basis for the robot to interpret and model its world in terms of the behavior it affords. Experiments conducted on a bimanual robot illustrate a progression of cumulative manipulation behavior addressing manual and visual skills. Such accumulation of skill over the long-term by a single robot is a novel contribution that has yet to be demonstrated in the literature.



Manjunatha Jagalur; *Discovery of Complex Regulatory Modules from Expression Genetics Data*; (David Kulp, Advisor); May 2010; Bioinformatics Scientist, Pacific Biosciences.

Mapping of strongly inherited classical traits has been immensely helpful in understanding many important traits including diseases, yield and immunity. But some of these traits are too complex and are difficult to map. Taking into consideration gene expression, which mediates the genetic effects, can be helpful in understanding such traits. Together with genetic variation data such a dataset is collectively known as expression genetics data. Presence of discrete and continuous variables, observed and latent variables, availability of partial causal information, and under-specified nature of the data make expression genetics data computationally challenging, but potentially of great biological importance. In this dissertation the underlying regulatory processes are modeled as Bayesian networks consisting of gene expression and genetic variation nodes. Due to the under-specified nature of the data, inferring the complete regulatory network is impractical. Instead, the following techniques are proposed to extract interesting subnetworks with high confidence.

The network motif searching technique is used to recover instances of a known regulatory mechanism. The local network inference technique is used to identify immediate neighbors of a given transcript. Application of these two techniques often results in identification of hundreds of individual networks. The network aggregation technique extracts the most common subnetwork from those networks, and identifies its immediate neighbors by collapsing them into a common network. In all the above tasks, simulation studies were carried out to estimate the robustness of the proposed methods and the results suggest that these techniques are capable of recovering the correct substructure with high precision and moderate recall. Moreover, manual biological review shows that the recovered regulatory network substructures are typically biologically sensible.



Jeffrey Johns; *Basis Construction and Utilization for Markov Decision Processes using Graphs*; (Sridhar Mahadevan, Advisor); Feb. 2010; Computing Innovation Fellow, Department of Computer Science, Duke University.

In reinforcement learning (RL), an agent takes actions in an environment and receives rewards. The agent must use its experience in order to learn how best to act in the future. One of the main challenges for an autonomous agent is in representing functions/features over very large and complex environments. The majority of successful, large-scale RL applications have required humans to provide such features to the agent; however, recent research suggests this process of feature construction can be automated and solved by the agent itself. Building on this idea, we propose two algorithms for scaling automatic feature construction to very large data sets. Once the features are computed, the agent must utilize those features to learn how best to behave. We introduce a new least-squares algorithm that allows for the agent to make efficient use of its experience in the environment. Furthermore, we evaluate feature selection methods that tailor the features to the agent's desired task. These feature selection methods encourage sparse solutions and provide regularization, both properties that are necessary when dealing with complex environments.



Victoria Manfredi; *Sensor Control and Scheduling Strategies for Sensor Networks*; (James F. Kurose, Advisor); Sept. 2009; Computing Innovation Fellow, Department of Computer Science, Boston University.

We investigate sensor control and scheduling strategies to most effectively use the limited resources of an ad hoc network or closed-loop sensor network. We first consider where to focus sensing in a meteorological radar network. We show that the main benefits of optimizing sensing over expected future states are when there are multiple small phenomena in the environment. We next investigate how to make sensing robust to delayed and dropped packets. We ground our analysis in a meteorological radar network and show that prioritizing sensor control traffic decreases the round-trip control-loop delay, and thus increases the quantity and quality of the collected data and improves application performance. Finally, we examine how to make routing robust to network changes. We propose a routing algorithm that selects a type of routing subgraph (a braid) that is robust to changes in the network topology. We analytically characterize the reliability of a class of braids and their optimality properties, and give counterexamples to other conjectured optimality properties in a well-structured (grid) network. Comparing with dynamic source routing, we show that braid routing can significantly decrease control overhead while only minimally degrading the number of packets delivered, with gains dependent on node density.



