

Aligning Open IE Relations and KB Relations using a Siamese Network Based on Word Embedding

Rifki Afina Putri
School of Computing, KAIST
rifkiaputri@kaist.ac.kr

Giwon Hong
School of Computing, KAIST
gch02518@kaist.ac.kr

Sung-Hyon Myaeng
School of Computing, KAIST
myaeng@kaist.ac.kr

Abstract

Open Information Extraction (Open IE) aims at generating entity-relation-entity triples from a large amount of text, aiming at capturing key semantics of the text. Given a triple, the relation expresses the type of semantic relation between the entities. Although relations from an Open IE system are more extensible than those used in a traditional Information Extraction system and a Knowledge Base (KB) such as Knowledge Graphs, the former lacks in semantics; an Open IE relation is simply a sequence of words, whereas a KB relation has a predefined meaning. As a way to provide a meaning to an Open IE relation, we attempt to align it with one of the predefined set of relations used in a KB. Our approach is to use a Siamese network that compares two sequences of word embeddings representing an Open IE relation and a predefined KB relation. In order to make the approach practical, we automatically generate a training dataset using a distant supervision approach instead of relying on a hand-labeled dataset. Our experiment shows that the proposed method can capture the relational semantics better than the recent approaches.

1 Introduction

Open Information Extraction (Open IE) aims at extracting key information from a large amount of text into a structured format, commonly in the form of triples, (*subject entity, relation, object entity*), where the relation denotes the type of a semantic relation between the entities. As opposed to the traditional Information Extraction that generates triples over a predefined relation set, Open IE can extract all possible relations without having to be restricted to a predefined set of relations. However, a relation from an Open IE system is merely a sequence of words coming from the sentence containing the entities, resulting in ambiguous and semantically redundant relations. For example, Open IE may extract "*died in*" and "*location of death*" as two distinct relations although they should be treated as semantically equal and expressed (or canonicalized) with a single relation type.

In order to address this problem, some methods have been proposed to canonicalize Open IE relations (Yates and Etzioni, 2009; Galárraga et al., 2014; Vashishth et al., 2018). Given that they rely on a clustering method, however, they tend to suffer from over-generalization. For example, the latest canonicalization method called CESI (Vashishth et al., 2018) would put "*is brother of*," "*is son of*," "*is main villain of*," and "*was professor of*" into the same relation cluster. While these relation phrases have a common pattern (to be + noun + of) and expresses that the subject entity has a certain role, the overarching relational category is too general to be useful.

Besides Open IE, Knowledge Base (KB) systems such as DBpedia, Freebase, and Wikidata, also store general facts in a triple format. Different from Open IE, the relations in a KB are already classified into distinct semantic categories. Although KB relations are better defined semantically than Open IE relations, they are limited in terms of quantity and coverage. Dutta et al. (2015) attempted to mitigate the weaknesses of the two approaches by aligning the relations of Open IE triples to those in DBpedia,

thereby adding semantics to Open IE triples. While useful, their approach is primarily based on the frequency of triples without explicitly taking into account the relational semantics.

In this paper, we propose a new model using a Siamese network for aligning relations from Open IE to those from KB (i.e. relation alignment task) for the purpose of providing more semantics to Open IE relations, which are to be used for question answering as in TriviaQA (Joshi et al., 2017). The Siamese network, a form of a neural network, takes two sequences of word embeddings representing an Open IE relation and a KB relation and compares them. The network is trained to learn the semantic similarities between an Open IE relational phrase and a KB relation type name that are considered identical in their meanings. By utilizing word embeddings as the input of the network and encode relational descriptions, we can incorporate their semantics information without an extra process of extracting linguistic features from the training data. In order to mitigate the problem of manually constructing training data, i.e. pairs of an Open IE relational phrase and a KB relation type name, we propose a distant supervision method that does not require manual annotations. Our contributions in this paper are:

- We propose a novel method of applying a Siamese network for the relation alignment task. To the best of our knowledge, our model is the first attempt that incorporates the semantic information of the textual descriptions of relations, specifically for the relation alignment task.
- We propose to automatically generate a training dataset using a distant supervision approach so that we avoid manual creation of training data, which can be prohibitive, thereby making the proposed approach practical.
- We experimentally confirm that our model better captures relational semantics than the clustering and the statistical rule-based approaches with a significant margin. We also analyze different variations of the Siamese network to provide insights about the relation alignment task.

2 Related Work

Open IE Canonicalization. Yates and Etzioni (2009) proposed a simple probabilistic method for identifying Open IE triples which has a similar meaning. They calculated similarity between two relation phrases and clustered them with a greedy agglomerative clustering method. Although their model works well in finding synonyms for relation phrases, it still suffers from the polysemy problem. Galárraga et al. (2014) canonicalized relation phrases by employing a rule mining algorithm called AMIE (Galárraga et al., 2013) to mine the relationship rules between relation phrases and clustered the relation phrases based on the generated rules. Recently, Vashishth et al. (2018) improved Dutta’s model by using relation embeddings and side information as the features for the clustering method. They canonicalized the Open IE relations by clustering the embedding. Our task is different from their task since we focus on adding more semantics to the Open IE relations by aligning them to the KB relations.

