Introduction to Data, Text, and Web Mining
for Business Analytics Minitrack

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

Hamed Zolbanin
Ball State University
hmzolbanin@bsu.edu

This mini-track has four papers that are about developing analytic systems for decision support by means of data, text, or web mining. One of the four papers focuses on text mining; two papers attend to developing predictive models to answer interesting and challenging business problems; and one focuses on extracting semantic relationships among entities in big document corpora.

The paper by Eickhoff and Wieneke develops a mixed-method approach for the analysis of large document collections. The proposed approach improves the drawbacks of the quantitative topic models, such as Latent Dirichlet Allocation (LDA), and provides context and meaning to topic models. The paper illustrates the proposed approach by conducting a case study on the thematic composition of the AIS Senior Scholars' Basket of Journals.

The paper by Freeze *et al.* develops a system to collect and integrate comprehensive datasets from prior outbreaks of diseases to facilitate multivariate, predictive analysis of disease patterns, intensity, and timing. The proposed system uses machine learning algorithms to predict the 1-week and 4-week forecasts of Dengue incidences in two cities in central and south America. The paper presents the prediction results along with the features of the proposed system.

The paper by Sifa et al. provides an interesting example for the value of predictive analytics to marketing initiatives. The paper uses a combination of data processing and predictive analytics to identify, retain, and upsell high-valued customers in non-contractual freemium settings. This is especially valuable for businesses where a small share of users often drives the largest part of revenue for firms and co-finances the free provision of the product or service to a large number of users.

Finally, the paper by Schmidt and Scholz uses a novel approach to develop a quantitative understanding of the relationship between multiple entities in a big corpus of documents. The paper points out that for a context-sensitive entity-based search engine, the traditional approach of identifying relatedness on the basis of appearance in one document is not feasible. The authors, instead, use a smaller unit of sliding window, which is shifted over the text, to determine if entities are associated with one another. The result is a context-sensitive auto-suggestion system that is based on some offline batch processing and some real-time computation at the time of query submission.

HĪCSS