Sarah Osentoski; *Action-Based Representation Discovery in Markov Decision Processes*; (Sridhar Mahadevan, Advisor); Sept. 2009; Postdoctoral Researcher, Computer Science Department, Brown University.

This dissertation investigates the problem of representation discovery in discrete Markov decision processes, namely how agents can simultaneously learn representation and optimal control. Previous work on function approximation techniques for MDPs largely employed hand-engineered basis functions. We explore approaches to automatically construct these basis functions and demonstrate that automatically constructed basis functions significantly outperform more traditional, hand-engineered approaches.

We specifically examine two problems: how to automatically build representations for action-value functions by explicitly incorporating actions into a representation, and how representations can be automatically constructed by exploiting a pre-specified task hierarchy. We first introduce a technique for learning basis functions directly in state-action space. The approach constructs basis functions using spectral analysis of a state-action graph which captures the underlying structure of the state-action space of the MDP. We show how our approach can be used to approximate state-action value functions when the agent has access to macro-actions: actions that take more than one time step and have predefined policies. We describe how state-action graphs can be modified to incorporate information about the macro-actions. Finally, we describe how hierarchical reinforcement learning can be used to scale up automatic basis function construction. We extend automatic basis function construction techniques to multi-level task hierarchies and describe how basis function construction can exploit the value function decomposition given by a fixed task hierarchy. We demonstrate that combining task hierarchies with automatic basis function construction allows basis function techniques to scale to larger problems and leads to a significant speed-up in learning.



Shichao Ou; *A Behavioral Approach to Human-Robot Communication;* (Roderic Grupen, Advisor); Feb. 2010; Senior Software Engineer, Network Equipment Technologies.

This dissertation focuses on how a robot can acquire and refine expressive and receptive communication skills with human beings. I hypothesize that communication has its roots in motor behavior and present an approach that is unique in the following aspects: (1) representations of humans and the skills for interacting with them are learned in the same way as the robot learns to interact with other “objects,” (2) expressive behavior naturally emerges as the result of the robot discovering new utility in existing manual behavior in a social context, and (3) symmetry in communicative behavior can be exploited to bootstrap the learning of receptive behavior.

Experimental results show that the robot successfully acquired a variety of expressive pointing gestures, and the perceptual skills with which to recognize and respond to similar gestures from humans. This illustrates the validity of the approach as a computational framework for learning increasingly comprehensive models and behavior for communicating with humans. Also, due to variations in human reactions over the training subjects, the robot developed a preference for certain gestures over others, showing that the approach can adapt to different human behavior. These results support the experimental hypotheses and offer insights for future studies.



M.S. Raunak; *Resource Management In Complex And Dynamic Environments;* (Leon J. Osterweil, Advisor); Sept. 2009; Visiting Assistant Professor, Department of Computer Science, Loyola College.

Resource management is at the heart of many diverse science and engineering areas. Often a relatively simple model of resources can suffice for work in a number of domains. The problems of resource specification and management become much more challenging, however, when working with a complex real-life domain, such as the emergency department of a hospital, with many heterogeneous resource types and intricate constraints on their utilization. This dissertation proposes an approach for modeling and managing resources in complex and dynamic environments, and presents an architecture that focuses on appropriate separation of concerns. To evaluate this approach we developed ROMEO, an implementation of the general approach proposed in the dissertation. ROMEO supports execution and simulation of complex real-world processes. We have studied the effectiveness of ROMEO’s well modularized separation of concerns by examining how well ROMEO supports execution and simulation of a wide variety of different real-world processes such as hospital emergency department processes, online dispute resolution processes, and web services development processes. Our studies suggest that our choices of concerns to separate offer some important advantages, such as ease of modification, and the ability to represent important fine-scale details.



