Learning to Transfer Relational Representations through Analogy

Gaetano Rossiello[*]
Department of Computer Science
University of Bari, Italy
gaetano.rossiello@uniba.it

Alfio Gliozzo
IBM Research AI
Yorktown Heights, NY, U.S.
gliozzo@us.ibm.com

Michael Glass
IBM Research AI
Yorktown Heights, NY, U.S.
mrglass@us.ibm.com

Abstract

We propose a novel approach to learn representations of relations expressed by their textual mentions. In our assumption, if two pairs of entities belong to the same relation, then those two pairs are analogous. We collect a large set of analogous pairs by matching triples in knowledge bases with web-scale corpora through distant supervision. This dataset is adopted to train a hierarchical siamese network in order to learn entity-entity embeddings which encode relational information through the different linguistic paraphrasing expressing the same relation. The model can be used to generate pre-trained embeddings which provide a valuable signal when integrated into an existing neural-based model by outperforming the state-of-the-art methods on a relation extraction task.

Introduction

Relation Extraction (RE) is the task of identifying semantic relationships between entities in unstructured textual corpora. RE is commonly addressed as a classification task, where a model is trained to classify relation mentions in text among a predefined set of relation types. Since a given relation can be expressed using different textual patterns surrounding entities, the RE models which follow this approach need a considerable amount of examples for each relation to reach satisfactory performance. Distant supervision (Riedel, Yao, and McCallum 2010) uses training examples from a knowledge base, guaranteeing a large amount of relation examples without human annotations, which can be used effectively by neural networks (Glass and Gliozzo 2018). However, even with this technique, the state-of-the-art RE systems present two limitations: (1) distantly supervised models are not accurate in extracting relations with a long-tailed distribution, because they typically have a small set of instances in knowledge bases; (2) in most domains, relation types are very specific and only a few examples of each relation are available.

To address these limitations, we propose a method able to encode information about the relations expressed in natural language into vector representations, that we call *analogy embeddings*, with the aim to integrate them in other RE models. Our assumption states that if two pairs of entities, (A, B) and (C, D), have at least one relation in common r,
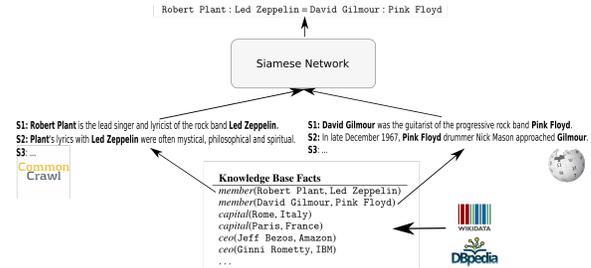
Figure 1: Learning relations by analogy through matching the facts from knowledge bases with web-scale textual corpora.

then those two pairs are analogous. Viceversa, solving proportional analogies, such as $A : B = C : D$, consists of identifying the implicit relations shared between two pairs of entities. Based on this idea, we propose a end-to-end neural model able to measure the degree of analogical similarity between two entity pairs, instead of predicting a confidence score for each relation type. An entity pair is represented through its mentions in a textual corpus, sets of sentences where entities in the pair co-occur. If a mention represents a specific relation type, then this relationship is expressed by the linguistic context surrounding the two entities. Thus, given two analogous entity pairs represented by their textual mentions sets as input, the model is trained to minimize the difference between the representations of relations having the same linguistic patterns. In other words, the model learns the different paraphrases expressing the same relation. We exploit alignments between knowledge bases and web-scale textual corpora to build a reliable set of analogous facts used as ground truth. Figure 1 provides an example of this process.

In our research hypothesis, a model trained in such way is able to recognize analogies between entity pairs by: (1) generalizing over the sequence of words in the mentions; (2) projecting the sequence of words in the mentions into a vector space representing relational semantics.

In our experiments, we integrate the pre-trained analogy embeddings into an existing end-to-end model based on convolutional networks (Glass and Gliozzo 2018), outperforming the state-of-the-art systems on two shared datasets commonly used for distantly supervised relation extraction.

Model

Since our goal is to train a model able to compute the relational similarity given two sets of textual mentions, we use siamese networks (Koch, Zemel, and Salakhutdinov 2015; Mueller and Thyagarajan 2016) to learn discriminative features between those two instances. A siamese network has symmetric twin sub-networks which share the same parameters, but are joined by an energy function at the head. Weight sharing forces the two similar instances to be mapped to very close locations in feature space because both of the sub-networks are optimized using the same function.

