

# Unsupervised Open Relation Extraction

Hady Elsahar<sup>1</sup>(✉), Elena Demidova<sup>2</sup>, Simon Gottschalk<sup>2</sup>, Christophe Gravier<sup>1</sup>,  
and Frederique Laforest<sup>1</sup>

<sup>1</sup> Univ Lyon, UJM-Saint-Etienne, CNRS, Laboratoire Hubert Curien, Lyon, France  
{hady.elsahar,christophe.gravier,frederique.laforest}@univ-st-etienne.fr

<sup>2</sup> L3S Research Center, Leibniz Universität Hannover, Hannover, Germany  
{demidova,gottschalk}@L3S.de

**Abstract.** We explore methods to extract relations between named entities from free text in an unsupervised setting. In addition to standard feature extraction, we develop a novel method to re-weight word embeddings. We alleviate the problem of features sparsity using an individual feature reduction. Our approach exhibits a significant improvement by 5.8% over the state-of-the-art relation clustering scoring a F1-score of 0.416 on the NYT-FB dataset.

**Keywords:** Relation extraction · Word embedding · NLP

## 1 Introduction

Relation extraction (RE) is the task of identification and classification of relations between named entities (such as persons, locations or organizations) in free text. RE is of utmost practical interest for various fields including event detection, knowledge base construction and question answering. Figure 1 illustrates a typical RE task. For the first two sentences, RE should identify the semantic relation type *birth place* between the named entity pairs regardless of the surface pattern used to express the relation such as *hometown is* or *was born in*. RE should also distinguish it from the album production relation between the same named entities in the third sentence.

Distant supervision techniques for RE [1, 4] have proven to be very efficient in solving that problem. However, distant supervision is limited to a fixed set of relations in a given knowledge base, which hinders its adaptation to new domains. Unsupervised approaches [3, 7] can potentially overcome these limitations by applying purely unsupervised methods enabling extraction of open relations (relations unknown in the knowledge base in advance). In this paper, we propose an unsupervised approach to extract and cluster open relations between named entities from free text by re-weighting word embeddings and using the types of named entities as additional features.

1. *David Bowie's hometown is London*, United Kingdom.
2. *Axel Rose*, also known as "William Bruce", was born in Lafayette, Indiana.
3. *David Bowie* produced his first album in London, United Kingdom.

Fig. 1. Sentences containing textual relations between *named entities*.

## 2 Proposed Method

Our system builds sentence representations based on the types of the involved named entities, and the terms forming the relations. For the latter, we use pre-trained word embeddings after re-weighting them according to the dependency path between the named entities. These representations are clustered so that different representations of the semantically equivalent relations are mapped to the same cluster. Figure 2 presents an overview of our system for unsupervised open relation extraction, consisting of four stages: preprocessing, feature extraction, sparse feature reduction and relation clustering described in the following.

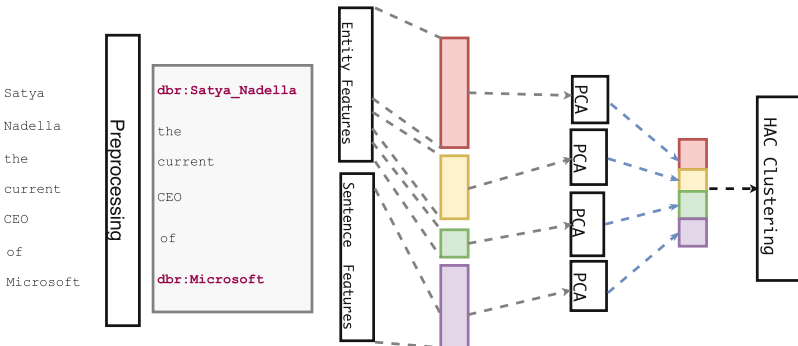


Fig. 2. System overview

**Preprocessing.** For each sentence in the dataset, we extract named entities using DBpedia Spotlight and consider all sentences containing at least two entities. For this set of sentences, the Stanford CoreNLP dependency parser is utilized to extract the lexicalized dependency path between each pair of named entities.

**Feature Extraction.** For each sentence, our method outputs a vector representation of the textual relation between each named entity pair. Features include word embeddings, dependency paths between named entities, and named entity types. Word embeddings provide an estimation of the semantic similarity between terms using vector proximity. Sentence representations are typically built by averaging word vectors. However, not all words in a sentence equally

contribute to the expression of the relation between two named entities. Therefore we develop a novel method to re-weight the pre-trained word embeddings. Terms that appear within the lexicalized dependency path between the two named entities are given a higher weight. Intuitively, shorter dependency paths are more likely to represent true relationships between the named entities. The vector representation  $s(W, D)$  of each sentence is calculated through the following function:

$$s(W, D) = \sum_{w_i \in W} f(w_i, W, D) \cdot v(w_i), \quad f(w_i, W, D) = \begin{cases} \frac{C_{in} \cdot |W|}{|D|}, & \text{if } w_i \in D \\ C_{out}, & \text{otherwise} \end{cases},$$

where  $W = \{w_1, \dots, w_n\}$  is the set of terms in the sentence,  $D \subset W$  is the set of terms in the lexicalized dependency path between the named entities in the sentence, and  $v(w_i)$  is the pre-trained word embedding vector for  $w_i$ .  $C_{in} \geq 1$  and  $C_{out}$  are constant values experimentally set to 1.85 and 0.02. We use Glove<sup>1</sup> word embeddings of size 100. As a baseline, we compare these representations with standard sentence representations features such as: TF-IDF, the sum of word embeddings, and the sum of IDF re-weighted word embeddings [5]. Intuitively, relations can connect entities of certain types. For example, a birth place relation connects a person and a location, although other relations between person and location are possible. Therefore, for each named entity, we use its DBpedia types and Stanford NER tags as features.