Bruno Ribeiro; *On the Design of Methods to Estimate Network Characteristics;* (Donald F. Towsley, Advisor); May 2010; Postdoctoral Research Associate, Department of Computer Science, University of Massachusetts Amherst.

Social and computer networks permeate our lives. Large networks, such as the Internet, the World Wide Web, and wireless smartphones, have indisputable economic and social importance. These networks have non-trivial topological features, i.e., features that do not occur in simple networks such as

lattices or random networks. Estimating characteristics of these networks from incomplete (sampled) data is a challenging task.

This thesis provides two frameworks within which common measurement tasks are analyzed and new, principled, measurement methods are designed. The first framework focuses on sampling directly observable network characteristics. This framework is applied to design a novel multidimensional random walk to efficiently sample loosely connected networks. The second framework focuses on the design of measurement methods to estimate indirectly observable network characteristics. This framework is applied to design two new, principled, estimators of flow size distributions over Internet routers using (1) randomly sampled IP packets and (2) a data stream algorithm.



Timothy Richards; *Generalized Instruction Selector Generation: The Automatic Construction of Instruction Selectors from Descriptions of Compiler Internal Forms and Target Machines;* (J. Eliot B. Moss, Advisor); Feb. 2010; Visiting Assistant Professor, Department of Computer Science, Trinity College.

One of the most difficult tasks a compiler writer faces is the construction of the instruction selector (IS). The IS is the part of the compiler that translates compiler intermediate representation (IR) into instructions for a target machine. Unfortunately, implementing an IS by hand is difficult, time consuming, and error prone. The details of the IR and target instruction set is carefully considered in order to generate correct and efficient code. This requires an expert in compiler internals as well as the target machine. In this dissertation we describe the instruction selector problem, cover previous attempts at solving it, and identify what we believe to be the most prominent factor inhibiting their widespread adoption.

This dissertation proposes a generalized approach toward generating instruction selectors automatically. We propose CISL, a common machine description language for specifying compiler IR and target instruction semantics, and GIST, a heuristic search procedure that discovers equivalent instruction sequences between compiler IR and target instructions. GIST leverages CISL’s well-defined semantics to discover IS patterns automatically. Adapter programs use GIST-generated selector patterns to output compiler specific implementation code. Our experiments show that IS patterns can be discovered automatically and independent of a particular compiler framework or target machine.

Alicia P. Wolfe; *Paying Attention to What Matters: Abstraction in Partially Observable Domains;* (Andrew G. Barto, Advisor); Feb. 2010.

Autonomous agents may not have access to complete information about the state of the environment. For example, a robot soccer player may only be able to estimate the locations of other players outside the scope of its sensors. However, even though all the information needed for ideal decision making cannot be sensed, all that is sensed is usually not needed. The noise and motion of spectators, for example, can be ignored in order to focus on the game field. Standard formulations do not consider this situation, assuming that all the can be sensed must be included in any useful abstraction. This dissertation extends the Markov Decision Process Homomorphism framework to partially observable domains, focusing specifically on reducing Partially Observable Markov Decision Processes (POMDPs) when the model is known. This involves ignoring aspects of the observation function that are irrelevant to a particular task. Abstraction is particularly important in partially observable domains, as it enables the formation of a smaller domain model and thus more efficient use of the observed features.

Levine and Liberatore help police apprehend Internet child sexual exploitation predators

Thanks to powerful new software developed by Professor Brian Levine and Research Scientist Marc Liberatore, state law enforcement officers across the country including the Massachusetts State Police now have an extraordinarily effective tool for collecting evidence against people who possess and share illegal images and produce child pornography for the Internet.

It is currently used in 58 out of 61 Internet Crimes Against Children (ICAC) Task Forces around the nation in more than 45 states. Liberatore presented a paper describing the project and its results on Aug. 3 to the 2010 Annual Digital Forensics Research Conference in Portland, Oregon. The previous day, the U.S. Department of Justice released its national strategy for preventing child exploitation, which names the UMass Amherst team as a primary partner in the new strategy. The project is featured in the U.S. Department of Justice's "National Strategy for Child Exploitation Prevention and Interdiction: A Report to Congress," released in August 2010 (www.projectsafefchildhood.gov/docs/natstrategyreport.pdf). Statistics gathered from the project are used in the report as estimates of the volume of child pornography trafficking in the United States.

