

# Homework 3: Summaries of Poster Presentations

Sanchit Nevgi

April 29, 2020

## Unsupervised semantic matching for job postings

In this project, the author tackles the problem of job-candidate matching. Candidates and jobs are represented by distributed representation. These representation are learned by a Graph Convolutional Network, where the graph is heterogenous and has information such as *job title*, *job skill*. Further, Neural topic modeling is used on the job description to obtain a representation. Finally these representations are compared using average cosine similarity to obtain a matching.

## Comparative study of BERT models on open-QA on the CORD-19 dataset

In this project, the authors tackle performances of BERT models on Question-Answering on medical documents. The architecture consists of an Encoder and Mapper. The Encoder consists of BERT encoder using sliding window on the document. The Mapper maps to the best cluster. The models tested are BERT, BioBert, and SciBert. The authors also experimented with different pooling approaches (Mean/Max pool). The embeddings generated by the BERT models are clustered. From the clusters, the the top-k documents are obtained and these are compared to manually annotated top documents. Euclidean distance is used as a similarity metric.

## Improving Graph embeddings

The project aims to improve the quality of Knowledge Graph embeddings. In this approach, the vectors are factorized into real and imaginary parts, and 2 boxes are generated for each (*head*, *relation*) pair. The intuition is that box representation can capture more semantic meaning. Next an intersecion box is calculated by averaging the embedding and multiplying the softmax height. A distance is computed to the tail vector to these boxes and an average is computed. The model is optimized for the average distance between the boxes and the tail vector.