Beyond Understanding and Prediction - Data Mining for Action

Jiuyong Li
University of South Australia
Adelaide, Australia
Jiuyong.li@unisa.edu.au

ABSTRACT

Association analysis and prediction are two major tasks in data mining, and they represent two foremost objectives: data exploration for understanding and model construction for prediction. Data mining is known as a process to convert raw data to useful information --- knowledge. However, what do we do with the knowledge discovered from data? We will need knowledge to enable actions, such as preventing diseases in health care, taking actions to retain customers, and etc.

One important attribute of knowledge is actionability. In order to be acted on, the knowledge must encode causal relationships to imply the mechanisms of the systems under consideration. Causal inference is a sophisticated topic that spans multiple disciplines, computer science, statistics, medicine, economics, and social science, to name a few. There are a number of well-established frameworks for causal inference in data with assumptions. Unfortunately, the assumptions are not directly testable in data, and hence there are limited off-the-shelf data mining methods for causal discovery. Many practitioners still use conventional machine learning methods for the tasks which actually requires causal inference [1].

It is desirable to have some simple data mining tools to explore causal relationships without or with few assumptions. The discoveries may not be proved causal, but are high quality candidates excluding many non-causal or spurious relationships. This will be a significant step forward in actionable data mining since it is well known that association rule mining generates too many spurious relationships and a classification model often provides a correct prediction based on non-interpretable (or even wrong) evidence. The spurious relationships and non-interpretable or wrong evidence hinder many data mining applications, especially in medical and social science areas.

We have been using some well-known causal inference principles to discover causal relationships and building causally interpretable data mining models, such as causal rules [2] and causal decision trees [3], and applying causal discovery methods to real world biological problems [4, 5]. In this talk, I will discuss some of our exploratory work in this promising direction.

General Terms

Algorithms; Experimentation; Theory.

WWW 2017 Companion, April 3-7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
http://dx.doi.org/10.1145/3041021.3053407

Keywords

Data Mining; Prediction and Action.

Keynote Speaker's Bio:

Dr Jiuyong Li is a Professor and an Associate Head of School at the School of Information Technology and Mathematical Sciences of University of South Australia. He leads the Data Analytics Group in the School. His main research interests are in data mining, bioinformatics, and data privacy. He has led six Australian Research Council Discovery projects and a Data to Decision CRC project. He has published more than 100 papers, mostly in leading journals and conferences in the areas. He has been a chair (or a PC chair) of multiple Australasian data mining and artificial intelligence conferences and actively serving PC (and senior PC) member for many international conferences in data mining. He has received senior visiting fellowships from Nokia Foundation, the Australian Academy of Science, and Japan Society of Promotion of Science.

ACKNOWLEDGMENTS

This research is supported by the ARC Discovery Project under grant No. DP140103617.

Reference:

[1] Susan Athey, Beyond prediction: using big data for policy problems, Science, 355(6324), 483-485, 2017.

[2] Jiuyong Li, Thuc Duy Le, Lin Liu, Jixue Liu, Zhou Jin, Bing-Yu Sun, and Saisai Ma, From observational studies to causal rule mining. ACM Transactions on Intelligent Systems and Technology, 7(2), 14:1-14:27, 2016.

[3] Jiuyong Li, Saisai Ma, Thuc Duy Le, Lin Liu, and Jixue Liu, Causal Decision Trees. IEEE Transsctions on Knowledge and Data Engineering, 29(2), 257-271, 2017.

[4] Thuc Duy Le, Lin Liu, Anna Tsykin, Gregory J. Goodall, Bing Liu, Bing-Yu Sun, and Jiuyong Li, Inferring microRNA-mRNA causal regulatory relationships from expression data. Bioinformatics 29(6):, 765-771, 2013.

[5] Weijia Zhang, Thuc Duy Le, Lin Liu, Zhi-Hua Zhou, and Jiuyong Li, Predicting miRNA targets by integrating gene regulatory knowledge with expression profiles, PLoS One, 11(4), 2016.