Giri Kumaran (Ph.D. '08) Tackles Data Science Issues

How does one estimate the price of a "Stride Rite Naiya Sandal Toddler Girls Size 7W" or a "Pivot Mach 429 Carbon Mountain Bike Frameset, Size Medium"? As a data science manager at eBay Inc., Giridhar ("Giri") Kumaran (Ph.D. '08, advised by Prof. James Allan) led a team that worked on the difficult problem of estimating the final selling price of the millions of diverse items listed for sale on eBay. "Sellers on eBay typically have a hard time determining how much their items are worth," said Kumaran. "They also struggle with how best to set their prices to optimize for sales, profit, sales velocity, seasonal trends, and so on." The real-time price guidance service that Kumaran's team built proved to be a hit with sellers. It led to more successful sales at better prices than before, thus impacting seller happiness and eBay's bottom line.

The real-time price guidance service is powered by a sophisticated machine learning model that utilizes hundreds of features. "Surprisingly, its roots are in very basic information retrieval," noted Kumaran. "The most impactful feature is what similar items sold for recently. These items are found by searching through past sales with the item's title as a query using BM25 as a ranking function!" Working on this problem, first as an individual contributor and later as a team lead, Kumaran found further validation that the key to a successful supervised model lay with the careful design and choice of features rather than the particular algorithm itself. Also playing a critical role in shaping the price guidance service was Kumaran's product management counterpart and fellow UMass Amherst CICS alum, Vikas Khandelwal (M.S. '01, also from the Center for Intelligent Information Retrieval (CIIR) and advised by Allan).

Prior to working on the pricing problem, Kumaran developed bidding algorithms for eBay paid search marketing campaigns. The development of a few features that captured the actual return on investment on bids placed on keywords dramatically changed the campaigns from loss to profit-making for eBay.

Today, Kumaran works as a senior director of data science and business intelligence at Expedia Inc., where he leads efforts to combat fraud and manage risk in online transactions. He believes that to be successful in a career in data science, it is important to get exposure to a wide variety of problems in various contexts. In just seven years since receiving his Ph.D. in Computer Science from UMass Amherst, his work has spanned web search, computational advertising, ecommerce, and now online travel. Kumaran began his career as a scientist at Microsoft Live Labs, where he conducted research in web search. The goal was to fold that research into Microsoft's products. He soon realized that he could have greater impact by joining a product group, in this case Bing, rather than working as a research partner. It was at Bing that Kumaran cut his teeth as a data scientist, teaching himself to use Microsoft's Map Reduce tools and experimenting with petabytes of data. While he believes that his foundation was built at UMass Amherst's CIIR research lab, where he spent months (and years!) analyzing search queries and wading through thousands of documents to identify patterns he could leverage to improve retrieval, it was at Microsoft where he got an opportunity to work at scale. While at Microsoft, Kumaran learned to build predictive models and scalable services, and evangelize new ideas; he also developed a knack to marry his work to business impact. "These are all the skills a person known as a data scientist today is expected to have," noted Kumaran.

Today data science is considered a hot field. More and more businesses are realizing the importance of unlocking the value in the mountains of data they possess, noted Kumaran. He believes that the training of data scientists is not complete if they have not made business cases for their project, gotten their hands dirty cleaning up data, trained a machine learning model with a business objective in mind, gone through the intricacies of deploying their models in a production environment, designed an A/B test, analyzed the online performance of their solutions, and incorporated feedback from performance to inform the next version of the product.

At a high level, a data scientist should have expertise in machine learning, empirical methods, etc., and, an ability to glean insights from data, and leverage those insights to make impact. According to Kumaran, these three skills are critical for a successful career in data science and are progressively difficult to acquire. Even amongst experienced data scientists, the ones who can consistently demonstrate the third skill are the most sought after.

"These are exciting times in the data science field, and it's a thrilling experience being in the frontlines," said Kumaran. "Barriers to entry are being broken, with several startups and big companies offering data science platforms as a service. New applications of data science in fields like healthcare, government, and geographical exploration are emerging. Continual breakthroughs are being made in data engineering and machine learning. All this is great news for the still-maturing field."

Save the Date: April 29, 2016

Recipients of the 8th Annual Outstanding Achievement and Advocacy Awards (OAA) will be honored on Friday, April 29, 2016. The awards recognize the remarkable accomplishments of graduates of the college's degree programs and acknowledge the support of important friends of the college. For details, go to: **cics.umass.edu/outreach/oaa2016**.

Graduation Celebrations

To celebrate the accomplishments of the newest class of computer science alumni, nearly 400 computer science student graduates, families, friends, faculty, and staff gathered at the Worcester Dining Commons on Friday, May 8, 2015 for the graduate student graduation celebration luncheon and on Saturday, May 9, 2015 for the undergraduate student graduation celebration luncheon.

Lori Clarke, then the chair of the School of Computer Science, Eliot Moss, graduate program director, and James Allan, master's degree program director, welcomed each M.S. and Ph.D. graduate to the stage during Friday's festivities. On Saturday, Clarke, Rod Grupen, undergraduate program director, and Tim Richards, chief undergraduate advisor, welcomed each undergraduate to the stage to receive their diploma holder. Benjamin Marlin, honors program director, joined the group on stage to recognize the computer science honors students. After the formal ceremony, undergraduates had the chance to take the microphone to talk about their experiences.

