

Introduction to the HICSS-56 Data Science for Digital Collaboration Minitrack

Lina Zhou
Business Information Systems
and Operations Management
The University of North Carolina
at Charlotte
Charlotte, NC 28223
lzhou8@uncc.edu

Souren Paul
College of Informatics
Northern Kentucky University
Highland Heights, Kentucky
Souren.paul@gmail.com

Florian Schwade
University of Koblenz-Landau
Faculty of Computer Science
Institute for IS Research
Universitaetsstrasse 1
D-56070 Koblenz
fschwade@uni-koblenz.de

The evolution of digital collaboration technologies over the past decades has given rise to a new application domain of data science. These emerging technologies range from business collaboration software such as enterprise collaboration systems and social networking tools, distributed ledger technology, and Internet of Things, to public collaboration platforms such as social media, crowdsourcing platforms, and open source communities (e.g., GitHub). Businesses and organizations can use these systems for collaboration in knowledge management, supply chain management, and customer relationship management. The public can use them for collaboration, information sharing, community support, and social goods. Applying data science in these domains allows for leveraging the digital trace data to gain actionable knowledge and insights on how people collaborate and how these platforms facilitate digital collaboration, which can lead to improvements in system design as well as collaboration processes for problem solving and decision making. Moreover, collaboration technologies can facilitate collaborative and distributed data science projects. This minitrack aims to address the various issues, opportunities and challenges of applying data science for digital collaboration.

Data science for digital collaboration can be addressed from a variety of perspectives. For instance, social analytics is used to generate insights into the topics, content and relationships that emerge through digital collaboration. Process mining is applied to discover sequences and interaction patterns of digital collaboration processes. Among other challenges associated with data science for digital collaboration, there are challenges regarding ethical and legal issues in corporate context. There has been a continued interest in finding ways to enable and strengthen digital collaboration by discovering knowledge and patterns and gaining actionable insights. This minitrack includes one paper session, consisting of two papers covering the following areas of interest: social analytics in digital collaboration, collaborative collection and management

of big data, collaborative analysis of big data, human-guided knowledge discovery, security and privacy analytics in digital collaboration, and ethical and legal issues in collaborative data science.

The first paper, “Federated Learning for Credit Risk Assessment”, proposes a prototype for credit risk assessment based on federated learning. As the underlying machine learning models are trained collaboratively, the prototype allows smaller financial institutions to compete. The prototype is evaluated using a historical mortgage dataset and benefits of federated learning for smaller financial institutions are demonstrated. In the evaluation, the authors find that their prototype improved accuracy, recall, precision and F1 scores of the machine learning models. The study findings contribute to existing literature on federated learning in financial services and demonstrates the use and benefits of federated learning in small financial institutions.

The second paper, “On Left and Right: Understanding the Discourse of Presidential Election in Social Media Communities”, develops an unsupervised learning approach EZClass to extracting political parties and detecting political stances from social media discourse. Drawing on the state-of-the-art zero-shot learning techniques, the proposed method addresses the limited availability of labelled data that challenges existing supervised learning methods for stance detection. Choosing the 2020 US presidential election as a case study, it collects relevant data from Reddit communities and uses the data to evaluate the proposed method. The analysis results reveal the political stance and party characteristics of the (sub)communities and identify post-election shift in political stances of the communities. Their findings further detect within-community variations with respect to political parties and stances. The findings of the study have practical implications for social media platforms, community moderators and members, and political stakeholders.