An Interview with Dr. Balaji Krishnapuram, Winner of SIGKDD Service Award

ABSTRACT:

Dr. Balaji Krishnapuram, director and distinguished engineer at IBM Watson Health, is honored for his contributions to society through the development of machine learning products to improve healthcare. The ACM SIGKDD Service Award is the highest service award in the field of knowledge discovery and data mining. It is conferred on one individual or one group for their outstanding professional services and contributions to the field of knowledge discovery and data mining. Dr. Krishnapuram sat down with *SIGKDD Explorations* to discuss how he first got involved in the KDD Conference, his work at IBM Watson Health and what exites him about the future of machine learning, data science and artificial intelligence.

CONGRATULATIONS ON RECEIVING THE SIGKDD SERVICE AWARD! TELL US ABOUT YOURSELF. WHO IS DR. BALAJI KRISHNAPURAM?

I am a distinguished engineer and director at IBM Watson Health, where I've led a team of scientists and software engineers since February 2015. I've been involved with SIGKDD since 1999 when I was in the Ph.D. program at Duke University. I've been a volunteer with KDD since 2007. Over the years, I have been involved as a KDD Cup chair, a sponsorship chair and one of the conference co-chairs at KDD 2016.

WHEN DID YOU DECIDE TO GO INTO THE FIELD OF DATA MINING AND MACHINE LEARNING?

I've always been very interested in analyzing data, ever since my high school days. I enjoyed learning how to set a function, for example a straight line or the quadratic curve, through given data samples, and then used the function to predict other samples. In my junior year of college, I was exposed to modern methods for analyzing data, which introduced completely new ideas to me. I was inspired by questions like, "What if you had a bunch of samples without any labels given to you?" I wondered whether that could help make better predictions. Ever since, I've been very drawn to the field because it has so many different applications in so many areas of modern science.

WHEN DID YOU FIRST DISCOVER KDD? WHAT ATTRACTED YOU TO THE COMMUNITY?

As a Ph.D. candidate at Duke, I reviewed research papers from many different conferences. Of all these conferences, KDD plays a particularly interesting role because it is not solely focused on researchers and academics. KDD is unique in that it bridges two different worlds together—theoretical and applied. The intersection of both is required to make an impact of society, and that's what drew me to this community.

YOU'VE HAD MANY VOLUNTEER ROLES WITHIN KDD, INCLUDING CO-CHAIR IN 2016. WHAT MADE YOU DECIDE TO GET INVOLVED IN KDD?

I've noticed a fundamental trend: As a society, we are undergoing the industrialization of data science. Essentially, it's the equivalent of what happened with manufacturing in the 1920s when the profession transitioned from artisanal work to mass production. Data science is becoming accessible to a much larger set of people. That's a fundamental transformation in the technology, but it also has tremendous implications for society. For conferences, the transformation means a substantial change in our demographics. Up until 2007, about 85% of conference attendees were researchers who had a paper to publish. Today, that is completely flipped; about 85% of our attendees are practitioners who want to put a new system of production in their companies. I'm inspired to explore how we can continue to best serve this evolving community and reach an even bigger audience.

WHAT EXCITES YOU ABOUT THE FUTURE OF KDD AND THE INDUSTRY AS A WHOLE?

As I previously mentioned, we are in the early stages of a major new industrial revolution. The first industrial revolution was steam engines, the second was electricity and manufacturing, the third was the arrival of chips and electronics. Machine learning, automated decision making, AI and algorithms make up the fourth industrial revolution. This is going to fundamentally transform society and I'm really excited about how we go about creating a new system that moves our whole society forward without leaving anybody behind. I'm particularly interested how this will relate to healthcare since that will help so many in their everyday lives.

CAN YOU EXPAND ON THAT? WHAT ABOUT HEALTHCARE AND DATA SCIENCE IS PARTICULARLY INTERESTING TO YOU?