Congratulations to the following Ph.D. graduates who have received tenure-track assistant professor appointments: **Thomas Helmuth** (Ph.D. '15), Department of Computer Science, Washington and Lee University; **George Konidaris** (Ph.D. '11), Departments of Computer Science and Electrical and Computer Engineering at Duke University; **Scott Kuindersma** (Ph.D. '12), School of Engineering and Applied Science, Harvard University; and **Scott Niekum** (Ph.D. '13), Department of Computer Science, University of Texas, Austin.

Ben Carterette (Ph.D. '08), associate professor of computer and information sciences at the University of Delaware, received the Best Paper Award at the 2015 ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR 2015) held in Northampton, MA in September.

Henning Schulzrinne, (Ph.D. '92), the Julian Clarence Levi Professor of Mathematical Methods and Computer Science at the Fu Foundation School of Engineering at Columbia University, has been named the recipient of the 2016 IEEE Internet Award for exceptional contributions to the advancement of Internet technology.

Recent graduate **Ranysha Ware** (M.S. '15), of the Architecture and Language Implementation (ALI) Lab, has taken a position at the MIT Lincoln Laboratory, working on an IARPA program in cybersecurity.

Tuomas Sandholm (M.S. '94, Ph.D. '96), computer science professor and director of the Electronic Marketplaces Laboratory at Carnegie Mellon University (CMU), was quoted in a *Pittsburgh Post-Gazette* article, describing his work on Claudico, a CMU-developed program that plays (and beats) human professional poker players.

Recent Computer Science Ph.D. Graduates (February 2015 and May 2015)

Md. Ashraful Alam; *Reconstructing Geometric Structures from Combinatorial and Metric Information;* Ileana Streinu, Advisor; Feb. 2015; Software Engineer, Intel Corporation

Gordon Anderson; *An Analysis of Two Student Learning Strategies in High Enrollment Computer-based College Courses;* Robert Moll, Advisor; May 2015; Lecturer, College of Information and Computer Sciences, UMass Amherst

Boulat Bash; *Fundamental Limits of Covert Communication;* Donald Towsley, Advisor; Feb. 2015; Postdoctoral Fellow, BBN Technologies

Bruno Castro da Silva; *Learning Parameterized Skills;* Andrew G. Barto, Advisor; Feb. 2015; Associate Professor, Computer Science, Federal University of Rio Grande do Sul, Brazil

Yung-Chih Chen; *Robust Mobile Data Transport: Modeling, Measurements, and Implementation;* Donald Towsley, Advisor; May 2015; Senior Performance Engineer, Akamai Technologies

Stefan Christov; *Model-Based Guidance for Human-Intensive Processes;* Lori A. Clarke, George Avrunin, Advisors; Feb. 2015; Assistant Professor of Software Engineering, Quinnipiac University

Boduo Li; *A Platform for Scalable Low-Latency Analytics using MapReduce;* Yanlei Diao, Advisor; May 2015; Researcher, NEC Laboratories America

Jason Naradowsky; *Learning with Joint Inference and Latent Linguistic Structure in Graphical Models;* David A. Smith, Advisor; Feb. 2015; Research Associate, Machine Reading Lab, University College London

Laura Sevilla Lara; *Long Range Motion Estimation and Applications;* Erik Learned-Miller, Advisor; Feb. 2015; Postdoctoral Researcher, Max-Planck Institute, Germany

Borislava Simidchieva; *Variation in Human-Intensive Systems: a Conceptual Framework for Characterizing, Modeling, and Analyzing Families of Systems;* Leon Osterweil, Advisor; May 2015; Scientist, Distributed Systems Group, BBN Technologies

Michael Wick; *Epistemological Databases for Probabilistic Knowledge Base Construction;* Andrew McCallum, Advisor; Feb. 2015; Senior Member of Technical Staff, Oracle Labs East

Dan Xie; *An Opportunistic Service Oriented Approach for Robot Search;* Allen Hanson, Roderic Grupen, Advisors; Feb. 2015; Senior Software Engineer, Autodesk Inc.

Limin Yao; *Universal Schema for Knowledge Representation from Text and Structured Data;* Andrew McCallum, Advisor; Feb. 2015; Software Engineer, Twitter Inc.

See details on the graduates' research at: **cics.umass.edu/phdgrads_febmay15**

CICS Undergraduate Dean's List – Spring 2015

See the full Spring 2015 Dean's List and the list of student citations (students recognized for their outstanding performance in the classroom) at **cics.umass.edu/people/spring-2015-deans-list-and-citations.**

In July, CICS faculty, graduate students, an alum, and even a future faculty member got together for a photo op at the International Conference on Machine Learning 2015 in Lille, France. Back (l. to r.): Philip Thomas, Akshay Krishnamurthy, Benjamin Marlin, Kevin Winner; Front (l. to r.): Yariv Levy (Ph.D. '12), David Belanger, Andrew McCallum, Tao Sun