Applied Stochastic Models in Business and Industry

(wileyonlinelibrary.com) DOI: 10.1002/asmb.2302 Published online 11 January 2018 in Wiley Online Library

Special issue on Data Fusion

Rapid developments of technologies and the availability of a large amount data from different sources have presented many new challenges and opportunities for statistics and data science. Effective approaches that can synthesize and combine information are in high demand. *Fusion inference*, sometimes called *fusion learning*, refers to synthesizing or fusing statistical inferences from multiple data sources, which can often yield more powerful findings than those obtained from individual sources alone. Meta-analysis is a well-known example of fusion inference. Fusion inference is of vital importance in light of the increasing trend to collect data from heterogeneous sources. As new challenges continue to arise from big data contexts, we anticipate continued interest in data fusion research.

In this special issue, we have four papers that discuss fusion inference and illustrate complexities that arise when combining multiple data sources. The paper by Grelaud, Mitra, Xie and Chen develops a fusion learning model to perform real-time detection of nuclear sources. The authors discuss how the model could be implemented within a network of taxicabs to provide surveillance monitoring for a city. The paper by Myhre, Jeske, Li and Hansen develops a methodology for combining pass/fail reliability test data that comes from two possibly different test environments. An attractively simple method is proposed and shown to do very well relative to what appear to be more sophisticated approaches. The paper by Hoegh and Leman provides a useful model fusion method to account for potential associations between individual models by using a Gaussian graphical model to explicitly characterize the associations between the models. It specifically provides a fusion technique for modular prediction with complex, unstructured data motivated by predicting civil unrest in Central and South America. The paper by Casleton, Osthus and Van Buren describes application of several different imputation methods for multi-source data. It provides a useful comparative study for integrating heterogeneous data sources in presence of missing data.

Daniel R. Jeske
Department of Statistics, University of California, Riverside California, USA
Email: *daniel.jeske@ucr.edu*

Min-ge Xie
Department of Statistics and Biostatistics, Rutgers University, Piscataway, New Jersey, USA
Email: *mxie@stat.rutgers.edu*