Taking a step back, essentially the whole healthcare system was designed in an earlier era when delivery of care centered on patients with acute conditions. However, today, the majority of patients have chronic conditions. Diseases that used to be terminal are now becoming chronic, so the care delivery model developed over 100 years ago is not well suited for treating a patient for the rest of their life. We need

to think about delivering care in a completely different way and technology empowers patients to have much more control. In the new delivery system, AI will play a significant role. But it's critical to understand, it is not going to replace doctors.

I see five different ways AI and machine learning can impact healthcare. The first is population management, which determines who is at risk and identifies interventions that will reduce the risk. The second is care management; designing individual care plans that can be managed by multiple members of a care team. The third way is patient self-management or enabling self-care plans by providing incentives for behavioral changes that lead to improved health. The fourth is optimizing the healthcare process through improved systems, ultimately reducing costs. The fifth and final is the implementation of machine learning to improve predictions for health outcomes.

AS THE DIRECTOR AND DISTINGUISHED ENGINEER FOR IBM WATSON HEALTH, CAN YOU SHARE SOMETHING YOU ARE CURRENTLY WORKING ON THAT REALLY INTERESTS YOU?

Our team established a product called Watson for Drug Discovery that helps pharmaceutical companies identify potential targets for new drugs more efficiently and at a lower cost than traditional methods. The Barrow Neurological Institute has used it to identify nine examples of RNA binding proteins whose genetic mutations were previously implicated in the increased risk of the onset of ALS. The product then analyzed patterns, published literature, 3D structures of proteins, gene sequences and other such information to automatically identify and rank all RNA binding proteins. Of these proteins, nine of the top 10 were found to be in elevated levels in patients with ALS, and five of those nine were previously unknown but newly verified in the lab to be relevant for this disease.

Watson for Drug Discovery can also find previously approved drugs that attract these proteins and their pathways, thus accelerating the sourcing of drugs which are known to be safe and potentially can be reused to solve ALS disease. This product is an excellent example of how machine learning and data mining can advance human health.

YOU ARE A SELF-DESCRIBED "SERVANT LEADER OF TEAM THAT DEVELOP SOFTWARE PRODUCTS USING MACHINE LEARNING: WHAT DOES THAT TITLE MEAN TO YOU?

To me, a servant leader is a leader that exists to serve the team and be a problem solver. Often many managers are fond of saying "don't come to me with problems, come to me with solutions to problems." I think that is the wrong approach. As a servant leader, it's my job to help the team by providing clarity, vision and goals. I assist with either technical issues – math or algorithms or architecture – or issues related to infrastructure, processes, tools, dependencies, timeline and budget. No matter how many ways we could potentially solve a problem, my role is to help the team identify the best approach and then succeed.

WHAT DO YOU BELIEVE ARE THE BIGGEST CHALLENGES FACING THE MACHINE LEARNING/DATA SCIENCE SECTORS?

At present there's a lot of confusion about data ownership and privacy. Most known AI and machine learning are fundamentally fueled by data, but who owns the data and surrounding legal rights can remain unclear. For example, in healthcare, how do we protect patient privacy and anonymize personal information while still being able to use data in general statistic models? Another concern is accountability – who bears the risk for a misdiagnosis or an accident involving a self-driving car?

I think we're heading towards a world where more things are going to be powered by data mining and machine learning, and our challenge is to make sense of all of this. We must ask ourselves how we perceive or make sense of data, make decisions using data, plan to deliver optimal results and act efficiently and economically.

IF YOU WERE TO GIVE THE NEXT GENERATION SOME ADVICE ABOUT GETTING STARTED IN THE FIELD, WHAT WOULD YOU SAY?

Welcome to the team! Together we are one large, international, collaborative community, so the next generation should feel free to talk to anyone. We are very happy to help our new members get involved. Keep participating and going to conferences because you'll find you learn much more rapidly than if you sit at home reading a bunch of papers. Tap into the community and let them help you. There are many ways to volunteer, so find a way that you can contribute. The more you contribute the more others will help you.

WHAT DOES WINNING THE SERVICE AWARD THIS YEAR MEAN TO YOU?

In my opinion it's a recognition of what's been going on in the whole community. We came together as a collective who together as one made this happen. This award is recognition for a large number of people who together made KDD the biggest conference of its kind.