Introduction to the Minitrack on Data Analytics, Data Mining, and Machine Learning for Social Media

Jeffry S. Babb
West Texas A&M University
jbabb@wtamu.edu

Kevin D. Mentzer
Nichols College
kevin.mentzer@nichols.edu

David J. Yates
Bentley University
dyates@bentley.edu

Abstract

Thank you for your interest in the Data Analytics, Data Mining, and Machine Learning for Social Media Minitrack. This year's paper contributions vary in scope and focus and thus provide a rich spectral view of developing insights and understanding from social media. According to Monique Thomas of Later, social media is experiencing a significant evolution: "We've said goodbye to simple status updates (we miss you!) and hello to social selling and growing digital communities." However, our digital communities continue to struggle with the challenge of being free and open, allowing anyone to express their thoughts and ideas, while at the same time working to combat misinformation and deep fakes. We are grateful to social media scholars who continue on various paths of sensemaking to better understand social media's impact on our society. This year's papers provide many new insights into this vital area of research.

1. Papers at a glance

Knowledge development from analytics will rarely exceed both the quality of the data and the quality and tuning of the models used. Lo, Lee, and Zhang (2023) explore this in detail in the paper *Is a Pretrained Model the Answer to Situational Awareness Detection on Social Media?* In evaluating the construct of situational awareness for first responders, the utility of pre-trained models is examined, and domain-specific vocabularies are emphasized in the findings. Often, the relationship between social media and product research and development is not accidental. Further advances in natural language processing are needed to develop useful summarization of dialog information. Zheng and Saga's (2023) paper, *A Lite Hierarchical Model for Dialogue Summarization with Multi-Granularity Decoder,* examines shortcomings in extant dialog summarization techniques by adjusting for the granularity of model inputs using a novel, hierarchical approach.

The application of machine learning models often extends into assisting critical product, marketing, and investment decisions. Training models to determine whether customer feedback via social media is informative is referred to as needmining. Stahlmann et al. (2023) approach the challenge of making useful comparisons between needmining models in their paper *What Do Customers Say About My Products? Benchmarking Machine Learning Models for Need Identification*. Kilroy, Caton, and Healy (2023) examine needmining in their paper entitled *The Trending Customer Needs (TCN) Dataset: A Benchmarking and Automated Evaluation Approach for New Product Development*. This paper takes a step back to broadly assess highly used key phrases in the consumer packaged goods market segment. A curated dataset of customer needs, expressed as key phrases, is applied in an online community case study.

Several papers examine public and behavioral impacts of machine learning and artificial intelligence. Noor and Turetken (2023) seek answers to the question *What Drives Sentiments on Social Media? An Exploratory Study on the 2021 Canadian Federal Election* in their paper so titled. This paper describes how sentiment analysis of nearly 800,000 tweets from recent elections in Canada produced text-clustering outcomes and results indicating the importance of message timing and negative sentiment. Cartwright et al. (2023) explore the origins and diffusion of COVID-19 misinformation in their paper entitled *Deploying Artificial Intelligence to Combat Covid-19 Misinformation on Social Media: Technological and Ethical Considerations*. The paper outlines an ongoing effort to provide tools to assist social media platforms, online service providers, and government agencies in identifying and responding to misinformation on social media and the balance to be struck versus privacy and freedom of speech considerations. Related to balancing, Pilgrim, Koss, and Bohnet-Joschko (2023) examine

HICSS

corporate social responsibility as it pertains to social media mining in their paper *CSR Communication on Twitter - A Scoping Review on Social Media Mining and Analytic Methods*. The paper is a review designed to identify critical factors using a systematic review of social media mining practices. Some actors in the public space are neither governmental nor corporate, or even formal organizations. Buz and de Melo (2023) examine the impact of WallStreeBets, an informal online community on Reddit, in their paper *WallStreetBets: An Analysis of Investment Advice Democratization*. As a potential extension of the crowdsourcing phenomenon, the wisdom of crowds is examined to determine whether an online community can analyze and predict financial markets.

Invariably, social media will also shape opinions and attitudes as they relate to everyday products and services. Often, this takes the form of user-to-user reviews. Dorwat, Namvar, and Akhlaghpour (2023) consider the relationship between review depth, features, and helpfulness in their paper *Revisiting Review Depth in Search for Helpful Online Reviews*. The paper draws conclusions regarding readability and content using an Amazon dataset of hundreds of thousands of reviews and reviewers. The importance of online reviews extends into healthcare services, where Hao et al. (2023) explore doctor reviews in China in their paper titled *What Can Online Doctor Reviews Tell Us? A Deep Learning Assisted Study of Telehealth Service*. This paper identifies model results that link review sentiment to service utilization. Given the volume of input to social media that arises from the ubiquitous and mobile nature of computing, recommendation models are commonly used in a geospatial context: "what is good to eat around here?" Xie et al. (2023) considers the implications of points of interest (POI) locators for geospatial service searches in their paper *Visualization of POI Category on the Dynamic Rasterized Map Tiles from Geo-Tagged Social Media (Twitter) with SZ-GAT*. A model for prioritizing POI placement and emphasis based on tweet data is explored using deep learning and data mining.

References

Buz, T., & de Melo, G. (2023). WallStreetBets: An Analysis of Investment Advice Democratization. In *Proc. of the Hawaii International Conference on System Sciences (HICSS)*, Lahaina, HI.

Cartwright, B., Frank, R., Weir, G., Padda, K., & Strange, S.-M. (2023). Deploying Artificial Intelligence to Combat Covid-19 Misinformation on Social Media: Technological and Ethical Considerations. In *Proc. of the HICSS*, Lahaina, HI.

Dorwat, S., Namvar, M., & Akhlaghpour, S. (2023). Revisiting Review Depth in Search for Helpful Online Reviews. In *Proc. of the HICSS*, Lahaina, HI.

Hao, H., Zhang, B., Zhan, Y., & Wu, J. (2023). What Can Online Doctor Reviews Tell Us? A Deep Learning Assisted Study of Telehealth Service. In *Proc. of the HICSS*, Lahaina, HI.

Kilroy, D., Caton, S., & Healy, G. (2023). The Trending Customer Needs (TCN) Dataset: A Benchmarking and Automated Evaluation Approach for New Product Development. In *Proc. of the HICSS*, Lahaina, HI.

Lo, S. L., Lee, K., & Zhang, Y. (2023). Is a Pretrained Model the Answer to Situational Awareness Detection on Social Media? In *Proc. of the HICSS*, Lahaina, HI.

Noor, H. M., & Turetken, O. (2023). What Drives Sentiments on Social Media? An Exploratory Study on the 2021 Canadian Federal Election. In *Proc. of the HICSS*, Lahaina, HI.

Pilgrim, K., Koss, J. & Bohnet-Joschko, S. (2023). CSR Communication on Twitter - A Scoping Review on Social Media Mining and Analytic Methods. In *Proc. of the HICSS*, Lahaina, HI.

Stahlmann, S., Ettrich, O., Kurka, M., & Schoder, D. (2023). What Do Customers Say About My Products? Benchmarking Machine Learning Models for Need Identification. In *Proc. of the HICSS*, Lahaina, HI.

Xie, H., Wang, Y., Li, D., & Kawai, Y. (2023). Visualization of POI Category on the Dynamic Rasterized Map Tiles from Geo-Tagged Social Media (Twitter) with SZ-GAT. In *Proc. of the HICSS*, Lahaina, HI.

Zheng, T., & Saga, R. (2023). A Lite Hierarchical Model for Dialogue Summarization with Multi-Granularity Decoder. In *Proc. of the HICSS*, Lahaina, HI.