Instead of relying on only one Open IE systems, Bovi et al. (2015) proposed a method called KB-Unify to integrate the triples from different Open IE systems into a single repository. Our work differs from their work since we attempt to align the Open IE and KB relations. Our task is mostly similar to the alignment task presented by Dutta et al. (2015), which was introduced in Section 1. They aimed to bring the benefits of Open IE and KB by mapping the Open IE triples to existing KB triples (DBpedia) by using a statistical rule-based approach. While their result seems promising, it only relies on frequency of the triples without considering semantics. Besides, it suffers from an efficiency problem arising from frequency calculation.

Word Sense Alignment. Gurevych et al. (2016) define Word Sense Alignment as linking senses or concepts that has an identical meaning from multiple Lexical Knowledge Bases (LKB). There has been a lot of work with various goals such as aligning WordNet, Cyc, and VerbNet for building knowledge representation (Crouch and King, 2005), aligning FrameNet, VerbNet, and WordNet for semantic parsing (Shi and Mihalcea, 2005), and building large-scale LKB alignments (Matuschek, 2015; Gurevych

et al., 2012; Navigli and Ponzetto, 2012). Although this task is conceptually similar with our relation alignment task, we focus on aligning the relation meaning of Open IE and KB, not word sense in general.

Relation Extraction using Distant Supervision. There are many works that used distant supervision method to generate the dataset for relation extraction task such as Mintz et al. (2009) and Riedel et al. (2010). Sorokin and Gurevych (2017) proposed a LSTM-based neural network to extract relation using another relation in the same sentence as a contextual information and utilized Wikidata to construct the dataset. Even though we also use Wikidata in our dataset generation method, however, we aim to align the Open IE relations to the extracted relations in Wikidata.

3 Model Description

3.1 Task Description

Let \mathbf{x}_{OIE} be an Open IE triple and \mathbf{x}_{KB} be a KB triple. Given \mathbf{x}_{OIE} and \mathbf{x}_{KB} as the input, the goal is to determine whether the relation in \mathbf{x}_{OIE} can be aligned to (i.e. expressed with) that of \mathbf{x}_{KB} . If they can be aligned, the model will give 0 ("semantically same") as the output, or 1 ("not semantically same" or "semantically different") otherwise. For example, given **(English, are language of, England)** as the Open IE triple and **(English, official language, England)** as the KB triple, we want to determine whether the relation *are language of* is semantically close enough to and hence can be replaced by the KB relation *official language*. Given the task, our proposed model essentially gives the distance of the Open IE and KB relations based on the weights learned for the network so that it predicts whether the pair is semantically same or not.

3.2 A Siamese Network for Relation Alignment

The concept of a Siamese network was introduced by Bromley et al. (1993) and typically used for measuring the similarity of two inputs. It consists of two identical sub-networks that extract the features from two inputs, respectively. Then the distance from the two sub-network outputs is calculated to determine the input similarity. Note that the two sub-networks will have been learned at the training stage in such a way that the distance between the semantically identical inputs is minimized. The overall model architecture is in Figure 1.

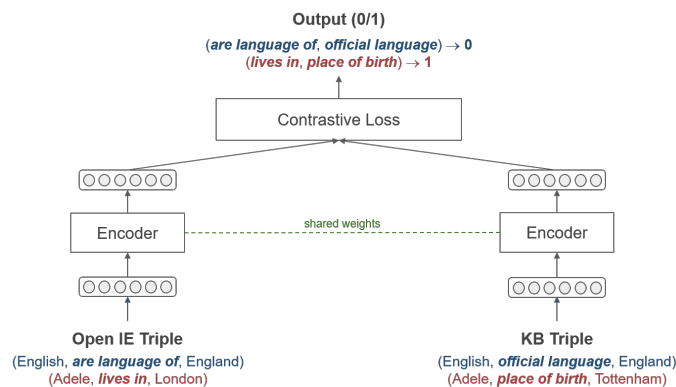


Figure 1: General architecture of the proposed model including the input and output example. The *blue* text represents positive example and the *red* text represents negative example.

In the proposed network, the first and second sub-networks attempt to capture the features from the Open IE and the KB relations, respectively. The embeddings of the input words on each sub-network are encoded to produce a new vector. Note that the encoders share the same weights. For training, we use a contrastive loss function. The details of the model are described in the following sub-section.

