

Building a biomedical literature knowledge graph and automatic screening of biomedical abstracts using knowledge graph embeddings

Iqra Muhammad
iqra.muhammad@liverpool.ac.uk

June 2020

1 Introduction

Biomedical literature is increasing at an exponential rate. Researchers in the biomedical community have to spend hours screening articles for inclusion in systematic reviews. In this study, we consider the problem of automating the manual search process of biomedical literature using knowledge graph embeddings. We develop a system that automatically generates a literature knowledge graph of biomedical abstracts. We use open information extraction based methodology to extract concepts and relations from a document collection of 55,000 biomedical abstracts. The concepts and relations identified using open information extraction, result in the creation of an open knowledge base. We use entity linking for disambiguation of biomedical concepts in the knowledge graph. The extracted relations and concepts are fed into a graph database. We propose to use graph neural networks for generating graph embeddings using the literature knowledge graph. Recently, Graph neural networks have been used on text classification, since GNN have the ability to handle complex structure [1]. Our aim is to use graph neural networks for generating graph embeddings from subgraphs in the knowledge graph. Each abstract in the knowledge graph, is in the form of a subgraph of concepts and relations. Graph neural networks will be used to convert subgraphs (for each abstract) into graph embeddings. These graph embeddings will be used to rank abstracts for screening and inclusion in systematic reviews [2]. Currently, we are in the process of collecting relevance judgements for abstracts and we plan to use these relevance judgements for our future experiments.

References

- [1] Lianzhe Huang, Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng WANG. Text level graph neural network for text classification. *arXiv*

preprint arXiv:1910.02356, 2019.

- [2] Christopher R Norman, Elizabeth Gargon, Mariska MG Leeflang, Aurélie Névéal, and Paula R Williamson. Evaluation of an automatic article selection method for timelier updates of the comet core outcome set database. *Database*, 2019, 2019.