MSc Defence

Thursday July 29, 2021 at 10AM via Zoom

Automated Generation of Privacy Policies Using Deep Models

Nastaran Bateni

Chair: Dr. Fangju Wang
Advisor: Dr. Rozita Dara
Advisory Member: Dr. Ritu Chaturvedi
Non-Advisory Member: Dr. Fei Song

ABSTRACT:

Privacy policies are statements about how websites, applications, and any other service providers collect, use, share and manage users' data. Nowadays, the contents of privacy policies have been affected by different regulations such as the General Data Protection Regulation (GDPR), which is a framework that enforces the protection of personal data and requires privacy policies to be more transparent for readers. There is a limited understanding of how GDPR has impacted the content of privacy policies. This study presented a comprehensive framework for evaluation of compliance of privacy policies with GDPR recommendations and best practices. This evaluation framework includes text feature analysis, coverage analysis, and content analysis.

Our findings suggest that although GDPR enforcement has improved the content of privacy policies, many of these legal agreements do not satisfy GDPR requirements. In addition to analysis of post-GDPR policies, we utilized machine learning methods for automatic generation of data practices. OPP-115 dataset was used, to train sequence-to-sequence models based on deep neural networks, including Long Short-Term Memory (LSTM) and bidirectional Long Short-Term Memory (bi-LSTM), to generate legal data practices and content in the three levels of paragraph, sentence, and data practice. Our findings have suggested that models trained on legal data practices using bi-LSTM algorithm are most similar to the original privacy policy data practices.