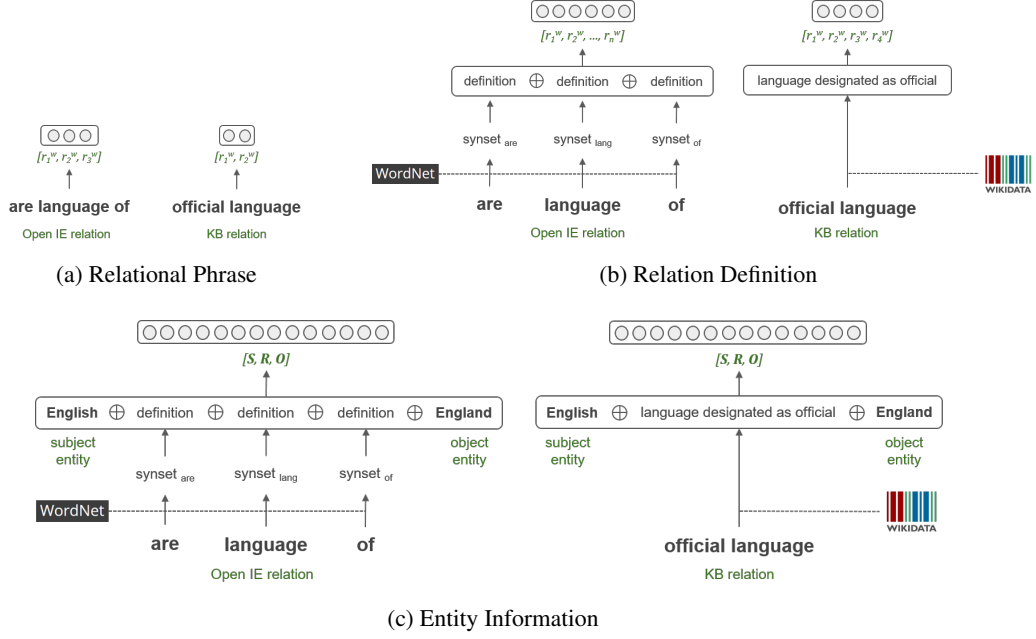


Figure 2: The illustration of input representation.

3.2.1 Input Representation

Before an input is fed into the encoder, each word in the input is converted into a fixed size n -dimensional word embedding. Given that relation phrases from Open IE and names (from a KB) can be too short to carry enough semantics, we also utilize relation definitions and entity information when input representations are computed. Therefore, we have three input representations as follows:

- **Relational Phrase.** For a relational phrase input (Figure 2a), each word w is transformed into real-valued vector $r^w \in \mathbb{R}^d$ where d is the dimension of the word embedding. For the encoding of the entire phrase, we concatenate the word vectors for the phrase to generate a phrase embedding, $\mathbf{x} = [r_1^w, r_2^w, \dots, r_n^w]$, where n is the number of words in the phrase.
- **Relation Definition.** For a relation definition input (Figure 2b), we utilize WordNet to obtain an Open IE relation phrase definition and Wikidata for additional relation description of each KB relation. For an Open IE relation, we transform each word in the relation phrase into WordNet synset using the Lesk algorithm¹. The definition of each synset is then obtained from the WordNet dictionary. For a KB relation, we utilize Wikidata API² to get its description that serves as the definition. Finally, the input representation is formed by concatenating of the word vectors in the definition text, i.e., $\mathbf{x} = [r_1^w, r_2^w, \dots, r_n^w]$ where n is the number of words in the definition.
- **Entity Information.** Besides relation information, we consider entity information as an additional feature in our model (Figure 2c). The entity information is construed as the context surrounding the relation and hence providing the semantics of the relation. The subject and the object entity phrases are concatenated to the relation phrase or definition, i.e., $\mathbf{x} = [\mathbf{S}, \mathbf{R}, \mathbf{O}]$ where $\mathbf{S} = [s_1^w, s_2^w, \dots, s_t^w]$, $\mathbf{R} = [r_1^w, r_2^w, \dots, r_u^w]$, $\mathbf{O} = [o_1^w, o_2^w, \dots, o_v^w]$; t and v denotes the number of words in the subject and object entities, respectively, and u denotes the number of words in the relation phrase.

¹https://www.nltk.org/_modules/nltk/wsd.html

²<https://query.wikidata.org>

3.2.2 Encoder

After we have the input representations for Open IE and KB relations, the next step is to feed them to the encoder. Considering the past success of CNN in extracting appropriate features for relation extraction from a sequence of words (Nguyen and Grishman, 2015), we opt for two encoders as follows:

- **Convolutional Neural Network (CNN).** CNN has three main parts: convolution, max-pooling, and fully-connected linear layers. In the convolution layer, we aim to extract the local features from the given input text. By extracting local features, we can create a subject, a relation, and an object representation similar to n-gram features. The max-pooling layer selects the most important features contained in the phrase. The output from the max-pooling layer is fed to a feed-forward fully-connected linear layer. Finally, its output is used as our final relation representation.
- **Piecewise Convolutional Neural Network (PCNN).** In PCNN, which was first introduced by Zeng et al. (2015), the original max-pooling layer is modified into piecewise max-pooling. We apply a max operation over a segment of the phrase so that the model can extract the important features without losing the information coming from the subject entity, the relation, and the object entity, separately. Finally, similar to the CNN encoder, the output from the feed-forward linear layer is used as the final relation representation.

