Unsupervised relation extraction using sentence encoding

Manzoor Ali[1], Muhammad Saleem[2], and Axel-Cyrille Ngonga Ngomo[1]

[1] DICE group, Department of Computer Science, Paderborn University
manzoor@campus.uni-paderborn.de, axel.ngonga@upb.de
[2] AKSW, University of Leipzig, Germany
saleem@informatik.uni-leipzig.de

Abstract. Relation extraction between two named entities from unstructured text is an important natural language processing task. In the absence of labelled data, semi-supervised and unsupervised approaches are used to extract relations. We present a novel approach that uses sentence encoding for unsupervised relation extraction. We use a pre-trained, SBERT based model for sentence encoding. Our approach classifies identical sentences using a clustering algorithm. These sentences are used to extract relations between two named entities in a given text. The system calculates a confidence value above a certain threshold to avoid semantic drift. The experimental results show that without any explicit feature selection and independent of the size of the corpus, our proposed approach achieves a better F-score than state-of-the-art unsupervised models.

1 Introduction

Relation extraction (RE) is a salient Natural Language Processing (NLP) task, which aims to extract the semantic relation between two named entities from natural language text. Relation extraction plays an essential role for many NLP applications such as question answering systems, knowledge bases creation and completion [9], etc. Supervised approaches require labelled data for relation extraction, which is an expansive and tedious task. In the absence of labelled training data, unsupervised approaches are used to extract relations from natural language text.

State-of-the-art (SOTA) unsupervised approaches use different strategies such as word embeddings [5], entity-type information [10], convolutional neural network [8], etc. to extracts relations from unlabeled corpora. These approaches may fail to extract the correct and complete relations. For example, *"... Nephew George P. Bush – son of Florida Gov. Jeb Bush..."* from this sentence, the word embedding approach will extract *birthplace(George P. Bush, Florida)* relation. In contrast, the actual relations in this sentence are *governorOf(Jeb Bush, Florida)* and *sonOf(Jeb Bush, George P. Bush)*. In such situations, it is mandatory to know the context of the sentence instead of only using word embedding. Therefore, we use sentence encoding for relation extraction.

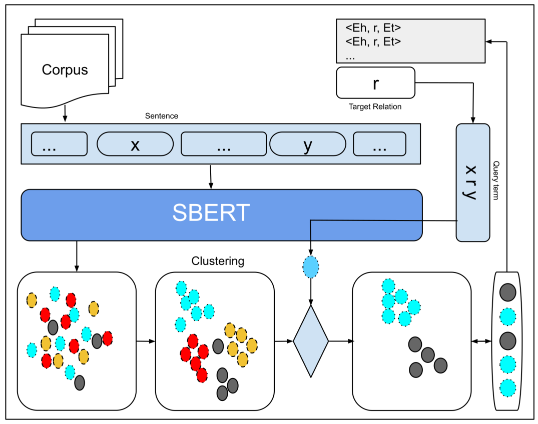

Fig. 1. The system Architecture

To the best of our knowledge, BERT-based [4] *sentence encoding* is yet not used for *unsupervised* relation extraction. This is because BERT-based sentence encoding and similarity is computationally expensive [7]. We propose a novel unsupervised approach dubbed **US-BERT** that uses the BERT-based sentence encoding [7] on a corpus that is already annotated for named entities (NE).

Our main contributions are as follows: we do not rely on any explicit feature selection for relation extraction; we achieve (SOTA) results for unsupervised relation extraction.

2 Our Approach

Figure 1 describes the US-BERT architecture. Input to our system is a target relation, named entities annotated corpus, and the two named entity types for the target relation. The system outputs all those sentences that include a target relation.

Our approach comprises four main modules. The candidate sentences selection module chooses all those sentences that contain two already specified named entity types. A sentence can also contain other entities and entity types. The sentence encoding module uses sentence-BERT to calculate 768-dimensional sentence encoding of the selected sentences. The clustering module creates clusters from similar vector representations of sentences. The relation extraction module uses a verb form to choose the most relevant cluster and extract semantically identical sentences based on cosine similarity.

