Keynote Speaker

Honest Data: Predicting Social Patterns and Collective Behavior

Masoud Makrehchi^{a,*}

^aUniversity of Ontario Institute of Technology, Canada

*Corresponding Author: masoud.makrehchi@uoit.ca

Abstract

Filters play essential roles in modeling many engineering problems. A filter can be a transmission line, computer memory, a learning machine, or a compression algorithm. We spend a great deal of effort to design trustworthy filters. Trustworthy filters don't lie, betray the source, or mislead the destination. But these all are based on one strong assumption: source is trustworthy and honest. Almost in all models, the honesty of the source is out of context. In the case of natural and man-made sources such as sensors, engineers model and estimate sampling and measurement error. But dishonesty is different from error. In social context, error is an honest mistake but with good faith. Error is not because of bad intention and mostly caused by poor judgement. On the other hand, dishonesty has two major characteristics: insincere intention and lack of fairness. Unfortunately in most studies on social and human behavior which are based on data collection from questionnaires and surveys, these two variables are hard to measure. As a matter of fact, we humans demonstrate our true intentions and honest opinion when we implicitly express our views. It means we don't manipulate our thought for the sake of interests or to prevent any threat. But most lab experiments, surveys, and questionnaires are based on this fact that participants know they are under study.

In this talk, first the theory of Honest Data is informally discussed and then we show that Social Data is honest. This is where Social Computing and Human Computation come to play. Some research projects are also introduced. We demonstrate why Twitter data can be a good signal to predict the trend of crime rate or stock market. To deal with Twitter sampling limitation, we discuss an interesting direction in social computing: User-centric social analytics. In designing user-centric solutions, we rely on heavy active users who are expert in the domain of our interest. This brings us three questions: Who is expert? How to deal with sparsity of user activity? and how to make sure the users are credible and trustworthy? The major portion of the talk will be dedicated to answer these three questions.