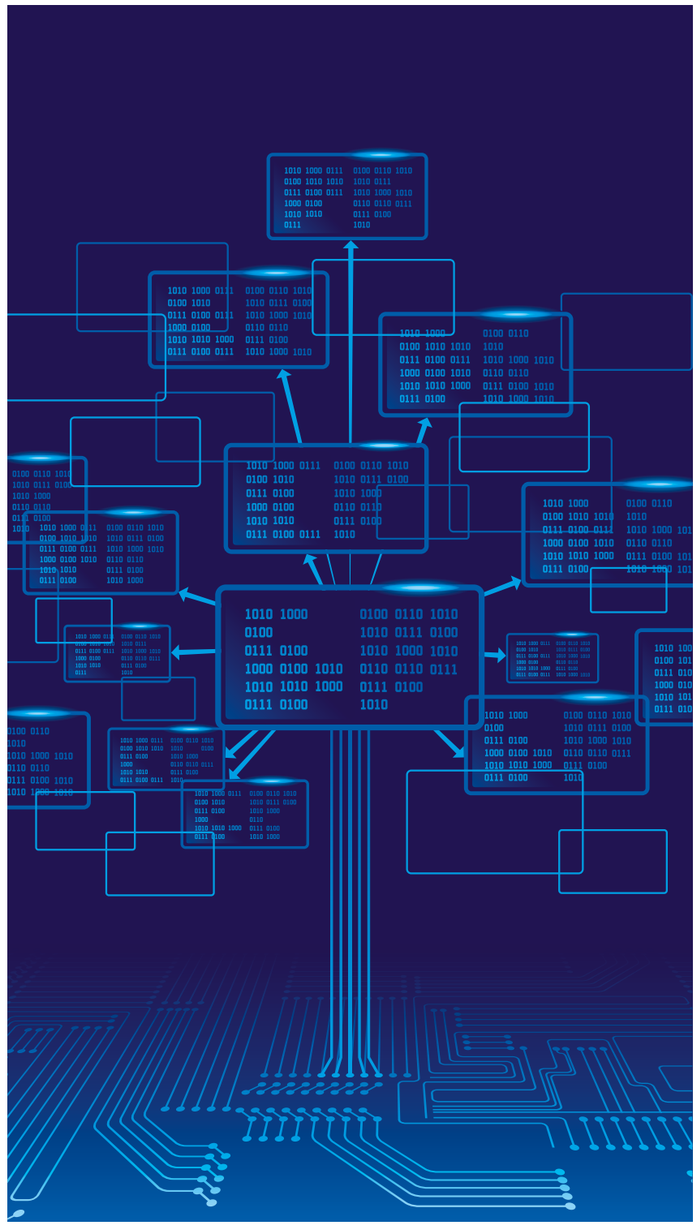
Biases from Big Data:

The prejudiced computer

Big Data and Machine Learning seem to be the modern buzzword answers for every problem. Areas such as healthcare, fraud prevention and sales are just a few of the places that are thought to benefit from self-learning and improving machines that can be trained on huge datasets. However, how carefully do we scrutinise these algorithms and investigate possible biases that could skew results? ***Professor Olfa Nasraoui*** *at the University of Louisville has demonstrated that this is done not nearly carefully enough and is developing tools to lift the lids on 'black box' algorithms and create truly fair alternatives.*



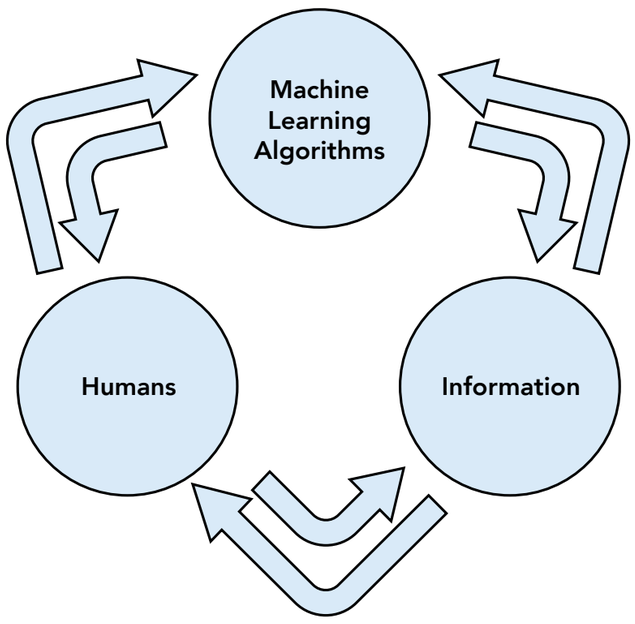
Big data is a generic term for any dataset that is large in volume or variety. It may also be large in 'velocity', a term for the rate at which new data is being added to the existing dataset. One example of a 'big data' dataset might be a census, with a huge number of entries (people) and a variety of information (age, gender, location).