**Sparse Feature Reduction.** Some of the features are more sparse than the others; concatenating them for each relation skews the clustering. In supervised relation extraction, this is not an issue as any learning algorithm is expected to do feature selection automatically using the training data. In unsupervised relation extraction there is no training data, hence we devise a novel strategy in order to circumvent the sparse features bias. Individual feature reduction of the sparse features is applied before merging them with the rest of the feature vectors. For feature reduction, we use Principal Component Analysis (PCA) [2].

**Relation Clustering.** We use Hierarchical Agglomerative Clustering (HAC) to cluster the feature representations of each relation, with Ward’s [6] linkage criteria<sup>2</sup>, which yields slightly better results than the k-means clustering algorithm.

### 3 Evaluation

To evaluate our system, we use the NYT-FB dataset [3]. This dataset contains approximately 1.8M sentences divided into 80%–20% test-validation splits and aligned automatically to the statements (triples) from Freebase. The alignment between sentences and the properties of the Freebase triples in this dataset is considered as the gold standard for the relation clustering algorithm.

<sup>1</sup> <http://nlp.stanford.edu/projects/glove/>.

<sup>2</sup> Accessing the clustering output by HAC at rank  $k$  giving  $k$  clusters.

We use the validation split to tune the parameters for re-weighting word vectors and the PCA algorithm, and the test set for evaluating relation discovery methods. We compare our method using the best identified feature combination with the state-of-the-art models for unsupervised Relation Discovery, namely the variational autoencoders model [3] and two other systems, Rel-LDA [7], and Hierarchical Agglomerative Clustering (HAC) baseline with standard features [8]. To make our results comparable we set the number of relations to induce (number of clusters  $k$ ) to 100, following the SOA systems.

Table 1 shows the performance of the clustering algorithm by relying only on sentence representations as features. Results demonstrate that our method of word embeddings re-weighted by the dependency path shows a significant improvement over other traditional sentence representations. Table 2 shows the performance when the dependency re-weighted word embeddings are merged with the rest of the proposed features and applying individual feature reduction. Our method outperforms the state-of-the-art relation discovery algorithm scoring a pairwise F1 score of 41.6%.

**Table 1.** Comparison between different features for clustering.

Feature	F <sub>1</sub>
TF-IDF	12.2
Word-Emb.	7.4
IDF-Emb.	10.3
Dependency Re-Weighted Emb.	<b>19.5</b>

**Table 2.** Pairwise F<sub>1</sub> (%) scores of different models on the test set of the NYT-FB dataset.

Var. Autoencoder	Rel-LDA	HAC	Our
35.8	29.6	28.3	<b>41.6</b>

## 4 Conclusion

In this paper, we proposed an approach for unsupervised relation extraction from free text. Our approach is based on a novel method of re-weighting word vectors according to the dependency parse tree of the sentence. As additional features, we use the types of named entities involved in the relations. A final HAC clustering is applied to the sentence representations so that similar representation of a relation are mapped to the same cluster. Our evaluation results demonstrate that our method outperforms the state-of-the-art relation clustering method by 5.8% pairwise F1 score. The code for feature building and dimensionality reduction is publicly available<sup>3</sup>.

<sup>3</sup> <https://github.com/hadyelsahar/relation-discovery-2-entities.git>.

**Acknowledgements.** This work was partially funded by H2020-MSCA-ITN-2014 WDAqua (64279), ALEXANDRIA (ERC 339233) and Data4UrbanMobility (BMBF).

## References

1. Augenstein, I., Maynard, D., Ciravegna, F.: Distantly supervised web relation extraction for knowledge base population. *Semant. Web* **7**(4), 335–349 (2016)
2. Jolliffe, I.T.: Principal component analysis. In: Lovric, M. (ed.) *International Encyclopedia of Statistical Science*, pp. 1094–1096. Springer, Heidelberg (2011). doi:[10.1007/978-3-642-04898-2\\_455](https://doi.org/10.1007/978-3-642-04898-2_455)
3. Marcheggiani, D., Titov, I.: Discrete-state variational autoencoders for joint discovery and factorization of relations. *Trans. ACL* **4**, 231–244 (2016)
4. Mintz, M., Bills, S., Snow, R., Jurafsky, D.: Distant supervision for relation extraction without labeled data. In: *Proceedings of ACL 2009*, pp. 1003–1011 (2009)
5. Rei, M., Cummins, R.: Sentence similarity measures for fine-grained estimation of topical relevance in learner essays. In: *Proceedings of the BEA Workshop* (2016)
6. Ward Jr., J.H.: Hierarchical grouping to optimize an objective function. *J. Am. Statist. Assoc.* **58**(301), 236–244 (1963)
7. Yao, L., Haghghi, A., Riedel, S., McCallum, A.: Structured relation discovery using generative models. In: *Proceedings of EMNLP 2011* (2011)
8. Yao, L., Riedel, S., McCallum, A.: Unsupervised relation discovery with sense disambiguation. In: *Proceedings of ACL 2012* (2012)