Introduction to Data, Text, and Web Mining for Business Analytics Mini-track

Dursun Delen
Oklahoma State University
dursun.delen@okstate.edu

This mini-track has eight papers that are about developing analytic systems for decision support by means of data, text, or web mining. Five of the eight papers focus on an increasingly more popular area of analytics (i.e., text mining, text analytics, and sentiment analysis). While some of these papers are aimed at improving the text mining methods others are focused on novel use of text mining for solving interesting and challenging business problems. The remaining three papers in this mini-track deals with search engine enhancement, document classification, and firm failure timeline prediction.

The paper by Pröllochs et al. is about detecting negation scope for financial news sentiments. Their proposed method tries to replicate human-like learning and thus appears to be a better suited method for natural language processing. Their results provide evidence that the proposed reinforcement learning can outperform rule-based approaches, revealing a prediction accuracy of up to 76.37% on a manually-labeled dataset.

The paper by Lee et al. is about augmenting text mining approaches with social network analysis to understand the complex relationships between users’ requests, which is presented as a case study of the android operating system. Their findings suggest that text mining approaches may be supplemented with other methods like network analysis to deliver better insights. Similarly, the paper by Wang et al. proposes a unified framework for fine-grained opinion mining from online reviews. Their experimental results on both English and Chinese online reviews demonstrate the effectiveness of proposed framework compared to the existing methods.

The paper by Jin et al. presents yet another novel application of analytics to stock markets. Considering the mature research for the network characteristics of the stock market, in this paper they attempt to employ a complex networks analysis method to study the risk of the stock market. Their experimental results show that price fluctuations of a small number of stocks may impact the other stocks, even the entire stock market. These influential stocks exists in almost all industries, but are mainly in finance, manufacturing and mining.

The paper by Soper and Turel is about analyzing and comparing European and North American IS research articles using a concept frequency analysis method. Therein, employing a text mining approach to the leading publications in IS field from North America and Europe (MISQ and ISR vs. EJIS and JIT from 1991 it 2013) they are seeking to identify possible convergence and/or divergence trends between the corpuses of the two continents.

The paper by Ye et al. is about enhancing retrieval and ranking performance for a media search engine by deep learning. In this paper, they propose a framework that utilizes a Deep Structured Semantic Model (DSSM) to build similarity features to enhance search engine relevance. To illustrate the effectiveness of the proposed deep learned similarity features, they applied their method to an Xbox Media Search Engine.

The last two papers of the mini-track focused on prediction type analytics problems. The paper by Timsina et al. is about using a semi-supervised learning for the creation of medical systematic review, where they explore semi-supervised learning based classifiers to identify articles that can be included when creating medical systematic reviews. The paper by Ryu is about firm failure timeline prediction using a math programming approaches. In this study Ryu applied two math programming methods to predict firm failure timeline: isotonic prediction and modified linear regression. With these two methods, Ryu also evaluated financial ratios that were previously used for fixed-term bankruptcy prediction.