3.3 Contrastive Loss Function

For learning, we apply a contrastive loss function defined as the sum of the loss of positive examples (semantically same relations) and the loss of negative examples (semantically different relations). More formally, the loss function is defined as:

$$L = (1 - Y)D^2 + Y(\max(0, m - D))^2; m > 0 \quad (1)$$

$$D = |\mathbf{x}'_{OIE} - \mathbf{x}'_{KB}| \quad (2)$$

where Y and D denote the label of the input pairs (0 for semantically same, 1 for semantically different) and the euclidean distance between the Open IE and the KB relation vectors (i.e. the output from the encoder explained in the previous section) respectively, with m being a margin. Note that the first term of Equation 1 is used for positive examples and the second term for negative examples. When training, we want to make the distance of the positive pairs smaller and the distance of the negative pairs inside the margin larger.

4 Dataset Generation using Distant Supervision

The distant supervision method for the task of relation extraction was first introduced by Mintz et al. (2009). It assumes that any sentence containing an entity pair participating in a triple of a known KB is likely to contain a relevant expression of the relation of the triple. As a result, it becomes possible to construct positive training instances for the relation in the triple by taking the expressions between the occurrences of the two entities. The collection of textual expressions can be used as revealing the target relation. By adopting this approach, we can obtain the sentences containing the target relation in KB and use them to extract Open IE triples with the relation. Once the Open IE triples are generated, we apply some rules to annotate them as positive or negative automatically so that we obtain training data for the KB relations used in collection the Open IE triples. The training data generation steps are as follows:

1. Select the top 200 most frequent relations³ in the KB and collect the KB triples containing one of the relations. We utilize Wikidata (Vrandečić and Krötzsch, 2014) as our KB.

³As of October, 2018

2. Crawl the sentences for each triple using the distant supervision method. In other words, we pick the sentences containing the two entities of the triple. In order to reduce ambiguities associated with the occurrences of the entities, we retrieve sentences from the Wikipedia page of each entity.
3. Apply Open IE to each sentence to extract Open IE triples. In this paper, we use the existing Stanford Open IE system (Angeli et al., 2015).
4. Align Open IE and KB triples. The triples *sharing the same entity pair* are labeled as *semantically same* or *positive* (0). But if *one of the entities is different*, it is labeled as *different* or *negative* (1).
5. From the previous step, we will get a small amount of positive examples but a high number of negative examples. To handle the data imbalance problem, we add more positive examples by swapping a pair of alignments when the other sides of the two alignments share the same relation but with different entities. For example, when we have positive examples as follows:

(Inn, country, Switzerland), (Inn, is river in, Switzerland), 0
 (Villavicencio, country, Colombia), (Villavicencio, is city in, Colombia), 0

we generate two additional positive examples by swapping the right hand side triples as follows:

(Inn, country, Switzerland), (Villavicencio, is city in, Colombia), 0
 (Villavicencio, country, Colombia), (Inn, is river in, Switzerland), 0

5 Experiments

The goal of our experiments are two-fold: the first is to examine the influence of different input representations and encoder variations of our model in capturing the semantics of the relations of the Open IE and the KB and the second is to compare our model against the existing approaches for aligning Open IE and KB relations. The existing approaches that serve as the baselines are:

- **CESI** (Vashishth et al., 2018): For this model, we adjust the clustering result so that it can be compared with our model for the evaluation tasks to be described below. If two relations are in the same cluster, then they are labeled as semantically same; otherwise different.
- **Dutta et al.** (2014): This model uses a statistical rule-based approach for aligning relations. It calculates a confidence score of every possible Open IE relations mapping to a KB relations based on occurrence statistics of the particular mapping. If the mapping has a higher confidence than the threshold determined by linear regression, it is labeled as semantically same; otherwise different. Because the code has not been shared by the authors, we implemented their method on our own.

Besides the above baselines, we also apply our alignment rule (denotes as **rule-based** in Table 2) used in the dataset generation process (see Section 4) for predicting the label, i.e., the triples sharing the same entity pair are labeled as semantically same; otherwise different. Note that this case is used as a reference point in explaining the performance of the proposed method and the other baselines. It also can be used to measure the quality of the distant supervision dataset.

Since there is no standard evaluation suit available for the relation alignment task, we provide three evaluations to reveal different aspects of the proposed model and compensate for the limitations of each.

1. **Internal Evaluation with Automatically Generated Dataset.** The goal is to examine different variations of the proposed model using automatically generated test data of a large quantity. It is internal because we only compare different variations of the proposed model, not against other methods. We split our automatically generated dataset into training, validation, and testing datasets (see Table 1 for details).

