An Interview with Dr. Thorsten Joachims, Winner of ACM SIGKDD 2020 Innovation Award

ABSTRACT

Thorsten Joachims, Ph.D. is the recipient of the 2020 ACM SIGKDD Innovation Award, the highest honor for technical excellence in the field of knowledge discovery and data mining. The Innovation Award is conferred annually on an individual or group of collaborators whose outstanding technical innovations have greatly influenced the direction of research and development in the field.

Dr. Joachims is recognized for his research contributions in machine learning, including influential work studying human bias in information retrieval, support vector machines (SVM) and structured output prediction. Notably, he pioneered methods for eliciting reliable preferences from implicit feedback and for unbiased learning-to-rank and ranking that provide fairness guarantees.

Dr. Joachims is currently a professor for the Department of Computer Science and Information Science at Cornell University, as well as an Amazon Scholar. He sat down with SIGKDD Explorations to discuss his career, his research and the biggest challenges facing the field.

CONGRATS ON RECEIVING THE ACM SIGKDD INNOVATION AWARD. TELL US ABOUT YOURSELF. WHO IS DR. THORSTEN JOACHIMS?

I grew up in Germany and was lucky enough to take a computer science program in high school, which was very unusual at the time. From an early age, I was fascinated with artificial intelligence (AI) and knew I wanted to make that my career.

After getting my start in computer science at the University of Dortmund, I spent what, at first, was supposed to be nine months at Carnegie Mellon University (CMU) working with Tom Mitchell, widely considered one of the founding fathers of modern machine learning for his contributions to machine learning, AI and cognitive neuroscience. This turned into two of the most formative years of my career. The web was just starting at the time, presenting new and interesting problems in machine learning.

After that, I went back to Germany to finish my undergraduate degree and then my Ph.D. However, professionally and personally, I got "stuck" in the U.S. meeting my spouse at CMU, and we both have been holding faculty positions at Cornell since 2001.

WHEN DID YOU DECIDE TO GO INTO THE FIELD OF KNOWLEDGE DISCOVERY AND DATA MINING?

There's actually a competition in Germany called the Bundeswettbewerb Informatik, which is a little bit like ACM programming contest in the U.S. As a young student, I loved taking part in these events, especially the challenges about creating intelligent behavior. Many of the problems were puzzles which required the use of search and backtracking to find a solution that was not necessarily the obvious one. In solving these challenges, you could actually think that the computer was better at it than the programmer, which fascinated me. The

transition to data-driven forms of intelligence came with the emergence of the Web, where data about real-life behavior became available at a large scale.

My own experience is a testament to how competitions and other outreach activities at the high-school level can be very effective in attracting young people to the field.

WHAT DOES THIS RECOGNITION FROM SIGKDD MEAN TO YOU?

It means a lot to receive the Innovation Award from ACM SIGKDD. More than any other community, SIGKDD manages to engage both industry and academia every year. I first attended KDD in 2002. I'd learned about the conference from my colleague Jon Kleinberg and thought it was interesting mix of new ideas and communities coming together, so I decided to submit a research paper. I've found KDD is extremely good at identifying new, challenging problems that exist in the world, doing the research into the topic and then actually getting the research out into the world. So, being recognized for innovation from this community is a big honor.

To me, this award recognizes the ability to build systems. I've always been motivated by the engineering aspect of building systems— from the fundamental research to their scalability in practice. I like to see my work deployed in the world, hopefully making a difference in people's life for the better. That motivation comes through in the work that I've done in research and is what SIGKDD itself exemplifies.

HOW HAS THE FIELD CHANGED SINCE THE BEGINNING OF YOUR CAREER?

Since I started out in computer science, all of these different and novel techniques have come together to build the online systems, like search engines and recommender systems, that are now widely used by everyone on a daily basis. We've not only learned many different techniques to make these systems perform well, but also gained insight into how these systems interact with humans. I've been intent on understanding how these systems are trained in interaction with humans by learning from human behavior while also providing humans with information that is useful to them. This is a long-running thread throughout my work.

