

How to Adapt Deep Learning Models to a New Domain: The Case of Biomedical Relation Extraction

Cómo adaptar un modelo de aprendizaje profundo a un nuevo dominio: el caso de la extracción de relaciones biomédicas

Jefferson A. Peña-Torres ¹, Raúl E. Gutiérrez ²,
Víctor A. Bucheli  y Fabio A. González ⁴

Recibido: 25 de septiembre de 2019

Aceptado: 19 de noviembre de 2019

Cómo citar / How to cite

J. A. Peña-Torres, R. E. Gutiérrez, V. A. Bucheli, F. A. González, "How to Adapt Deep Learning Models to a New Domain: The Case of Biomedical Relation Extraction", *TecnoLógicas*, vol. 22, pp. 49-62, 2019. <https://doi.org/10.22430/22565337.1483>



- ¹ Ingeniero de Sistemas, Escuela de Ingeniería de Sistemas y Computación, Universidad del Valle, Cali- Colombia, jefferson.amado.pena@correounivalle.edu.co
- ² PhD en Ingeniería, Escuela de Ingeniería de Sistemas y Computación, Universidad del Valle, Cali- Colombia, raul.gutierrez@correounivalle.edu.co
- ³ PhD en Ingeniería, Escuela de Ingeniería de Sistemas y Computación, Universidad del Valle, Cali- Colombia, victor.bucheli@correounivalle.edu.co
- ⁴ PhD en Ingeniería, Departamento de Ingeniería de Sistemas e Industrial, Universidad de Nacional de Colombia, Bogotá- Colombia, fagonzalezo@unal.edu.co

Abstract

In this article, we study the relation extraction problem from Natural Language Processing (NLP) implementing a domain adaptation setting without external resources. We trained a Deep Learning (DL) model for Relation Extraction (RE), which extracts semantic relations in the biomedical domain. However, can the model be applied to different domains? The model should be adaptable to automatically extract relationships across different domains using the DL network. Completely training DL models in a short time is impractical because the models should quickly adapt to different datasets in several domains without delay. Therefore, adaptation is crucial for intelligent systems, where changing factors and unanticipated perturbations are common. In this study, we present a detailed analysis of the problem, as well as preliminary experimentation, results, and their evaluation.

Keywords

Semantic Extraction, Deep Learning, Relation Extraction, Natural Language Processing.

Resumen

En este trabajo estudiamos el problema de extracción de relaciones del Procesamiento de Lenguaje Natural (PLN). Realizamos una configuración para la adaptación de dominio sin recursos externos. De esta forma, entrenamos un modelo con aprendizaje profundo (DL) para la extracción de relaciones (RE). El modelo permite extraer relaciones semánticas para el dominio biomédico. Sin embargo, ¿El modelo puede ser aplicado a diferentes dominios? El modelo debería adaptarse automáticamente para la extracción de relaciones entre diferentes dominios usando la red de DL. Entrenar completamente modelos DL en una escala de tiempo corta no es práctico, deseamos que los modelos se adapten rápidamente de diferentes conjuntos de datos con varios dominios y sin demora. Así, la adaptación es crucial para los sistemas inteligentes que operan en el mundo real, donde los factores cambiantes y las perturbaciones imprevistas son habituales. En este artículo, presentamos un análisis detallado del problema, una experimentación preliminar, resultados y la discusión acerca de los resultados.

Palabras clave

Extracción semántica, Aprendizaje profundo, Extracción de relaciones, Procesamiento de lenguaje natural.

1. INTRODUCTION

In this study, we address the Relation Extraction (RE) problem as follows: For a given sentence S , the RE problem is a classification problem, where the goal is to predict a semantic relation r between e_1 and e_2 , both entities in S , following previous research, mainly [1], as our baseline model. RE plays a key role in information extraction from unstructured text, and it has a wide range of applications in many domains [2]-[4].

The rapid growth of unstructured text data and the valuable knowledge recorded in them has generated considerable interest in automatic detection and extraction of semantic relations [5].

