

Introduction to the Data Science for Collaboration Minitrack

Lakshmi S. Iyer
The University of North
Carolina Greensboro
Lsiyer@uncg.edu

Souren Paul
Nova Southeastern University
Souren.paul@gmail.com

Lina Zhou
University of Maryland
Baltimore County
zhoul@umbc.edu

With explosive growth in unstructured and structured data, organizations are looking for ways to innovate through the use of data science and analytics. Data science for collaboration is the study of the generalizable extraction of knowledge from data to support human collaboration within and across groups and organizations. The new knowledge gained is expected to be actionable for achieving collaborative goals such as generating, choosing, negotiating, and executing. Data Science for collaboration requires a deep understanding of formulating problems valuable for collaboration, engineer effective solutions to the collaboration problems, and ways to effectively communicate findings across roles ranging from business managers to data analysts. The emerging heterogeneous, voluminous, and unverified data present both opportunities and new challenges for addressing collaboration problems. There is increasing interest in organizations looking for ways to increase value from data science and using it to address business challenges. This minitrack fills the void in research that focuses on data science for collaboration.

The second offering of the minitrack includes one paper session, consisting of three papers covering the following topics of interest: collaboration for success in crowdsourced innovation, social collaboration analytics, decision support for collaborative software performance engineering. The first paper, “Collaboration for Success in Crowdsourced Innovation Projects: Knowledge Creation, Team Diversity, and Tacit Coordination”, uses organizational knowledge creation theory (OKCT) and previous findings on crowdsourced innovation projects (CIPs) to examine and explain the impacts of team diversity, team coordination, and task complexity on task performance in CIPs. To test the proposed research hypotheses, the researchers employ both performance data and discussion forum data collected from a real-world website. The analysis shows that team coordination is likely a factor for success but collaboration was not a requirement in the test environment but competing as a team is likely a better choice for teams with medium level of performance.

The second paper, “Social Collaboration Analytics for Enterprise Collaboration Systems: Providing Business Intelligence on Collaboration Activities”, explores an emerging type of organizational collaboration software: Enterprise Social Software, and develop and evaluate an approach to social collaboration analytics (SCA). Following the guide of CRISP-DM, the authors develop a SCA framework that consists of six elements: level of analysis, content type, content component, action type, time, and further filters. The framework can assist researchers and managers in understanding and analyzing user activity and interaction in digital workplaces (communities).

The third paper entitled “Collaborative Software Performance Engineering for Enterprise Applications”, introduces a design science approach to develop a decision support for collaborative software performance engineering activities. They propose the design of a knowledgebase that integrates the activities of application performance monitoring (APM) and software performance engineering (SPE) for enterprise applications having global implementations across different organizations. A prototype of the knowledge base was developed and evaluated by the authors. The data from the monitoring logs of different implementations of standard EA were collected and organized in a database and several machine learning approaches were applied of the data to develop different prediction models. Models based on random forest machine learning approach produced low errors and were considered for further evaluation in this paper. For this model type, the mean relative error was 23.19% across the different standard transactions. The authors show how trained models can be used to predict the performance of enterprise applications. This is a cost efficient approach for new implementations of enterprise applications. The artifact proposed in the paper is simple and ignores many complexities, such as network congestions and faults in the applications which may be pursued in a future research. These papers will help to stimulate thinking about future research agenda in data science for collaboration.