- **CNN_no_def**: This version uses CNN as the encoder and relational phrases and relation names (no definitions) as the input representation for both Open IE and KB relations.
- **CNN_def**: This is the same as CNN_no_def except that relation definitions are added.
- **CNN_def_ent**: This version is the same as CNN_def except that entity information is added.
- **PCNN_no_def_ent**: This version uses PCNN as the encoder and relational phrases and relation names for Open IE and KB relations as the input, respectively, as well entity information.
- **PCNN_def_ent**: This is the same as PCNN_no_def_ent except relation definitions are added.

Set	# triples	# sentences	# alignments		
			Positive	Negative	Total
Training	86,178	102,863	430,364	430,364	860,728
Validation	34,726	42,874	184,621	184,621	369,242
Testing	32,309	39,477	168,257	168,257	336,514

Table 1: Statistics of our dataset generated by the distant supervision method.

Since we use a large number of sentences and triples extracted thereof, this evaluation allows us to test different variations for all the relations exhaustively.

2. **Manual Evaluation.** This evaluation is intended to overcome a drawback of the internal evaluation, which relies on the assumption that the gold standard generated by distant supervision is always correct. Another limitation is that it does not include external evaluation. Therefore, in this evaluation, we use a manually annotated test data set and use it as the gold standard to make the evaluation more reliable and compare the performance of the proposed model with the two existing approaches mentioned above.⁴ An added value is that we can indirectly examine the reliability of the internal evaluation method by comparing the relative ordering of the variations. To build the dataset, we randomly sampled 400 alignments from the distant supervision testing data. The dataset has all unique entity pairs and it covers 90 unique KB relations. For each pair of relations, one from Open IE and the other from KB, we asked three annotators to decide whether the relations were semantically same or not, resulting in 258 "same" and 142 "different" relation pairs. The inter-judge agreement was 81.88% in Fleiss' Kappa.
3. **Qualitative Analysis.** The goal is to examine the strengths and weaknesses of the proposed model by looking at different lexico-syntactic complexities of the relational phrase patterns, relative to the two baselines. We chose a smaller sample of the alignment result than the above "manual evaluation", including ten semantically same relation pairs and five semantically different ones. For the semantically same relation pairs, we divide the set into two categories: lexical similarity vs difference. Lexical similarity means the relations share at least one similar word, for example "*died in*" and "*place of death*" relations. Lexical difference means the relational phrases do not share a lexically similar word at all, for example "*'s son is*" and "*child*" relations.

For the evaluation metric, we use precision (P), recall (R), F1, and accuracy (Acc) scores.

$$P = \frac{T_P}{T_P + F_P} \quad R = \frac{T_P}{T_P + F_N} \quad F1 = 2 \cdot \frac{P \times R}{P + R} \quad Acc = \frac{n_{correct}}{n_{total}} \quad (3)$$

where T_P denotes the number of true positives, F_P the number of false positives, F_N the number of false negatives, $n_{correct}$ the number of correct predictions, and n_{total} the number of total testing data. Note that the score presented in this paper is the best score over multiple runs.

In the training process, we applied the filter height of 1 and 2 with 100 feature maps for the convolutional layer. For the input, we used pre-trained fastText (Bojanowski et al., 2017) with 300 dimension size and update the weight of the word embeddings. For learning, we applied a stochastic gradient descent algorithm using Adam optimizer (Kinga and Adam, 2015) with 0.001 as the learning rate. The

⁴The dataset and code are available at: <https://github.com/rifkiaputri/rel-aligner>

batch with size was 128. Also, we employed dropout in the feed-forward linear layer with a probability of 0.5. For the loss function, we set the margin m as 2.

6 Result and Discussion

6.1 Internal Evaluation

The relative performance differences among the five versions are summarized in Figure 3a. In addition to the ordering of the five variations, the difference between CNN and PCNN encoders is most notable. Detailed analyses are as follows.

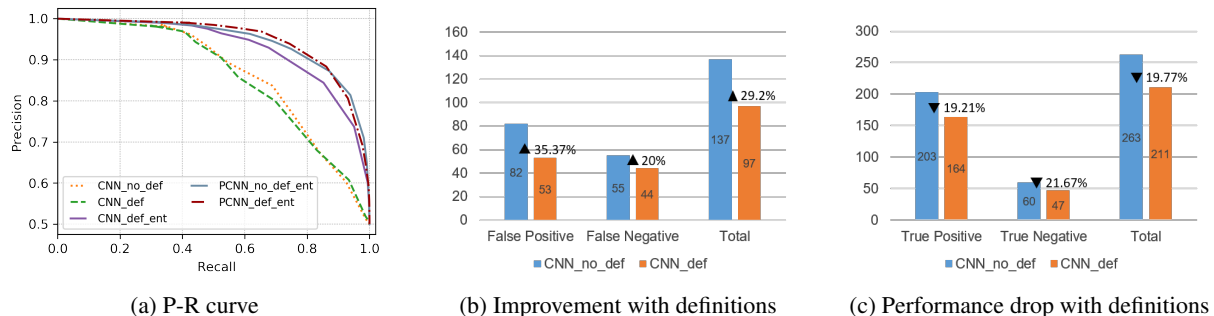


Figure 3: Internal evaluation result.

