MSc Defence

April 17, 2019 at 10:30AM in Reynolds, Room 1101

Using Convolutional Neural Networks to Extract Keywords and Keyphrases from Foodborne Illness Scientific Articles

Jingjing Wang

Chair: Dr. Xiaodong Lin
Advisor: Dr. Fei Song
Co-Advisor: Dr. Rozita Dara
Non-Advisory Committee Member: Dr. Fangju Wang

Abstract:

Although taxonomies have been created for many domains such as biology, where biological organisms are organized into a hierarchy of groups based on common characteristics, research on generating a taxonomy for the field of foodborne illnesses is still quite limited. There is no standard set of keywords and keyphrases for foodborne illnesses that can be used to construct a taxonomy for this domain. Keywords and keyphrases can be manually extracted by domain experts, but this process is time-consuming. Thus, it is desirable to extract them automatically using statistical, machine learning, or deep learning methods. In this paper, we propose a new approach by using Convolutional Neural Networks (CNN) to identify keywords and keyphrases for foodborne illnesses. To accomplish this, we collected science articles on foodborne illnesses from PubMed and constructed a set of annotated keywords and keyphrases based on TF-IDF. Moreover, we compared the results with several semi-supervised/supervised (KEA, GuidedLDA) and unsupervised (LDA) machine learning algorithms. Our experiments showed that the new solution outperforms other algorithms when tested on a large dataset by selecting more relevant keywords and keyphrases for foodborne illnesses.