Candidate sentences selection: The proposed approach chooses candidate sentences based on NE types for a particular relation for reducing the computation time. A set of candidate sentences is created from all annotated (named entity types annotation) sentences that include E_h and E_t, and the type of E_h and E_t is according to a particular relation. E_h and E_t represent the subject entity and object entity, respectively. For example, for a relation `birthPlace`, we only filter those sentences that include entity type Person (PER) and Location (LOC). Before sentence encoding, both entities E_h and E_t, are removed from the sentence.

Sentence encoding: We consider the SBERT based pre-trained model that achieves SOTA performance for sentence encoding [7]. SBERT uses siamese and triplet network to produce sentences that are semantically meaningful and also comparable for cosine-similarity. We used pre-trained SBERT model, *distilbert-base-nli-stsb-mean-tokens*. It is trained on 570,000 sentences.

Clustering: To aggregate semantically similar vectors for a particular relation, we performed clustering. Clustering combines similar vectors and reduces the number of sentences for comparison to the query term. The system only selects those clusters that are semantically close to the query term.

Unsupervised algorithms like K-Mean and K-medoids require manual selection of the number of clusters. In relation extraction, two entities can have a variable number of relations in the real world. Therefore, we choose Affinity propagation for clustering. Affinity propagation does not need the number of cluster to be specified in advance. It selects an exemplar vector and creates a cluster around the exemplar.

Query encoding and relation extraction: We adopt a query-based approach to extract a relation from a cluster. In our approach, a query is a sentence that contains two entities X and Y. Also, the query contains a relationship in phrasal verb form. For example, the complete query we use for the relation birthPlace is "X born in Y". We use sentence encoding to convert the query to a vector representation. We compute the cosine similarity between all the centroids and the query vector. If the cosine similarity between the centroid of a cluster and query vector crosses the threshold value, we select that cluster for further computation.

To increase recall and avoid semantic drift, we use two iterations. Semantic drift is the change in the actual meaning of a word with time [3]. In the first iteration, the system selects only those vectors from a cluster with a high cosine similarity score to the query term q. V_p represents the selected vectors in the first iteration. While in the second iteration, those vectors are selected that have high similarity with the list of selected vectors (V_p) in the first iteration.V_s represent the selected vectors in the second iterations. This two-step iteration increases the recall but sometimes causes semantic drift. To avoid semantic drift, we use a threshold value in the first iteration. In the second iteration, we score the vectors according to the following equation and only select those vectors with a P_{score} higher than zero.

$$P_{score} = Cosine(V_s, q) - (1 - Cosine(V_p, V_s))^2 \tag{1}$$

Sometimes the sentences include semantically similar meanings, but the actual relation exists far from the two entities occurrences that cause a decrease in the precision[2]. To address this issue and increase the precision, our model uses a window-based approach like Snowball [1] to minimize the false-positive. The window consists of words around two entities, `Before` E_h, `Between` E_h and E_t, and `After` E_t. The system creates vectors representing the selected tokens using the sentence encoding module and finding the cosine similarity with the initially selected query term. Only those vectors are filtered, which have a higher score than a certain threshold.

3 Evaluation

We evaluate our proposed approach on the NYT-FB [6] dataset. The NYT-FB dataset is extracted from New York Times articles and aligned with freebase. The NYT-FB dataset consists of 253 relations. The initial evaluation result of our approach with the (SOTA) unsupervised systems is shown in Table 1. We run RelLDA1 on their reported parameter on the NYT-FB dataset for only StanfordNER based annotated sentences. In contrast, we run the other two models Simon and EType+, for both StanfordNER and AllenNLP NER.

Table 1. Precision (P) Recall (R) and F1 score of different systems using two NER annotation techniques on the NYT-FB.

