Jiawei Tang (Research Portfolio)

Research Internships

— **Massachusetts Institute of Technology (MIT)**, USA (2022/06-08)

Mentor Professor Samuel Madden

Project Use deep learning models to solve the problem of entity resolution. Entity resolution is the task of deciding whether two data records refer to the same real-world object. It has diversified application domains such as banking, insurance, e-commerce, health care, and many others. For example, an e-commerce company wants to know if two products from different suppliers are the same so they can be displayed on the same product page; two banks sharing data need to identify and reconcile common customers.

Contributions I was responsible for system design, implementation, and testing for two tasks: determining the accuracy of foundation models for entity resolution and designing a deep-learning model for generic entity resolution.

Publication **First author** of paper "Generic Entity Resolution Models" accepted by Table Representation Learning Workshop @ NeurIPS 2022, where NeurIPS is one of the most prestigious and competitive international conferences in machine learning and computational neuroscience.

Link https://openreview.net/pdf?id=tRkVo1jMas

— **Qatar Computing Research Institute**, Qatar (2021/06-08)

Mentor Dr. Mourad Ouzzani

Project Build an end-to-end data visualization system that acts as a virtual assistant to allow novices to create visualizations through either natural language or speech.

Contributions Designed and implemented two main components: Speech-to-Text which is based on Google Cloud Speech-to-Text Rest API, and Text-to-VIS, which uses an end-to-end neural machine translation model.

Publication **First author** of paper "Sevi: Speech-to-Visualization through Neural Machine Translation" accepted by ACM SIGMOD International Conference on Management of Data, where SIGMOD is a leading international forum for database researchers. I presented and demonstrated this work in SIGMOD 2022 @Philadelphia.

Link https://dl.acm.org/doi/pdf/10.1145/3514221.3520150

— **Tsinghua University**, China (2020/06-08)

Mentor Professor Guoliang Li

Project Construct a benchmark of (natural language, data visualization) pairs and use this benchmark to train a deep learning model that translate a natural language query into a data visualization.

Contributions Used Python toolkits to clean and annotate data. Used PyTorch and Transformer models to train a deep learning model to support the translation from natural language queries to data visualizations.

Publication **Co-author** of paper "Natural Language to Visualization by Neural Machine Translation" accepted by IEEE Transactions on Visualization and Computer Graphics 2021, a top journal for data visualization.

Link https://ieeexplore.ieee.org/document/9617561

Note: Please see the following three pages for the abstracts of the above three papers.