Fairness, Accuracy and Transparency in Machine Learning

Overview

Automated decision making is a key component of the so called smart solutions, and like all other countries, India too is in the process of putting in place smart solutions in diverse areas– be it monitoring sewage treatment plants, or in scheduling public transport, in disaster management, or in controlling the spread of a deadly disease, etc. Automated decision making now is essentially data driven– where the data is huge– and what is needed is to find clusters, make use of classifiers, or, more generally, find patterns in a large dimensional huge data space. We need to use machine learning to carry out these tasks. As huge data drive decision making in all aspects of our life, whether it is for marketing, information gathering and search, or even for medical diagnosis, prison sentencing or for financial decision making, machine learning either already is, or is going to be everywhere.

Automated decision making through machine learning is very attractive: it is much faster than human decision making, is able to find patterns in much larger collections of data, and can integrate different sources in a way that humans find difficult, if not impossible. But is it any better? It has often been argued that decisions based on machine learning are less biased than human decision making because mathematics is blind. But a number of high profile embarrassments for automated decision making have shown that while algorithms may not always be biased in the same way that humans are, they are often trained to reflect biases and prejudices in the underlying data they are trained on. What is worse is that as machine learning algorithms become more and more complicated, it is even harder to understand why they make their decisions the way they do. In other words, it is becoming increasingly clear that machine learning algorithms may not be fair, they are not accountable, and they are far from transparent.

The objective of this course is to expose students in data analysis to the many challenges in making machine learning tools fair, accountable and transparent, and to discuss the currently available solutions. The course is for 2 credits with 15 one-hour lectures and 5 one-hour tutorial/discussion sessions.

Modules

- Course Start Date: 11/12/2016; Course End Date:20/12/2016
- Number of Participants (maximum): 50 (Preference will be given to the participants registering against 2 Credits)
- How are machine learning algorithms evaluated? How do we interpret their results? (1 Lecture)

- Fairness: (6 lectures)
  - What is fairness?
  - Fairness-preserving methods: modifying the classier, or modifying the data, or modifying the results.
- Interpretable machine learning: (4 Lectures)
  - How can we interpret the results of machine learning in a user-friendly way,
  - Can we generate causal explanations as opposed to correlations?
- Verifiable learning (2 Lectures)
  - Can we verify the results of a machine learning task even if we can't compute the results ourselves?
- Case studies of bias in machine learning (2 Lectures)
- Tutorials/Discussions (5 Hours)

Who can attend...

- Faculties, Engineers, Scientists, and Researchers from academic, industrial and government organizations including R&D laboratories from India or abroad.

- Students at all levels (BE/BTech/MSc/ME/MTech/PhD/Other) from academic and technical institutions/universities from India or abroad.

  Number of participants for the course will be limited to fifty. Preference will be given to the participants opting against credits.

Fees

The participation fees for taking the course is as follows:

Participants from abroad : US $500
Industry/Research Organizations: INR 10,000
Academic Institutions: INR 2,000 (half for SC/ST students)

The above fee include all instructional materials, computer use for tutorials and assignments, laboratory equipment usage charges, 24 hours free internet facility. The participants will be provided with accommodation on payment basis.

The Faculty

Dr. Suresh Venkatasubramanian is a BTech from IIT Kanpur and a PhD from Stanford University, and is currently an Associate Professor in the School of Computing, University of Utah. After his PhD in 1999, Suresh joined the AT&T Labs-Research where he was till 2007. He has been a Visiting Faculty at the Aarhus University, and a Visiting Scientist at Google, Inc., and in Simons Institute for Theoretical Computer Science in UC, Berkeley. He has been awarded the NSF Career award (2010-15), and held John E. and Marva M. Warnock Presidential Endowed Chair for Faculty Innovation in Computer Science during 2007-13. Suresh is in the Editorial Board of Internet Mathematics and an Associate Editor of International Journal of Computational Geometry and Applications, and the Academic Editor of Peer J Computer Science. He has been in the Program Committees of SWAT, SIAM Conf. on Data Mining, IEEE Intnl. Conf. on Data Mining, STACS, NIPS, ACM SIGKDD, SODA, etc. His research is in the area of algorithms and computational geometry, with current focus on data mining and large data (and large dimension) geometric questions.

Prof. Somenath Biswas obtained his B.Tech (Electronics and Electrical Communication Engg.) from IIT Kharagpur in 1973, M.Tech and Ph.D in Computer Science, from IIT Kanpur, in 1976 and in 1980, respectively. Joined the faculty of the Computer Science and Engineering Department of IIT Kanpur in 1980, where currently he is the Sanjay and Rachna Pradhan Chaired Professor. He is a Fellow of the National Academy of Sciences, India. Has been a Visiting Associate Professor at the Aarhus University (1989) and a Visiting Professor at University of Nebraska, Lincoln (1999). Has been the Department Head (1986-87, 1995-97), and the Dean, Faculty Aairs (2005-07), IIT Kanpur. Has been the President of the Indian Association for Research in Computing Science (2000-2002). His Research interest includes computational complexity theory, randomized algorithms, computational biology, logic in computer science.

Course Coordinator

Prof. Somenath Biswas
Phone: 0532-2922027
Email: sb {at} iiita {dot} ac {dot} in

.....

https://gian.iiita.ac.in/