

Exploring the Power of Source Reliability in Information Integration

Houping Xiao, Ph.D. candidate, SUNY Buffalo State

Date: March 14, 2018, 1:00-2:00pm **Room:** JMH342

Abstract: In the era of Big Data, data entries, even describing the same objects or events, can come from a variety of sources. There are some sources that typically provide accurate information, but due to various reasons such as recording errors, device malfunction, background noise, or even intent to manipulate the data, some other sources may contain noisy or even erroneous information. Therefore, during information integration, it is critical to identify reliable sources that more often provide accurate information. Unfortunately, there is no oracle telling us which information source is more reliable *a priori*.

In this talk, I will present my work on the development of novel information integration methods that incorporate the estimation of source reliability. I explored the power of source reliability estimation in both data-level and model-level information integration. In data-level information integration, the objective is to jointly estimate which source is reliable and which piece of information is correct. To this end, I developed a new approach that iteratively calculates source reliability and true value estimates. In addition, the proposed approach outputs the confidence in the true value estimates, which provide useful information for decision making. In model-level information integration, I proposed to integrate models derived from a collection of distributed data sources. The proposed approach combines these models via weighted aggregation, in which weights reflect source reliability and are inferred jointly with the model combination procedure. An effective solution is designed through solving a constrained optimization problem by alternating direction method of multipliers. In both work, I proved some nice properties of the proposed approaches via theoretical analysis, and demonstrated their impact on some real applications such as indoor floorplan construction and crowdsourced question answering. I will also briefly discuss my other work on information integration, and their applications in health care, cyber security, and urban computing.

Bio: Mr. Houping Xiao is a Ph.D. candidate in the Department of Computer Science and Engineering, University at Buffalo, State University of New York. He received B.S. degree in Statistics from Beijing Normal University. His research interests are data mining and machine learning, with a focus on truth discovery, information integration, information trustworthiness analysis, privacy-preserving data mining, and educational data mining. In particular, his thesis work focuses on the development of novel information integration methods that incorporate the estimation of source reliability. His research emphasizes both theoretical performance and their applications in the real world. His publications appear in top conferences in both data science and other domains including KDD, WWW, SDM, MobiHoc and SenSys. He has internship and co-op experience in industry research labs, such as IBM T.J. Watson Research Center and Tencent AI Lab. More information can be found from his website: <http://www.acsu.buffalo.edu/~houpingx/>