

Introduction to Data Analytics and Data Mining for Social Media Minitrack

Dominique M. Haughton
Bentley University, USA
dhaughton@bentley.edu

Jennifer J. Xu
Bentley University, USA
jxu@bentley.edu

David J. Yates
Bentley University, USA
dyates@bentley.edu

Xiangbin Yan
Harbin Inst. of Technology, China
xbyan@hit.edu.cn

As the HICSS conference turns 50, this minitrack turns five. The papers at HICSS in 2017 remind our attendees and readers of the many possible real-world applications of data analytics and data mining for social media. Last year, for example, we explored critical success factors for online petitions for social causes, and also analyzed brand sentiment analysis for Twitter using neural networks. In 2015, our papers showed how to identify uptake, sessions, and key actors in socio-technical networks as well as how to measure NBA players' mood by mining athlete-generated content.

This minitrack begins by examining the relationship between social media engagement and financial performance of the global fast fashion company, H&M. Ravi Vatrapsu and his Copenhagen Business School team find, measure and deconstruct the positive relationship between social buzz generated by style icons like David Beckham and company sales.

The second paper describes the Lariat system, developed at the University of Washington. Lariat is a visual analytics tool that facilitates exploratory data analysis through integrated grouping and visualization of social media data. Its design is informed by the results of formative studies and sensemaking theory, which together indicate that exploratory processes for social media require search, comparison, verification, and iterative refinement.

In the third paper, "A domain oriented LDA model for mining product defects from online customer reviews," a Virginia Tech team defines key attributes, such as components and keywords used to describe a defect, and builds a novel latent Dirichlet allocation (LDA) model to identify and acquire integral information about defects. Their case studies include defects reported in vehicles tracked by the National Highway Traffic Safety Administration.

"Exploring time series spectral features in viral hashtags prediction" utilizes frequency domain tools to analyze time series such as the number and type of hyperlinks in a particular Tweet stream. Through experiments with two sets of tweets, Doong and Chung show that wavelet spectral features can improve prediction performance, and using a discrete

fast wavelet transform (FWT) yields results as good as more complicated methods.

The fifth paper, "Birds of a feather talk together: user influence on language adoption," is from a Lancaster University and Loughborough University collaboration. These authors examine temporal and spatial trends in out-of-vocabulary (OOV) adoption prediction within the Reddit and Twitter social networks. Their results show that the time at which a user adopts a term is dependent on the local structure, however, a large part of the influence comes from the global structure and that influence between users and communities is not significantly dependent on network structures.

"A large-scale analysis of the marketplace characteristics in Fiverr" studies the Fiverr crowdservices platform as a unique marketplace and characterize the sellers, buyers, and the interactions among them. An analysis of the services of the top sellers indicates that the more diverse type of products a seller sells, the more sales they can make. This IIT Kharagpur team also presents interesting findings with respect to service competition as well as service differentiation.

The seventh paper, "A peer-based approach on analyzing hacked Twitter accounts," is by Murauer, Zangerle and Specht. This team analyzes the possibilities of including the reactions of hacked Twitter accounts' peers into a detection system. An empirical analysis of conversations embedded in six million tweets shows that 30% of the users that are allegedly being hacked reply to accusations, often within an hour, suggesting that these users acknowledged that their account was hacked.

Our final paper presents and examines interactive visualizations of political polarization found on Twitter during the 2014 Brazilian presidential campaigns. This paper is from a research team at the Pontifical Catholic University of Rio Grande do Sul. Their study identifies features, benefits as well as understanding gaps of data visualizations that might be used by journalists and media professionals covering political issues and events. They conclude that for data analysis and visualization to be more effective in such applications, further research is needed.