Levine says that since January, 1,201 search warrants have been issued as a result of his and Liberatore's work, with 2,995 cases open or completed, bringing police investigators closer to apprehension of contact offenders. Contact offenders are people who sexually exploit children to create and distribute pornographic images. Police have made 639 arrests as a result of these cases since January.

With collaborator Clay Shields, a computer scientist at Georgetown University, Levine and Liberatore received funding from the U.S. Department of Justice's National Institute of Justice to design and build software for network forensics. Levine and Liberatore have received additional funding from the National Science Foundation for this digital forensics research collaboration with the Crimes Against Children Research Center at the University of New Hampshire. Levine and Liberatore, along with Assistant Professor Gerome Miklau, also received UMass President's Office Science and Technology Initiative funding to advance digital forensics technology and analysis by establishing a Commonwealth Center for Digital Forensics & Society, a partnership of UMass Amherst and the Massachusetts State Police Crime Laboratory.

The UMass Amherst computer scientists created a program they call RoundUp that allows law enforcement officers to observe and search open peer-to-peer (p2p) networks on the Internet and gather evidence of criminal possession



The graphic shows the approximate location and approximate number of people sharing child pornography over the Internet during a single minute in the United States on Aug. 3, 2010. *Graphic courtesy of the Massachusetts State Police.*

and sharing of images. The software was designed specifically for the challenges posed by investigations of child pornography and it is exclusively used by law enforcement.

The software does not allow law officers to hack into an individual's private computer, Levine and Liberatore are quick to point out. It simply provides law enforcement with an "optimized interface for observation" which allows an investigator to watch the open activities of remote peers on the network. It is a situation they liken to a police officer observing a drug transaction on a street corner. "It's not magic and it's not hacking," says Levine. "This allows regular shoe-leather, routine police work, the steps of which can be tracked and verified just as in any other search for evidence."

The software is used by law enforcement officials who pair it with a watch list of files of interest. RoundUp alerts investigators when p2p users announce they are sharing such files. A unique aspect of the RoundUp system is its ability to aggregate information discovered by investigators using the software in one place. Using the aggregate data, the ICAC Task Forces are able to track the volume of online child pornography trafficking on an almost hourly basis.

In fact, law enforcement partners specifically asked that the software should not be highly automated except for its ability to identify the location of suspicious activity. This is to allow police investigators to retain the ability to use expert judgment, including their knowledge of the law, when following leads. "A real investigator with his or her years of experience and finely tuned sense of what is criminal activity and what is not, is always in charge of this investigative tool," Liberatore points out. "That is a very important aspect of this new trend of using the enormous capabilities of computerized data analysis to fight crime."

Nevertheless, RoundUp is an extremely powerful tool that is generating literally millions of leads worldwide every day. Law enforcement agencies from Interpol and the FBI to big-city police departments across the globe receive tips daily from leads screened by ICAC Task Force members.

Faculty News



Effective this September, **Deepak Ganesan** and **Erik Learned-Miller** were promoted to Associate Professor with tenure, **Brian Levine** and **Sridhar Mahadevan** were promoted to full Professor, and Associate Professor **Yannis Smaragdakis** received tenure. Effective in January, **Arun Venkataramani** will be promoted to Associate Professor with tenure.



Professor **Shlomo Zilberstein** became the President of the ICAPS Executive Council, an international organization that oversees and sponsors the annual International Conference on Automated Planning and Scheduling.



Professor **James Allan** was elected to a three-year term as the Chair of the ACM Special Interest Group on Information Retrieval (SIGIR).



