

Introduction to the HICSS-51 Data Science and Digital Collaborations Minitrack

Lakshmi S. Iyer
 Appalachian State University
 Boone, NC 28608
iyerLs@appstate.edu

Souren Paul
 Nova Southeastern University
 Fort Lauderdale, FL 33314
Souren.paul@gmail.com

Lina Zhou
 University of North Carolina at
 Charlotte
 Charlotte, NC 28223
lzhou8@uncc.edu

Data science and analytics for collaboration is the study of generalizable extraction of knowledge from structured and/or unstructured data to support collaboration within and across groups and organizations. The new actionable knowledge gained should support achieving collaborative goals such as innovation, idea generation, decision making, negotiation, and execution. In addition, it requires deep understanding of formulating problems valuable for collaboration, engineering effective solutions to collaboration problems, and effectively communicate findings across roles ranging from business managers to data analysts. One promising way for businesses and organizations to enhance their performance or competitiveness is by investigating how data science and analytics can facilitate collaboration both internally and externally. For example, businesses are trying to understand how data science and analytics can help engage customers and improve operation efficiency and how it can use social media to support corporate knowledge management. Another example is collaborative creation of ideas and solutions through crowdsourcing and online communities (such as dominodatalab.com).

This minitrack includes one paper session, consisting of the following three papers covering the above topics of interest. The first paper, “Developing a User Typology for the Analysis of Participation in Enterprise Collaboration Systems”, aims to propose a user typology for Enterprise Collaboration Systems (ECS). The typology includes the ECS user types, the dimensions of ECS use and the collaborative action types in ECS. Five dominant categories of ECS users identified in the paper are creator, contributor, lurker, inactive and non-user. The user types are characterized by differences along some dimensions, such as type of use, frequency of use, variety of use, choice of content type and platform preferences. The user typology was implemented as SQL scripts to enable the analysis of live data from the ECS UniConnect, which is a collaboration platform based on IBM Collaboration suite. It is provided for universities with more than 3500 user bases. The transactional data in the event log of

UniConnect was analyzed to identify dominant types of users and their activities in the ECS. Analyses were conducted at platform level and community level and by content types. The majority of users are lurkers at platform level and for different types of content. The results for the analysis at the community level show that the distribution of user types depends on the type of community. The findings have practical implications for platform owners and community managers.

The second paper, “Divergence Based Non-Negative Matrix Factorization for top-N Recommendations”, proposes a personalized top-N recommendation using Matrix Factorization (TNMF) method based on non-negative matrix factorization using K-L divergence as a point-wise ranking loss function. In the method, the learning problem is formulated with regularized divergence as a constrained non-convex minimization problem and the divergence problem is addressed by developing a projected stochastic gradient descent optimization algorithm. The results of empirical evaluations with six datasets demonstrate that TNMF outperform state-of-the-art methods in terms of RMSE on most of the test datasets. In addition, latent factors learnt from existing data contribute to improved hit-rate of recommendation predictions.

The third paper, “I Understand What You Are Saying: Leveraging Deep Learning Techniques for Aspect Based Sentiment Analysis”, investigates the dynamics of aspect based sentiment in online reviews by proposing a semi-supervised, deep learning facilitated analytical pipeline. Previous studies of online reviews are dominated by methods that treat aspect extraction and sentiment analysis as two separate issues. This study not only addresses the two issues simultaneously, but also introduces a novel perspective for understanding sentiment dynamics by analyzing aspect based sentiment in time series. The methods and results of this research can be used to improve product/service offerings and customer relationship management by facilitate communications between consumers, product/service providers and online retailers.