Abstract

Commonsense knowledge is crucial in today's artificial intelligence research area. It contains truths and information of our daily situations, for example "Sugar is sweet" or "Breathing makes a person alive", that most of the humans are expected to own by the time they grow up. It is an unsolved artificial intelligence problem till this era. Many working commonsense knowledge programs are needed to assist decision making in AI expert systems. Thus, commonsense knowledge becomes an essential joint for such AI expert systems. The importance of commonsense reasoning was recognized since many of these systems were able to reason, but they were also vulnerable because they frequently offered nonsensical responses when faced with unexpected problematic data.

This becomes our core motivation to dive into this area of research. Speaking of diving into this field, there are many approaches to make machines gain commonsense knowledge by means of natural language processing, computer vision, etc. However, from all the efforts made over years, it is obvious that teaching machines to have commonsense knowledge is a time-consuming and expensive process. Two main reasons why natural language processing is chosen to proceed for this thesis are:

1. Natural language processing can be treated as a base brick for further applications such as audio recognition or computer vision applications that need commonsense knowledge,

2. It is less costly compared to other approaches.

In recent years, knowledge graph embedding algorithms become popular for knowledge base completion tasks. Those prior studies enlightened in a way that if we can use knowledge graph models for knowledge base completion tasks, it can also be helpful in commonsense knowledge mining task along with certain enhancements. Each embedding model has its unique performance when they are dealt with different parts of a dataset. Among various well-known embedding methods, state-of-the-art models [32] were chosen to use in our experiments.

From the aforementioned points, we conducted our thesis with following main contributions:

1. We presented the proposal of using the CKBC dataset to testify the ability of Knowledge Graph Embedding Models,

2. We reproduced the results for 24 knowledge graph embedding models which are built-in models from PyKeen,

3. We found out the different results on each model, select the best ones and implemented ensemble from the results from step 2,

4. Our ensemble method shows that combining the results can produce better performace.

Keywords: Natural Language Processing, Knowledge Graph Embedding Models.