Relational Phrase vs. Relation Definition. While our intuition was that the additional information obtainable from the relation definitions would help compensate for the lack of semantics in a short relation phrases and names, it turns out that the overall gain shown in Figure 3a is not as significant as our expectation. A further analysis shows, however, that definitions help reduce incorrect predictions as in Figure 3b. Out of 137 errors (82 false positives and 55 false negatives) made by CNN_no_def, 40 were predicted correctly by including definitions (CNN_def), resulting in 29.2% improvement. On the other hand, out of 263 correct prediction in CNN_no_def (203 true positives and 60 true negatives), 52 were predicted incorrectly in CNN_def, resulting in 19.77% drop. This suggests that adding definition has potential to enrich the semantics; more sophisticated approaches are left for future research.

Impact of entity information. We observe that the performance of CNN_def_ent is significantly higher than that of the CNN_def model. From this result, we can conclude that adding entity information contributes to predicting the similarity between two relations. It suggests that entity information provides the context with which relation phrases and names can be aligned more accurately. It is consistent with the result in Zeng et al. (2015) that also shows the importance of including entities in relation classification.

CNN vs. PCNN. Compared to the performance CNN_def_ent, PCNN_def_ent is clearly better, strongly suggesting that for the relation alignment task, the PCNN encoder is better than CNN, regardless of whether relation definitions are used. A rational explanation for this result is that we lose important information when we apply max-pooling to the entire input representation including entities and relational phrases in CNN. Note that in PCNN, piecewise max-pooling allows the model to extract major features from three different segments of the representation (i.e. subject entity, relation, and object entity). Therefore, this result confirms that the piecewise max-pooling helps in preserving more meaningful features resulting from the convolutional layer for the relation alignment task.

6.2 Manual Evaluation

For more reliable evaluation of the proposed model, we compared it against the baselines using the 400 gold standards labeled by human. The summary result for predicting the semantically same and different

Model	Semantically Same			Semantically Different			Overall
	P	R	F1	P	R	F1	Acc
Dutta	0.740	0.496	0.594	0.427	0.683	0.526	56.25%
CESI	1.000	0.066	0.124	0.371	1.000	0.541	39.75%
<i>Rule-based</i>	<i>0.645</i>	<i>1.000</i>	<i>0.784</i>	<i>0.000</i>	<i>0.000</i>	<i>0.000</i>	<i>64.50%</i>
CNN_no_def	0.742	0.547	0.629	0.443	0.655	0.528	58.50%
CNN_def	0.726	0.678	0.701	0.478	0.535	0.505	62.75%
CNN_def_ent	0.659	0.891	0.758	0.451	0.162	0.238	63.25%
PCNN_no_def_ent	0.678	0.880	0.766	0.523	0.239	0.329	65.25%
PCNN_def_ent	0.669	0.922	0.775	0.545	0.169	0.258	65.50%

Table 2: Manual evaluation result.

relation pairs in Table 2 clearly shows the proposed model (CNN and PCNN) outperforms the baselines, CESI and Dutta, in predicting the semantically same pairs. Although CESI has the highest precision score, it has the lowest recall among all models variations due to the bias of predicting most of the data as semantically different. While it is possible to apply a different threshold in forming clusters for different precision and recall pairs, the low F1 value precludes its moderate performance for relation alignment. Compared to CESI models, Dutta’s model shows better performance, especially in recall and F1 scores for predicting semantically same pairs. Based on this result, it is obvious that using a simple probabilistic rule-based approach is better than using the clustering approach for the relation alignment task. However, it is much worse than our model variations, with a high number of false negatives, resulting in the low recall and F1 scores.

The scores of the *rule-based* model in Table 2 is provided as a reference point of our proposed models. Since the rule-based model predicts the label using the alignment rules in our distant supervision dataset generation, a pair (one from Open IE and the other from KB) sharing the same entities in respective triples is judged to be semantically same by this model. That is, all the 400 pairs are predicted to be semantically same. From the alignment task perspective, it gives 100% recall for the semantically same case (all of the 258 "same" pairs are predicted correctly) and 0% recall for semantically different case. Since all the 142 "different" pairs are predicted as "same", the overall accuracy score is 64.5%. It shows that the training process is still needed since we cannot only rely on the alignment rules in our dataset generation. Moreover, since the *rule-based* model predictions are made with the distant supervision rule, we can also infer the quality of our distant supervision dataset based on the scores of the model (64.5% alignments are correctly labeled by distant supervision). Note that the low F1 score for predicting semantically different pairs in the proposed model is attributed to the high number of false negatives in the dataset. However, it still has the highest overall accuracy score compared to the two baselines.

