WSDM'15 Workshop Summary / Scalable Data Analytics: Theory and Applications

Kaizhu Huang
Dept. of Electronic and Electric Engineering
Xi'an Jiaotong-Liverpool University
Ren'ai Road 111
Higher Education Town, SIP
Suzhou, Jiangsu, China
kaizhu.huang@xjtlu.edu.cn

Haiqin Yang[1,2], Irwin King[1,2], Michael R. Lyu[1,2]
[1] Shenzhen Key Laboratory of Rich Media Big
Data Analytics and Application
SZRI, Chinese University of Hong Kong
[2] Dept. of Computer Science & Engineering
Chinese University of Hong Kong, Hong Kong
{hqyang,king, lyu}@cse.cuhk.edu.hk

ABSTRACT

The SDA workshop at WSDM 2015 is the fifth International Workshop on Scalable Data Analytics, following the previous four workshops of SDA respectively held at IEEE Big Data 2013, PAKDD 2014, IEEE Big Data 2014, and IEEE ICDM 2014. This series of workshops aims to provide professionals, researchers, and technologists with a single forum where they can discuss and share the state-of-the-art theories and applications of scalable data analytics technologies. In particular, in the era of information explosion, the scientific, biomedical, and engineering research communities are undergoing a profound transformation where discoveries and innovations increasingly rely on massive amounts of data. The characteristics of volume, velocity, variety and veracity originated in the massive big data then bring challenges to current data analytics techniques. The focus of the fifth SDA is to discuss how we can scale up data analytics techniques for modeling and analyzing big data from various domains.

Categories and Subject Descriptors

I.2.6 [**ARTIFICIAL INTELLIGENCE**]: Learning—*Knowledge acquisition* ; H.m [**Information Systems**]: MISCELLANEOUS

Keywords

Big Data, Scalable, Data Analytics

1. INTRODUCTION

With the fast evolving technology for data collection, data transmission, and data analysis, the scientific, biomedical, and engineering research communities are undergoing a profound transformation where discoveries and innovations increasingly rely on massive amounts of data. New prediction

techniques, including novel statistical, mathematical, and modeling techniques are enabling a paradigm shift in scientific and biomedical investigation. Data become the fourth pillar of science and engineering, offering complementary insights in addition to theory, experiments, and computer simulation. Advances in machine learning, data mining, and visualization are enabling new ways of extracting useful information from massive data sets. The characteristics of volume, velocity, variety and veracity bring challenges to current data analytics techniques. It is desirable to scale up data analytics techniques for modeling and analyzing big data from various domains.

This series of workshop features the following topics of interest, but not limited to:

- Distributed data analytics architectures
- Theory and algorithms for scalable descriptive statistical modeling
- Theory and algorithms of scalable predictive statistical modeling
- Scalable analytics techniques for temporal and spatial data
- Scalable data analytics algorithms in large graphs
- Novel applications of scalable machine learning

Following the previous four workshops, the fifth International Workshop on SDA aims to provide professionals, researchers, and technologists with a single forum where they can discuss and share the state-of-the-art theories and applications of scalable data analytics technologies. The specific focus of the fifth SDA will be described in the next section.

2. WORKSHOP FOCUS

This fifth International Workshop on SDA focuses on scalable extensions of many famous machine learning algorithms as well as their interesting applications in big data. In particular, the workshop discusses distributed and scalable versions of metric learning algorithms, scalable and discriminative feature generation algorithms, scalable text clustering and document representation, fast and multilevel coclustering algorithms, scalable non-negative matrix factorization, and scalable point-of-interest recommendation. Applications of these scalable extensions in social networks,

document classification, pattern recognition are also the focus of this year's workshop. Expert researchers in the above mentioned fields submitted position papers in the fifth SDA workshop and they exchange ideas specifically in the following aspects:

- Retweet scale prediction model

- Efficient influence maximization algorithm in social networks

- Scalable text clustering and fast multilevel co-clustering for social network data

- Fast and hybrid deep belief network for document representation

- Distributed non-negative matrix factorization

- Accelerated information-theoretic metric learning

- Heterogeneous information network based ranking and clustering of mobile apps

- Scalable and discriminative random nonlinear features from data

- Scalable point-of-interest recommendation for social networks

- Other fundamental and theoretic analysis in big data

3. CONCLUSION

Following the previous successful four SDA workshops respectively at IEEE Big Data 2013, PAKDD 2014, IEEE Big Data 2014, and IEEE ICDM 2014, the fifth International Workshop on Scalable Data Analytics aims to provide professionals, researchers, and technologists with a single forum where they can discuss and share the state-of-the-art theories and applications of scalable data analytics technologies. More specifically, this workshop accepts position papers that investigate scalable extensions of many famous machine learning algorithms (e.g., scalable metric learning, scalable non-negative matrix factorization, scalable feature extraction, and scalable and fast text clustering algorithms) as well as their interesting applications in massive data (e.g. social networks, document classification, and large-scale pattern recognition). New research trends in Scalable Data Analytics will be also extensively discussed in the workshop.

Acknowledgments