In Cornell's computer science department, we have a tradition of bringing together all faculty to have lunch and coffee with faculty candidates. During my own interview in 2001, I remember claiming over coffee that all this data we are collecting from search engines – they were just starting to collect click-data at the time – was soon going to lead to public awareness of privacy implications and, subsequently, regulation. I got the job, despite my predictions being a few decades off.

Today, there is more awareness about the huge impact of artificial intelligence and machine learning on people's lives. It's past time to think about how we as data scientists can design these systems to be fair and equitable to all participants. This ideas goes deep down into the machine learning methods and how we train these systems. There is new and existing research around how we actually ensure systems behave in a desirable way, how we implement that from a technical perspective, and how we make policies around what these systems are

supposed to do. I think its an interesting and important research agenda.

WHAT DO YOU BELIEVE ARE THE BIGGEST CHALLENGES FACING THE INDUSTRY?

Some of the biggest challenges of course are the pressing issues of consumer privacy and issues of fairness, especially around factors like race and gender. I'm also curious about how the AI systems we are building will be effectively audited. We must understand and update our current policies and laws, which were written to emulate actual people making their own decisions in everything from driving to education. It is very unclear how we as a society will come to terms with regulating the AI systems that help automate decisions for people, be they consequential or everyday decisions. Ultimately, we will need to address these questions and update our polices so the systems we create will have positive societal implications for all people.

WHAT EXCITES YOU ABOUT THE FUTURE OF THE INDUSTRY?

For online systems, like search, product and movie recommendations, we've developed techniques that are very effective. We are now in the process of taking some of these techniques into other domains, which is exciting. At the same time, other domains may not be as forgiving. If we think about it, getting a bad movie recommendation is one thing but getting a bad job recommendation is a whole other thing.

There is a lot of promise in transferring the systems we've developed to consequential questions in hiring, college decisions, banking/loans and criminal justice. But, clearly, there are different levels of importance, even on the level of individual decision making. The question is how do we make sure the promises of these techniques are actually able to be realized, so we are not stuck with any of the unintended consequences?

CAN YOU SHARE SOMETHING YOU ARE CURRENTLY WORKING ON THAT INTERESTS YOU?

One thing that I've been working on for the past years is the question of fairness in ranking systems. For example, if you are an e-commerce provider, you actually have multiple types of participants. On one side is the traditional consumer that comes into the system, types in a query, finds the product they are looking for and then makes a purchase. Most of the work in machine learning about search and recommendation has been purely focused on servicing these consumers by providing guarantees and good performance to that side of the market.

But, of course, there is also the other side: the sellers in the marketplace. We're looking more closely at what we can do to guarantee that the sellers are treated fairly within our systems. Ultimately, the way an e-commerce provider is presenting products to the users influences how much money the seller will make. So, we've been thinking about ranking techniques that not only provide good performance to consumers but also provide certain guarantees to sellers that they are getting the exposure they deserve. That's one interesting aspect of fairness that I think we'll see much more of when we discuss these ranking systems.

HOW HAS COLLABORATION FURTHERED YOUR CAREER AND RESEARCH?

One thing I've always enjoyed, and found to be very important for my research, is actually visiting organizations and working with collaborators from different companies. I get inspired learning about their problems, both current and what they think their problems will be in five years.

Right now, I'm working part time for Amazon music. It's fascinating to get an inside view and role in how these systems are actually built and designed. These experiences always present interesting new challenges which in turn, informs my research.

BESIDES THE SIGKDD INNOVATION AWARD, YOU HAVE EARNED NINE BEST PAPER AWARDS AND FOUR TEST-OF-TIME AWARDS. WHAT ADVICE DO YOU HAVE FOR DATA SCIENTISTS AT THE BEGINNING OF THEIR CAREERS?

If you are at the beginning of your career, I think you should look closely at the interesting new challenges of the day. You should think about what is going to be important to society and business not just tomorrow, but five or ten years from now.

The other piece of advice is to publish less and dedicate more time to making each paper as strong and convincing as it can be. There is always pressure to publish more, but I think it pays off to go the extra mile on each paper.