However, in our setting each instance consists of a set of mentions, therefore it is inherently a multi-instance learning task. We propose a Hierarchical Siamese Network (HSN) with an attention mechanism at both word level and at the mention level in order to select the textual mention which better describes the relation, where attentive Gated Recurrent Units (GRU) are used to encode each mention. Our model learns how to map patterns of words across the sentences containing two pairs of entities so as to capture the semantics of the relation. To the best of our knowledge, this is the first application of a siamese network by pairing sets of instances, so it can be considered a novelty of this work.

Experiments

We train our model on T-REX (ElSahar et al. 2018), a large scale alignment dataset between Wikipedia abstracts and Wikidata triples, having 685 unique relations. We select a balanced number of positive and negative examples out of the training split based on these rules: (1) for each relation, we randomly extract a set of 20 entity pairs; (2) out of this set, we generated all possible combinations, $\binom{20}{2} = 190$, as positive pairing examples; (3) for each combination, we create a negative example by randomly selecting an entity pair from another relation. After these steps, we collect a bucket of proportional analogy training examples. We iterate this process throughout the training phase by selecting a different buckets at each iteration to prevent overfitting. The training is monitored using a validation set build on the least frequent 40% relations.

Once the HSN model is trained, we use it as feature extractor to generate entity-entity vectors given sets of sentences as input. We first evaluate our pre-trained analogy embeddings on NYT-FB (Riedel, Yao, and McCallum 2010), a standard benchmark for distantly supervised relation extraction. We concatenate these embeddings to the penultimate layer of a relation extraction model, PCNN-KI (Glass and Gliozzo 2018), based on a convolutional neural network, which is the state-of-the-art for this benchmark. The final fully-connected layer uses the representation from HSN in combination with its own learned multi-instanced vector representation to predict a confidence score for each relation. During the training of this joint model, PCNN-KI+ANALOGY, we freeze our analogy embeddings to avoid the loose of the "knowledge transfer" capability.

Figure 2 reports the results of our evaluation. The model which uses the features generated by HSN largely improve the performances of PCNN-KI, despite the HSN is trained
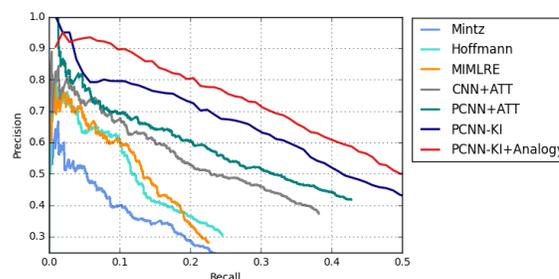

Figure 2: Precision-Recall curves comparison on NYT-FB.

on a different corpus and using a different knowledge base. In the same chart, we also report a compared evaluation for other approaches proposed in the literature for this dataset.

	AUC	F1
PCNN-KI	0.437	0.468
PCNN-KI+ANALOGY	**0.500**	**0.504**

Table 1: Area Under Curve (AUC) and F1 results on CC-DBP.

Finally, we run the evaluation also on CC-DBP (Glass and Gliozzo 2018), a larger dataset for distantly supervised RE, and the results confirm the improvements obtained by PCNN-KI model if it integrates our pre-trained embeddings. The results in terms of Area Under Curve (AUC) and F1 are reported in Table 1.

Conclusion and Future Work

In this work, we proposed a novel approach to learn representations of relations in text. Alignments between knowledge bases and textual corpora are used as ground truth in order to collect a set of analogies between entity pairs. We designed a hierarchical siamese network trained to recognize those analogies. The model can generate low-rank representations can help existing neural-based models designed for other tasks. As future work, we plan also to explore the use of analogy embeddings in other tasks, such as reading comprehension, question answering, and knowledge base population.

References

ElSahar, H.; Vougiouklis, P.; Remaci, A.; Gravier, C.; Hare, J. S.; Laforest, F.; and Simperl, E. 2018. T-rex: A large scale alignment of natural language with knowledge base triples. In *LREC*.

Glass, M., and Gliozzo, A. 2018. Discovering implicit knowledge with unary relations. In *ACL*.

Koch, G.; Zemel, R.; and Salakhutdinov, R. 2015. Siamese neural networks for one-shot image recognition. In *ICML Deep Learning Workshop*.

Mueller, J., and Thyagarajan, A. 2016. Siamese recurrent architectures for learning sentence similarity. In *AAAI*.

Riedel, S.; Yao, L.; and McCallum, A. 2010. Modeling relations and their mentions without labeled text. In *ECML PKDD*.