Assistant Professor **Hanna Wallach** and co-authors (Ryan Adams and Zoubin Ghahramani) received the Best Paper Award at the 2010 Thirteenth International Conference on Artificial Intelligence and Statistics for their paper “Learning the Structure of Deep, Sparse Graphical Models.”



Associate Professor **Ramesh Sitaraman** was named an Akamai Fellow in recognition of his pioneering contributions to Internet Content Delivery. On a past leave of

absence from academia, he helped architect and build Akamai’s highly-distributed cloud platform that currently serves hundreds of billions of web user requests per day, utilizing over 73,000 servers in 70 countries and nearly 1000 networks.



Robert Moll, Associate Professor and Associate Department Chair for Academic Programs, received a Student Choice Award at the 2010 Residential First Year Academic Experience Academic Awards Banquet.



Professor **Rod Grupen** gave a keynote address at the Ninth IEEE International Conference on Development and Learning (ICDL 2010), held in Ann Arbor, Michigan. Grupen and the Laboratory for Perceptual Robotics (LPR) organized the 2010 New England Manipulation Symposium, attended by over 60 researchers from the Northeastern U.S. who convened to discuss new directions in robot manipulation. The LPR is teaming up with iRobot and UPenn to participate in the DARPA ARM-S program to discover how robots can manipulate objects in the natural world autonomously.



Associate Professor **David Jensen** gave a keynote address, “Computational Social Science,” at the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2010), the premier international forum for data mining researchers and practitioners.



Assistant Professor **Kevin Fu** was a participant at the President’s Council of Advisors on Science and Technology (PCAST)/President’s Innovation and Technology Advisory Committee (PITAC) Golden Triangle Workshop that took place in Washington, D.C. in June. The workshop focused on information technology, biotechnology, and nanotechnology.



Professor **Sridhar Mahadevan**, the late Professor **Paul Utgoff**, and Adjunct Professor **Ileana Streinu** were recently selected



as three of the top twelve Notable Alumni in Academia of the Rutgers University Computer Science Department.

Assistant Professor **Rui Wang** is the conference co-chair of the 2011 ACM Symposium on 3D Graphics and Games, a top-tier conference in computer graphics.



Adjunct Professor **Lee Spector** received a National Science Foundation grant, “Evolution of Robustly Intelligent Computational Systems,” that is extending the science of automatic programming, using concepts derived from evolutionary biology and software engineering, to permit the evolution of general and robust computational systems with multiple interacting functionalities and interfaces.



Osterweil receives Chancellor’s Award

Chancellor Robert Holub presented Professor Leon Osterweil with the Chancellor’s Award for Outstanding Accomplishments in Research

& Creative Activity during the sixth annual UMass Amherst Faculty Convocation held in October. He was recognized for his research in software engineering that promises to help reduce medical errors, improve election process, make electronic commerce safer, and enable other advances in the technological world.

In addition to Osterweil, Adjunct CS Professor Jane Fountain and CS alum Lixin Gao (Ph.D. ’97) were among the eight members of the UMass Amherst faculty receiving the award at this year’s ceremony.

Distinguished University Professor **Bruce Croft** received a Yahoo! Faculty Research and Engagement Award for his project “Social Applications on the web.” Croft also received a UMass President’s Office 2010 Science & Technology Initiatives Fund Award for a collaborative project with the



UMass Medical School. This summer, Croft was interviewed by John Moe of American Public Media’s Future Tense radio program for a segment on the future of search.

For their years of service at UMass Amherst, Professors **Bruce Croft** (30 years), **Neil Immerman** (20 years), **Edwina Rissland** (30 years), and **Chip Weems** (25 years) each received campus Length of Service Awards.



Distinguished Emeritus Professor **Arnold Rosenberg** gave a lecture at Missouri University of Science and Technology as part of the ACM Distinguished Speakers Program. He recently joined Northeastern University as a Research Professor in the College of Computer and Information Science. In addition, Rosenberg completed his fifth consecutive Falmouth Road Race in August.