6.3 Qualitative Analysis

We selected a sample of the alignment result and examined the label of each model to obtain insights about success and failure cases. As in Table 3, the variations of the proposed model tend to perform very well in predicting positive examples. For negative examples (labeled as "semantically different" in Table 3), almost no model predicts the alignment label perfectly, with an exception of the CESI model. Note that CESI has the tendency of predicting most alignments as semantically different, generating many false negatives. The "correct" decisions made for the semantically different pairs are likely to be attributed to this tendency.

Furthermore, the proposed model appears to make correct predictions for a pair where the relation expressions are lexically different but semantically same, as in $\langle 's\ son\ is,\ child \rangle$, $\langle 's\ serial\ is,\ notable\ work \rangle$, and $\langle was\ first\ married\ to,\ spouse \rangle$. However, CESI and Dutta’s fail to predict them correctly because they are difficult to be predicted the models using symbolic representation of words. This result indicates that distributed representation of words as in embeddings has a clear advantage in dealing with semantics even when radically different words are used in relation phrases. For semantically and lexically similar relational phrase pairs such as $\langle died\ in,\ place\ of\ death \rangle$, $\langle died\ of,\ cause\ of\ death \rangle$, and $\langle was\ filmed\ in,\ filming\ location \rangle$, almost all models predict the alignment correctly, except CESI, again

Open IE Relation	KB Relation	Dutta	CESI	CNN			PCNN		Gold
				no_def	def	def_ent	no_def_ent	def_ent	
<i>Semantically same - lexically similar</i>									
died in	place of death	0	1	0	0	0	0	0	0
died of	cause of death	0	1	0	0	0	0	0	0
founding member of	member of	1	0	0	0	0	0	0	0
was filmed in	filming location	0	1	0	0	0	0	0	0
permanent capital of	capital of	1	0	0	0	0	0	0	0
<i>Semantically same - lexically different</i>									
' son is	child	1	1	0	0	0	0	0	0
's serial is	notable work	1	1	0	0	0	0	0	0
was first married to	spouse	1	1	0	0	0	0	0	0
is main villain of	present in work	0	1	0	1	0	0	0	0
first aired with	original language of work	1	1	1	1	0	0	0	0
<i>Semantically different</i>									
was born in	place of death	0	1	0	0	0	0	0	1
have won	nominated for	1	1	0	0	0	0	0	1
began trip to	place of death	1	1	1	0	1	1	0	1
was born in	work location	0	1	0	0	0	0	0	1
leave	father	1	1	1	1	1	1	1	1

Table 3: Alignment results of some Open IE and KB relations.

due to its tendency of judging pairs as semantically different. In another example, $\langle \textit{founding member of}, \textit{member of} \rangle$ and $\langle \textit{permanent capital of}, \textit{capital of} \rangle$, Dutta’s model makes an incorrect prediction which attributed to the fact that the model relies heavily on the frequency of the training instance and that the frequency of the pairs is low.

Note that most models fail to correctly predict the cases where the pairs look similar but are in fact semantically different as in $\langle \textit{was born in}, \textit{place of death} \rangle$. We argue that this is caused by the existence of noisy instances in our training dataset. As explained in Section 4, when we automatically labeled the dataset with distant supervision, it assumed that the triples sharing the same subject and object entities would have a semantically same label. Obviously, this assumption does not always hold. Out of 4,274 examples of $\langle \textit{was born in}, \textit{place of death} \rangle$ pair in the training set, around 96% of the instances are labeled as semantically same because the triples share the same entities. In other words, the false positive problem is due to the distance supervision used for constructing training instances.

7 Conclusion and Future Works

In this paper, we present a Siamese network for aligning the relations of Open IE (Stanford Open IE) and relations of KB (Wikidata). As a way to overcome the difficulty of acquiring a large number of training instances, we built an extensive amount of training dataset using the distant supervision method which does not require manual annotation. In the experiments, we first confirm that using word embedding as the input of Siamese network is effective in extracting the semantics information compared to the probabilistic rule-based model and the clustering-based model. Adding a textual definition and entity information as the additional feature may also help to reduce the false positive and false negative errors that occur when we only use short relational phrases input.

Despite the superiority of the performance over the baselines, the dataset resulting from the distant supervision method still suffers from noises, i.e., incorrectly labeled alignment instances. The model variations presented in this paper have not been able to handle this problem, which we leave for future work. Another thing to consider is the number of KB relations. In this paper, we covered the top-200 most frequent relations over a total of more than 1000 relations in Wikidata. Even though we can include all the relations, the number of triples and sentence examples of the last relation is not as much as those of the first relation. Future work will have to investigate on how to handle the imbalance number of relation instances and increase the number of KB relations that are aligned through it.

Finally, relation alignment can be useful for several downstream tasks such as KB completion and question answering. In KB completion, we can combine the Open IE and KB relations by aligning

semantically same relations to the existing KB and adding new relations from Open IE, which are not semantically the same as any of the KB relations. In question answering task, less ambiguous triples resulting from the alignment process can be also used for question answering systems.

