University of Fribourg / Faculty of Sciences and Medicine / Department of Informatics

Human-AI Collaborative Approaches for Open-Ended Data Curation

Ines Arous

Collecting answers to open-ended questions is essential for many applications ranging from business through influencer finding to science through document mining. Many efforts have been dedicated to automatically curating such data through machine learning (ML) models. However, these methods have a limited performance since responding to open-ended questions requires intuition, domain knowledge, and reasoning abilities, which are still missing in state-of-the-art ML methods. Moreover, those models require large-scale and high-quality annotations, whose collection is a long and laborious process. Crowdsourcing provides a cost-effective way to answer open-ended questions in a short amount of time. Furthermore, human annotators on crowdsourcing platforms have diverse skill sets, which allow them to collect diverse results. Nonetheless, challenges arise since the answers supplied by workers can be prone to errors. Therefore, it is crucial to design solutions to ensure the quality of open-ended crowdsourced data.

This thesis proposes human-AI collaborative approaches to curate —collect, and clean— open-ended data. Overall, our frameworks comprehend human computation and AI model components that interact with each other. In the human computation component, we model workers' performance and optimize their involvement in open-ended data collection to minimize the cost and maximize the data quality. While in the AI model component, we leverage the task's and answers' features in addition to the worker model to learn the quality of their answers. The human computation component and the AI model are updated iteratively, allowing their learning processes to benefit from each other until an agreement on the quality of the answers is reached. Thus, the interaction between the human computation component and the AI model is bidirectional, which is fundamental to ensuring the effectiveness of the human-AI team.

We consider the methods introduced in this thesis as a step towards better collaboration between machine learning and human computation. Our work establishes principled optimization algorithms that allow the machine learning model's parameters to be updated using worker's modeling and vice-versa. The developed methodologies can be used as a foundation to build more interpretable methods and provide explanations for both their learning process and results. We envision that crowdsourced open-ended data curation can establish a new research direction to solve complex cognitive tasks and allow workers to "learn, not just earn" through these tasks.

Jury:
Prof. Dr. Rolf Ingold, University of Fribourg (Switzerland), president of the jury.
Prof. Dr. Philippe Cudré-Mauroux, University of Fribourg (Switzerland), thesis supervisor.
Dr. Mourad Khayati, University of Fribourg (Switzerland), thesis co-supervisor.
Prof. Dr. Jie Yang, University of TU Delft (Netherlands), thesis co-supervisor.
Prof. Dr. Jiliang Tang, Michigan State University (United States), external examiner.