Research Professor **Beverly Woolf**, Research Scientist **Ivon Arroyo**, and UMass Amherst Engineering co-author Hasmik Mehranina received the Best Paper Award at the 2010 Third International Conference on Education and Data Mining. In September, *The New York Times* featured an article, “Automated Computer Tutors With a Human Touch,” about Woolf’s and Arroyo’s affective tutors research. This fall, their research was also the subject of a *Wired Magazine* article, and Woolf was interviewed on CNN’s *Chalk Talk*.



This past summer, the World Economic Forum (WEF), based in Geneva, Switzerland, appointed Adjunct Professor **Jane Fountain** chair of Global Advisory Council on the Future of Government. Fountain, who is director of the National Center for Digital Government on campus and the interdisciplinary Science, Technology and Society Initiative, served as a Global Advisory Council member of the WEF for two years before receiving this significant leadership appointment.



Researcher News

The department welcomed three Postdoctoral Research Associates this fall: **Laura Dietz**, **William Yeoh**, and **Ting Zhu**.

Working with the Advanced Computer Research Group, Ping Yi is a Visiting Associate Professor from Shanghai Jiao Tong University.

Imran Zualkernan, Associate Professor of Computer Science and Engineering at the American University of Sharjah, is a Visiting Professor working with the Center for Knowledge Communication.

Working with Associate Professor Chip Weems, **Jung Wook Park** is a Visiting Scholar from Yonsei University.

Research in action

Graduate student Shane Clark is working with donated defibrillators for research on trustworthy computing. Clark’s work is supported by a National Science Foundation Graduate Research Fellowship.



A student at the University of Lugano, **Maryam Esmacili** is a Visiting Scholar working with the Multi-Agent Systems Lab.

Dimitri Ogniberne is a Visiting Scholar from ISTC-CNR, Italy who is working with the Resource-Bounded Reasoning Research Group.

Jeremie Cabessa is a Visiting Scholar from the University of Grenoble, working with Associate Professor Hava Siegelmann.

Student News

Xiaojuan Wu, a first year graduate student working in the Resource-Bounded Reasoning Research Group, received the inaugural Paul Utgoff Memorial Graduate Scholarship in Machine Learning. Wu’s area of interest is artificial intelligence. He received his M.S. from Mississippi State University and a B.S. from Wuhan University in China.

CS alum (B.S. ’10) and current graduate student **Eric Ssebanakitta** was awarded a scholarship from the UMass Amherst National Science Foundation S-STEM (Scholarships for Science, Technology, Engineering, and Mathematics) Scholars Program. The scholarship is renewable for up to four consecutive semesters.

Pallika Kanani, IESL graduate student, received a Best Student Paper Runner-up Award at the 14th Pacific-Asia Knowledge Discovery and Data Mining conference (PAKDD 2010) for her paper, “Resource-bounded Information Extraction: Acquiring Missing Feature Values On Demand,” co-authored by Andrew McCallum and Shaohan Hu.

Graduate students **Aruna** and **Niranjan Balasubramanian** welcomed the birth of their daughter, Nila, born on September 11.

Graduate student **Scott Kuindersma** spent the summer at the NASA Johnson Space Center working on Robonaut 2 (R2). R2 will be on the next Space Shuttle launch going to the International Space Station. Kuindersma is part of the Autonomous Learning Lab and the Laboratory for Perceptual Robotics.

UMass Guide, a new free iPhone app designed by undergrad **Daniel Stewart**, is now available through iTunes for the iPhone. The app provides a convenient way to navigate campus.

Staff News

Valerie Caro (25 years), **Steve Cook** (20 years), **Glenn Stowell III** (20 years), and **Barbara Sutherland** (25 years) each received Length of Service Awards from UMass Amherst.

