North Carolina Agricultural and Technical State University

Doctoral Dissertation Defense

Yang Liu

Candidate for Doctor of Philosophy
in Computer Science

Thursday, March 21, 2024

4:30 PM

Martin Complex, Room 354

Computer Science Department
Conference Room

Advisor & Committee Chair:

Dr. Mohd Anwar

Committee Members:

Dr. Xiaohong Yuan

Dr. Letu Qingge

Dr. Hossein Sarrafzadeh

Dr. Abebe Kebede

Pandemic Surveillance Using Social Media Data, Natural Language Processing, and Machine Learning

Abstract

Social media and online social networking sites have become key platforms for daily social interactions. During the peak of the COVID-19 pandemic, the stay-at-home policy and quarantine measures played a crucial role in both reducing the transmission rate of the pandemic and increasing social media usage. Researchers have increasingly turned to conversational social media as a valuable alternative source of insights into public behavior and sentiment during crises. However, research in this field needs to harness data from multiple social media platforms to build a comprehensive public health surveillance system, showcasing its multi-domain use cases.

In this dissertation, we aim to utilize social media datasets, natural language processing (NLP), and machine learning (ML) to address the following research questions (RQs): (RQ1) To what extent can conversational social media serve as an alternative data source to understand public concerns about a pandemic? (RQ2) To what extent can conversational social media be used as an alternative data source for tracking pandemic-related information? (RQ3) How effective are machine learning models in forecasting the trends of a pandemic? (RQ4) To what extent can machine learning models based on conversational social media data help gauge public sentiment about a pandemic? We employ a multi-case study approach to comprehensively address the research questions. Each case study is carefully crafted to investigate various scenarios and leverage diverse data sources. Additionally, we explore the potential benefits of employing quantum computing for a sentence classification task. This comprehensive design enables us to derive robust conclusions and valuable insights for monitoring and mitigating the impact of a pandemic at a population level.

Our research makes significant contributions to monitoring the spread of a pandemic, identifying trends and patterns, predicting surges or peaks, and forecasting outbreak hotspots. Our work lays the foundation for comprehensive analysis by leveraging social media as a data source, showcasing the potential of NLP and ML methods in public health surveillance and response strategies, and providing valuable perspectives for addressing health-related crises on different social media platforms.

Human-Centered AI (HCAI) Lab | Quantum Computing Research Center <QC|RC>