Such large datasets are becoming increasingly common as large-scale data storage has become more practical as well as an increasing number of possibilities for tracking user behaviour on websites and app usage. While such complex datasets may contain valuable information on why customers choose to buy certain products and not others, the size and scale of the available data makes it unfeasible for a human being to analyse it and identify any patterns present.

This is why machine learning is often touted as the answer to the 'Big Data Problem.' Automation of the analysis is one approach to deconstructing such datasets, but conventional algorithms must be pre-programmed to compare particular factors and look for certain levels of significance. An automated algorithm capable of learning and adapting to the dataset offers much greater levels of flexibility in the analysis and can offer far deeper, and potentially original, insights into any trends. This is what is inspiring the use of machine learning in an increasing number of areas, such as education, justice and criminal investigation.

Dedicated mentoring, teamwork, hard work, service, outreach and research dissemination are pillars at the Knowledge Discovery & Web Mining Lab. At a research symposium (left to right): Wenlong Sun, Mahsa Badami, Prof Olfa Nasraoui, Behnoush Abdollahi, Gopi Nutakki.

There is a pressing need for transparency in machine learning models



Humans and algorithms are tightly coupled within a feedback loop. They influence each other via the information or the data generated by humans and by algorithms who guide them,

While a self-teaching, developing algorithm may sound wonderful, machine learning algorithms are often only as good as the datasets on which they have been trained. It also appears that the computer may not be a dispassionate, impartial analysis tool either. With her students and collaborators, Professor Olfa Nasraoui at the University of Louisville has been investigating the issue of how bias can affect machine learning, rendering results unreliable, and how the behaviour of such algorithms can be monitored.

POLARISING DATA

Bias in machine learning models is a crucial issue as such results are now being used in systems such as informational filtering and personalisation. This means there is a continuous feedback loop between the user and system and the algorithm can eventually restrict the information available to the user. This also raises questions about the ethics of using such models if they, even inadvertently, lead to the manipulation of users and perhaps to discrimination against others.

There are several ways that bias can creep into machine learning algorithms. One is through biases in the sampling to create the dataset. For example, if the sampling for a dataset that was supposed to be representative of the population's shopping habits was only performed at a university, the results from the final dataset would be inherently biased because the sampling would over-represent the student demographic.

Iterative bias, common in user-recommendation systems such as those found on most online shopping platforms, is created by a feedback loop between the

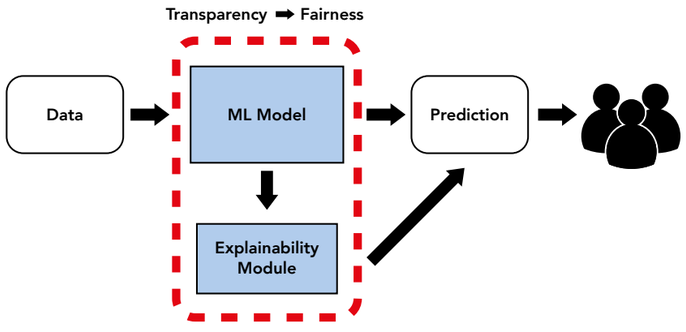
Prof Olfa Nasraoui's students at the Knowledge Discovery and Web Mining lab. Rrom left to right: Wenlong Sun, Mahsa Badami, Gopi Nutakki.

user and system. For recommendation, it is impossible to train the algorithm on stale benchmark datasets which is why Professor Nasraoui and her students Wenlong Sun and Mahsa Badami, along with collaborator Prof Patrick Shafto, have created cognitive models to try and benchmark such rating systems and also what are called counter-polarising systems. These break the positive feedback loop between the user and system and encourage the recommendation of items that will be genuinely new to the user, freeing them from their algorithmic chains and feedback loops.

OPENING THE BOX

All of these factors are why Professor Nasraoui feels there is a pressing need for transparency in machine learning models. Many models are 'black boxes', such as deep learning networks and matrix factorisation. This means that the model cannot give explanations for why certain results are achieved. The results obtained may be accurate, but it is not clear how they were obtained so it is difficult to probe their reliability.