	StanfordNER			**AllenNLP NER**		
Models	P	R	F1	P	R	F1
RelLDA1	0.30	0.47	0.36	-	-	-
Simon	0.32	0.50	0.39	0.334	0.497	0.399
EType+	0.30	**0.62**	**0.40**	0.31	**0.64**	0.417
US-BERT	**0.35**	0.45	0.39	**0.38**	0.61	**0.468**

Our proposed model (US-BERT) outperforms all the models in precision for StanfordNER based annotated sentences. However, the overall F1 score is less than the EType+. For AllenNLP NER based annotated sentences, our system achieves the highest F1 score. One of the reasons for the low recall we observed is the NER system. Wrongly annotated sentences reduce the recall. This did not penalize US-BERT since all evaluated systems were used with the same NER pre-processing steps.

4 Conclusion and Future Work

We used pre-trained sentence encoding to extract high-quality relations without any explicit feature's selection. We achieved the best F1, and precision score compares to the (SOTA) unsupervised methods. To further investigate the relation extraction, we will use some feature selection, compare the results with our work and see the impact also, we will compare our approach with some other (SOTA) approaches in our future work, mainly to the relation extraction systems based on language models.

Acknowledgement

This work has been supported by the EU H2020 Marie Skłodowska-Curie project KnowGraphs (860801), the BMBF-funded EuroStars projects E!113314 FROCKG (01QE19418) and E! 114154 PORQUE (01QE2056C).

References

1. Agichtein, E., Gravano, L.: <i>snowball</i>: Extracting relations from large plain-text collections. In: Proceedings of the Fifth ACM Conference on Digital Libraries. p. 85–94. DL '00, Association for Computing Machinery, New York, NY, USA (2000). https://doi.org/10.1145/336597.336644, https://doi.org/10.1145/336597.336644
2. Batista, D.S., Martins, B., Silva, M.J.: Semi-supervised bootstrapping of relationship extractors with distributional semantics. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. pp. 499–504 (2015)
3. Blank, A.: Why do new meanings occur? A cognitive typology of the motivations for lexical semantic change, pp. 61–90. De Gruyter Mouton (2013). https://doi.org/doi:10.1515/9783110804195.61, https://doi.org/10.1515/9783110804195.61
4. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding (2019)
5. Elsahar, H., Demidova, E., Gottschalk, S., Gravier, C., Laforest, F.: Unsupervised open relation extraction. In: Blomqvist, E., Hose, K., Paulheim, H., Ławrynowicz, A., Ciravegna, F., Hartig, O. (eds.) The Semantic Web: ESWC 2017 Satellite Events. pp. 12–16. Springer International Publishing, Cham (2017)
6. Marcheggiani, D., Titov, I.: Discrete-state variational autoencoders for joint discovery and factorization of relations. TACL **4**, 231–244 (2016), https://www.aclweb.org/anthology/Q16-1017
7. Reimers, N., Gurevych, I.: Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084 (2019)
8. Simon, É., Guigue, V., Piwowarski, B.: Unsupervised information extraction: Regularizing discriminative approaches with relation distribution losses. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. pp. 1378–1387. Association for Computational Linguistics, Florence, Italy (Jul 2019). https://doi.org/10.18653/v1/P19-1133, https://www.aclweb.org/anthology/P19-1133
9. Surdeanu, M., Tibshirani, J., Nallapati, R., Manning, C.D.: Multi-instance multi-label learning for relation extraction. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. pp. 455–465. Association for Computational Linguistics, Jeju Island, Korea (Jul 2012), https://www.aclweb.org/anthology/D12-1042
10. Tran, T.T., Le, P., Ananiadou, S.: Revisiting unsupervised relation extraction. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 7498–7505. Association for Computational Linguistics, Online (Jul 2020). https://doi.org/10.18653/v1/2020.acl-main.669, https://www.aclweb.org/anthology/2020.acl-main.669