One of the most popular methods of converting unstructured data into a structured format

is by relation extraction. To perform relation extraction, identification must be performed for

each entity within the data. This is then followed by identifying whether each pair of entities

are semantically related. Finally, through analysis, the identified semantic relations are classified accordingly. While several previous attempts performed toward smaller sets of text data have shown promising results, performances against larger sets of data tend to suffer. This shortcoming is especially easy to spot when having to identify global mention-level relations. Global mention-level relations are relations that stretch among long contexts, usually described through the course of many other intermediate sub-relations [3,4]. This kind of short-sightedness may also affect the accuracy of local relation extractions as well, since the model may misinterpret ambiguous relations or even omit them entirely if there are types of relations that only become coherent when viewed on a more global scale such as metaphors or sarcasm.

In this article, we propose a relation extraction method that focuses on extracting global relations by utilizing knowledge graph construction and processing. The model largely consists of two parts:

a context-aware long short-term memory(LSTM) [5] based relation extraction module and

a knowledge graph constructor module that produces knowledge graphs from the given document.

By constructing a knowledge graph from the input document, the constructed knowledge graph is used to extract the input document’s global relations through graph embedding. To do this, both modules go through a synchronization process. We achieved this by training the two modules in an alternating fashion. This allows the proposed model to extract and classify local relations while concerning its possible global relations. For this research, we have constructed a new set of data consisted of 10,470 lines of documented Korean text of history [6] for this task. The dataset was constructed purely through human annotations and contains the information of both local relations per each sentence and knowledge graphs that represents each document.

2. RelatedWorks

2.1. Relation Extraction

Relation Extraction can largely be divided into two types: global level relation extraction and

mention-level relation extraction [7], while the former deals with ‘global’ level relations that span across multiple relations and large bodies of text, and mention-level ‘local’ relations are smaller scale relations expressed directly within short sentences.

The best example that shows the difference between the two can be shown by syllogism. Given two sentences “All cars have wheels.’ and “A truck is a car”, a local relation extraction will only be able to extract the information (Car, possess,Wheel) and (Truck, is, Car). Thus, it will never be able to extract and represent what the two sentences combined mean: “A truck has wheels”. On the other hand, global relation extraction would consider these global level relations and would be able to extract the information (Truck, Possess,Wheel).

Currently, most relation extraction modules that yield high performances usually only deal with

‘local relations’. However, when representing a large amount of text such as a document, representations only through local relations are thoroughly limited. Considering the fact that the most valuable information within documents is those that have been described in detail and are spanned across the document, the importance of global level relation extraction is not something to be looked over so lightly.

However, the task itself is nowhere near simple. The main cause that prevents current

state-of-the-art local relation extraction methods to properly perform global relation extraction is due to the bottleneck problem [8]. With so much information to process for a singular relation, the loss of information is inevitable—thus hindering the model’s capability to properly judge the relation or capture it at all. On the other hand, if the model is focused only on extracting global relations, then, in an attempt to avoid the bottleneck problem [8], its performance against local relation extraction drops due to the sparsity of information [9]. Previous attempts on global relation extraction have tried several different methods to tackle this problem. For instance, Gerber et al. [10] attempted this by simply using co-references to gain access to other sentences without actually having to model them or Quirk et al. [11] who applied distant supervision to general global relation extraction as well.

Ontology:

The ontology is a formal specification of all the relevant concepts and their meaningful associations in a given application domain. It allows objects and relationships to be categorised into distinct types, and for generic properties about those types to be expressed.