Open, or 'white box' systems, are typically less accurate but the rules and decision trees which are utilised in the process are interpretable. This offers several advantages. It is possible to assess the validity of a prediction or, if there are errors, understand why these prediction errors occurred. Professor Nasraoui and her former doctoral student Behnoush Abdollahi have been developing such a system for recommendations that continues to proactively learn to make explainable predictions, overcoming issues with accuracy, but tries to remain more explainable in its decisions and results than alternative black box methods.



Explainability plays a critical role in transparency. The latter is a pillar of fairness in machine learning models.

Prof Nasraoui (centre) at the Doctoral hooding ceremony with her former PhD students, Behnoush Abdollahi (left) and Gopi Nutakki (right).

CAREFUL DECISIONS

Professor Nasraoui's work as part of the Knowledge Discovery and Web Mining Lab has far-reaching implications in the fields of big data and machine learning. The issue of bias in data, either through sampling or issues with feedback loops in the algorithm, means that the results of any machine learning approach should be carefully considered and Professor Nasraoui has been developing tool and algorithms to do that. She is also working on alternative approaches to black box methods that may help users make fully informed decisions about the data they are using, something that is becoming increasing critical with the growing reliance on the results of these types of analysis.

Behind the Bench

Professor Olfa Nasraoui

E: olfa.nasraoui@louisville.edu **T:** +1 901 491 3851 **W:** http://webmining.spd.louisville.edu

Knowledge Discovery & Web Mining Lab
Dept. of Computer Engineering & Computer Science
Speed School of Engineering
University of Louisville
Louisville, Kentucky 40292 USA

Bio

Olfa Nasraoui is a Professor of Computer Engineering and Computer Science, Endowed Chair of e-commerce, and the founding director of the Knowledge Discovery and Web Mining Lab at the University of Louisville. She received her PhD in Computer Engineering and Computer Science from the University of Missouri-Columbia in 1999. She has more than 160 refereed publications, including over 40 journal papers and book chapters and eight edited volumes.

Research Objectives

Professor Nasraoui's work focuses on Big Data. She examines how Machine Learning can lead to unreliable and biased models, problems around explainability and whether increased personalisation contributes to polarisation of opinions.

Funding

- National Science Foundation
- Kentucky Science & Engineering Foundation

Collaborators

Students: Wenlong Sun
Former students: Behnoush Abdollahi , Mahsa Badami, Gopi Nutakki
Colleague: Prof Patrick Shafto, Rutgers University, who collaborated with Professor Nasraoui on her work on filter bubbles.

Q&A

Can you discuss an example of biased machine learning results causing poor decisions?

I can think of two cases:
Filter bubbles: Suppose that an algorithm learns that you like a certain category of news simply because you happened to have clicked on a few popular items at some initial point, and then all news starts getting filtered through this narrow lens built by the model. If all the news you see happens to be visible to you because it passed through the algorithmic filter, and you therefore do not click on any alternative views, the algorithm will perceive your limitation in discovery as a narrow interest and will keep reinforcing its filter, hiding even more diverse options from your recommended items.

Unfair predictions: Suppose that an algorithm learns a predictive model for some risk scoring using data about people that includes certain demographic attributes. If the data itself hides some systemic societal biases, then the predictive model will simply learn and echo those biases. One example is a model to predict which individuals are likely to be indicted for consuming illegal drugs when people from certain ethnic backgrounds tend to be suspected, screened, arrested, and prosecuted at a higher rate.

Do users notice when extensive content filtering is occurring with feedback loops?

Most users are not aware that advanced algorithms act like gateways between them and the information they could potentially discover. Often users are missing out on information that could be discovered without even realising it. This is the biggest danger to discovery.

Are white box algorithms more difficult to create than black box ones?

White box models are easier to create. However they tend to be less powerful than black box models in making accurate predictions. This is one reason why black box models are popular.

Given the importance of machine learning algorithms, will there be standards and legislation around using 'fair' algorithms?

This has already started: the European Union recently passed a law requiring that algorithmic predictions that have an impact on humans must provide an explanation for the reasoning behind the prediction. The city of New York is also considering a bill that will assign a task force to monitor the fairness of predictive algorithms that influence decisions concerning people to try to prevent biased and unfair algorithms that discriminate.