**Newsletter of the
Department of Computer Science
College of Natural Sciences
at the University of Massachusetts Amherst**

140 Governors Drive
University of Massachusetts Amherst
Amherst, MA 01003-9264

"Significant Bits" is published twice a year by the
Department of Computer Science, University of Massachusetts
Amherst (www.cs.umass.edu).

Your news, suggestions, comments, and contributions are
welcome. Please mail them to the address above or send
them electronically to bits@cs.umass.edu.

Department Chair Andrew Barto
Editor Jean Joyce
Art Direction North Haven Design
Graduate Student Filip Jagodzinski,
Liaisons Borislava Simidchieva
Contributors Eric Brown, David Miller, Erik Learned-Miller,
UMass Amherst News Office staff

NONPROFIT ORG.
U.S. Postage

PAID

Permit No. 2
AMHERST, MA

Thanks for your support

The following alumni and friends have actively supported the Department of Computer Science from April 2010 through September 2010. Such financial support is greatly appreciated and helps maintain a world-class instructional and research program. Contributions from alums and friends help to fund important special activities that are not supported through the state budget.

Those interested in helping the department should visit www.cs.umass.edu/about/donate for online donations or send a check made out to *UMass Amherst* to: Department of Computer Science, Attn: Jean Joyce, University of Massachusetts Amherst, 140 Governors Drive, Amherst, MA 01003-9264
Please state that your gift is restricted to Computer Science.

Mr. John F. Adler ('85)
Dr. James P. Ahrens ('89)
Dr. Kevin D. Ashley ('88)
Dr. Brendon D. Cahoon ('02)
Ms. Jennifer M. Cannan ('03)
Thomas & Jennifer Chistolini
Charitable Gift Fund
Dr. Yuan-Chieh R. Chow ('77)
Carla & Todd Comeau ('85)
Mr. Paul J. Connolly ('57)
Dr. Edmund H. Durfee ('87)
Dr. Zhengzhu Feng ('05)
Dr. Claude L. Fennema, Jr. ('91)
Mr. Michael B. Friedman ('81)

Mr. Dennis P. Gove ('06)
Mrs. Corinne Roy Griswold ('93)
Mr. Kenneth E. Groder, III ('80)
Yong-Qing Cheng &
Cheng-Feng Han ('96)
Sharon & William Hingley
Mr. Charles E. Hurlburt II ('84)
Dr. Ping Ji ('03)
Mr. Christian J. Johnstone ('09)
Mr. Aristotelis Karageorgos ('08)
Mr. Stephen A. Kelley ('78)
Aiko Nomura &
Adam Lavine ('91)
Mr. David J. Lee ('81)

Dr. Zhihong Lu ('99)
Mrs. Woen-Ru L. Ma ('81)
Mr. Michael C. Monks ('87)
Susan & Brian Moriarty ('75)
Dr. Jitendra D. Padhye ('00)
Mr. Charles N. Paliocha ('83)
Dr. Ron Papka ('99)
Dr. Achilleas Papakostas ('93)
Ms. Desislava I. Petkova ('08)
Mr. Rodion M. Podorozhny ('97)
Lucia & Deraldo Portugal
Dr. Jonathan K. Shapiro ('00)
Mr. Lin Song ('01)

Ms. Julia D. Stoyanovich ('98)
Mrs. Renee B. Venne
Dr. Rukmini Vjaykumar ('88)
Mr. Harteg S. Wariyar ('07)
Mr. Bradley E. Wehrwein ('03)
Ms. Bonnie J. Adams &
Dr. Thomas D. Williams ('74)

**Corporate and
Matching Gifts**
Cisco Systems, Inc.
EMC Corporation
Intuit
Mass Mutual Life Insurance Co.

McDonald's Corporation
Microsoft Corporation
UnumProvident Corp.
Yahoo!

Paul Utgoff Memorial Fund
Marina & Philip Braswell
Mrs. Alexandra Utgoff Taylor ('74)
Mrs. Karen Utgoff