Acknowledgement

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2013-0-00179, Development of Core Technology for Context-aware Deep-Symbolic Hybrid Learning and Construction of Language Resources) and Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT (2017M3C4A7065962).

References

- Angeli, G., M. J. J. Premkumar, and C. D. Manning (2015). Leveraging linguistic structure for open domain information extraction. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Volume 1, pp. 344–354.
- Bojanowski, P., E. Grave, A. Joulin, and T. Mikolov (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics* 5, 135–146.
- Bovi, C. D., L. E. Anke, and R. Navigli (2015). Knowledge base unification via sense embeddings and disambiguation. In *EMNLP*, pp. 726–736. The Association for Computational Linguistics.
- Bromley, J., I. Guyon, Y. LeCun, E. Säcker, and R. Shah (1993). Signature verification using a "siamese" time delay neural network. In *Proceedings of the 6th International Conference on Neural Information Processing Systems, NIPS'93*, San Francisco, CA, USA, pp. 737–744. Morgan Kaufmann Publishers Inc.
- Crouch, D. and T. H. King (2005). Unifying lexical resources. In *Proceedings of the Interdisciplinary Workshop on the Identification and Representation of Verb Features and Verb Classes*, pp. 32–37.
- Dutta, A., C. Meilicke, and H. Stuckenschmidt (2014). Semantifying triples from open information extraction systems. In U. Endriss and J. Leite (Eds.), *STAIRS*, Volume 264 of *Frontiers in Artificial Intelligence and Applications*, pp. 111–120. IOS Press.
- Dutta, A., C. Meilicke, and H. Stuckenschmidt (2015). Enriching structured knowledge with open information. In *Proceedings of the 24th International Conference on World Wide Web, WWW '15*, Republic and Canton of Geneva, Switzerland, pp. 267–277. International World Wide Web Conferences Steering Committee.
- Galárraga, L., G. Heitz, K. Murphy, and F. M. Suchanek (2014). Canonicalizing open knowledge bases. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, CIKM '14*, New York, NY, USA, pp. 1679–1688. ACM.
- Galárraga, L. A., C. Teflioudi, K. Hose, and F. Suchanek (2013). Amie: Association rule mining under incomplete evidence in ontological knowledge bases. In *Proceedings of the 22Nd International Conference on World Wide Web, WWW '13*, New York, NY, USA, pp. 413–422. ACM.
- Gurevych, I., J. Eckle-Köhler, S. Hartmann, M. Matuschek, C. M. Meyer, and C. Wirth (2012). Uby: A large-scale unified lexical-semantic resource based on lmf. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 580–590. Association for Computational Linguistics.

- Gurevych, I., J. Eckle-Kohler, and M. Matuschek (2016). *Linked Lexical Knowledge Bases: Foundations and Applications*. Morgan & Claypool Publishers.
- Joshi, M., E. Choi, D. S. Weld, and L. Zettlemoyer (2017). Triviaqa: A large scale distantly supervised challenge dataset for reading comprehension. *arXiv preprint arXiv:1705.03551*.
- Kinga, D. and J. B. Adam (2015). A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, Volume 5.
- Matuschek, M. (2015). *Word sense alignment of lexical resources*. Ph. D. thesis, Technische Universität.
- Mintz, M., S. Bills, R. Snow, and D. Jurafsky (2009). Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - Volume 2*, ACL '09, Stroudsburg, PA, USA, pp. 1003–1011. Association for Computational Linguistics.
- Navigli, R. and S. P. Ponzetto (2012). Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence* 193, 217–250.
- Nguyen, T. H. and R. Grishman (2015). Relation extraction: Perspective from convolutional neural networks. In *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, pp. 39–48.
- Riedel, S., L. Yao, and A. McCallum (2010). Modeling relations and their mentions without labeled text. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 148–163. Springer.
- Shi, L. and R. Mihalcea (2005). Putting pieces together: Combining framenet, verbnet and wordnet for robust semantic parsing. In *International conference on intelligent text processing and computational linguistics*, pp. 100–111. Springer.
- Sorokin, D. and I. Gurevych (2017). Context-aware representations for knowledge base relation extraction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 1784–1789.
- Vashishth, S., P. Jain, and P. Talukdar (2018). Cesi: Canonicalizing open knowledge bases using embeddings and side information. In *Proceedings of the 2018 World Wide Web Conference, WWW '18*, Republic and Canton of Geneva, Switzerland, pp. 1317–1327. International World Wide Web Conferences Steering Committee.
- Vrandečić, D. and M. Krötzsch (2014, September). Wikidata: A free collaborative knowledgebase. *Commun. ACM* 57(10), 78–85.
- Yates, A. and O. Etzioni (2009, March). Unsupervised methods for determining object and relation synonyms on the web. *J. Artif. Int. Res.* 34(1), 255–296.
- Zeng, D., K. Liu, Y. Chen, and J. Zhao (2015). Distant supervision for relation extraction via piecewise convolutional neural networks. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 1753–1762.