Special Session: A Technical Research Agenda in Data Ethics and Responsible Data Management

Julia Stoyanovich
Drexel University
stoyanovich@drexel.edu

Bill Howe
University of Washington
billhowe@uw.edu

HV Jagadish
University of Michigan
jag@umich.edu

SESSION DESCRIPTION

Recently, there has begun a movement towards fairness, accountability, and transparency (FAT) in algorithmic decision making, and in data science more broadly [1–4]. The database community has not been significantly involved in this movement, despite "owning" the models, languages, and systems that produce the input to the machine learning applications that are often the focus in data science.

If training data are biased, or have errors, it stands to reason that the algorithmic result will also be unfair or erroneous. Similarly, transparency of just the algorithm is usually insufficient to understand why certain results were obtained: one needs also to know the data used. In short, FAT depend not just on the algorithm, but also on the data. This observation raises several important questions:

What are the core data management issues to which the objectives of fairness, accountability and transparency give rise? What role should the database community play in this movement? Will emphasis on these topics dilute our core competency in techniques and technologies for data, or can it reinforce our central role in technology stacks ranging from startups to the enterprise, and from local non-profits to the federal government?

This special session features leading researchers from machine learning, software engineering, security and privacy, and natural language processing, who are doing exciting technical work in FAT. The goal of this session is to outline a technical research agenda in data management foundations and systems around data ethics.

PARTICIPANTS

Julia Stoyanovich is an Assistant Professor of Computer Science at Drexel University, and an affiliated faculty at the Center for Information Technology Policy at Princeton. She is a recipient of an NSF CAREER award and of an NSF/CRA CI Fellowship. Julia's research focuses on responsible data management and analysis practices: on operationalizing fairness, diversity, transparency, and data protection in all stages of the data acquisition and processing lifecycle. She established the Data, Responsibly consortium, serves on the ACM task force to revise the Code of Ethics and Professional Conduct, and is active in the New York City algorithmic transparency effort. In addition to data ethics, Julia works on management and analysis of preference data, and on querying large evolving graphs. She holds M.S. and Ph.D. degrees in Computer Science from Columbia University, and a B.S. in Computer Science and in Mathematics and Statistics from the University of Massachusetts at Amherst.

Bill Howe is Associate Professor in the Information School, Adjunct Associate Professor in Computer Science & Engineering, Senior Data Science Fellow and Founding Associate Director of the UW eScience Institute, Director of the UW Urbanalytics Group, and Founding Chair of the UW Data Science Masters Degree. He has received two Jim Gray Seed Grant awards from Microsoft Research for work on managing scientific data, has had two papers selected for VLDB Journal's Best of Conference issues, and co-authored what are currently the most-cited papers from both VLDB 2010 and ACM SIGMOD 2012. Howe developed a first MOOC on data science that attracted over 200,000 students across two offerings, and founded UW's Data Science for Social Good program. He has a Ph.D. in Computer Science from Portland State University and a Bachelor's degree in Industrial & Systems Engineering from Georgia Tech.

HV Jagadish is the Bernard A Galler Collegiate Professor of Electrical Engineering and Computer Science, and Distinguished Scientist at the Institute for Data Science, at the University of Michigan in Ann Arbor. Prior to 1999, he was Head of the Database Research Department at AT&T Labs, Florham Park, NJ. He is a fellow of the ACM (since 2003), fellow of AAAS (since 2018) and serves on the board of the Computing Research Association (since 2009). He has been an Associate Editor for the ACM Transactions on Database Systems (1992-1995), Program Chair of the ACM SIGMOD annual conference (1996), Program Chair of the ISMB conference (2005), a trustee of the VLDB foundation (2004-2009), Founding Editor-in-Chief of the Proceedings of the VLDB Endowment (2008-2014), and Program Chair of the VLDB Conference (2014). Since 2016, he is Editor of the Morgan & Claypool "Synthesis" Lecture Series on Data Management. Among his many awards, he won the ACM SIGMOD Contributions Award in 2013 and the David E Liddle Research Excellence Award (at the University of Michigan) in 2008. He has developed a popular MOOC on Data Science Ethics that is carried by both Coursera and EdX.

REFERENCES

[1] Serge Abiteboul, Gerome Miklau, Julia Stoyanovich, Gerhard Weikum. Data, Responsibly. Dagstuhl Seminar 16291 available at http://drops.dagstuhl.de/opus/volltexte/2016/6764/

[2] H. V. Jagadish, Ethics of Data Science, MOOC available for free at https://www.edx.org/course/data-science-ethics-michiganx-ds101x

[3] Julia Stoyanovich, Serge Abiteboul, Gerome Miklau. Data Responsibly: Fairness, Neutrality and Transparency in Data Analysis. Tutorial at EDBT 2016.

[4] Julia Stoyanovich, Bill Howe, Serge Abiteboul, Gerome Miklau, Arnaud Sahuguet, Gerhard Weikum. Fides: Towards a Platform for Responsible Data Science. SSDBM 2017: 26:1-26:6.