Although many studies have been conducted to develop supervised relation extraction models [6]-[9], neural network-based approaches have been proposed for relation extraction, in particular with deep learning, e.g., Recursive Neural Networks [10]-[12], Recurrent Neural Networks (RNNs) [13]-[15], and Convolutional Neural Networks (CNNs) [1], [2], [16]-[18].

Deep Learning (DL) has demonstrated its efficiency in improving the RE task.

Specifically regarding relations in English language, the deep learning (DL) models have been trained with little or no domain knowledge, and several studies have implemented DL methods for relation extraction from texts. However, depending on the language and domain of deep learning models for relation extraction, the following challenges may arise: (a) a lack of training samples in some languages and domains and (b) the generalization of model in a domain with different types of relations.

Against this backdrop, some researchers have been successful in performing RE for a specific domain. They have utilized large amounts of labelled data. However, there are insufficient labelled data for certain domains and languages. Therefore, domain adaptation,

domain shift, domain bias, and domain transfer are used to perform relation extraction on an unseen target domain or language. However, the factors and conditions that are appropriate for training and testing DL models with different types of datasets in a target domain or language should be explored. Therefore, transferring well-trained DL models to other domains remains a challenge.

This paper presents a baseline DL model [1] and an experiment conducted using multiple representations from the biomedical domain. To address both challenges mentioned above, we describe the impact of some biomedical datasets, the generalization capability of the deep learning model, and compare its performance when some representations of the baseline model are modified for the biomedical domain.

This paper is organized as follows.

Section 2 presents a DL model based on [1], [16] for relation extraction, considering the potential of DL for RE tasks in bioinformatics research (biology, biomedicine, and healthcare). Section 3 describes a suitable way to adapt a well-trained model to the biomedical domain.

Section 4 details the experimental setup and the evaluation of the baseline model on multiple public PPI, DDI and CPI corpora. The last two sections discuss and summarize the representations of the impacts and the behavior of the baseline model on the datasets.

2. RELATED WORK

Traditionally, Relation Extraction (RE) has been a classification problem that occurs between two or more named entities in the same sentence that have a semantic relationship. Depending on the number of semantic relation classes, RE tasks can be binary or multi-class. In this study, we considered a binary relation extraction task in the biomedical domain. [19]-[22].

Said task achieves a high performance with supervised approaches; however, it needs annotated data, which is time consuming and entails intensive human labor. Recently, models based on deep neural networks, such as CNNs and RNNs, have shown promising results for RE.

For example, in [1], the authors explored RE without exhaustive pre-processing. They employed a CNN and observed that any automatically learned features yielded promising results and could potentially replace the manually designed features.

In turn, in [10], [23] other authors proposed a DL approach with a RNN architecture and a matrix-vector word representation to explore the impact of the lack of explicit knowledge about the type of relation. Likewise, a study [11] compared the capabilities of CNN and RNN for the relation classification task. We reviewed other articles on relation extraction without specific domain. We also surveyed some models in the literature classified by network architecture and dataset. We hope this survey provides an overview to select a baseline model.

In Table 1, the baseline is marked in boldface, the CNN and RNN-based Models are learned on SemEval 2010, and the category Others is relevant for Deep Learning and RE tasks.

Most of the studies we reviewed are concerned with English language, and their models have not been extended to other languages or domains. There is a variety of possible relations between domains and languages, characterized by their own syntactic and lexical properties.

Nevertheless, the notion of a relation, what it “means”, is inherently ambiguous [35]. Many efforts have been devoted to biomedical relation extraction, whose goal is to discover valuable knowledge about proteins, drugs, diseases, genes, adverse effects, and other biological interactions from unstructured free text [36].

3. METHODOLOGY

In this section, we introduce a model architecture based on Convolutional Neural Networks (CNN) and modifications that allow other representations.

3.1. Model

Based on the literature review above, with better performance architecture for RE tasks is CNN [37]. The model used in this paper contains three components: (a) a convolutional layer with multiple windows sizes, (b) a max pooling layer, and (c) a fully connected layer with dropout and softmax. Additionally, the input data undergoes a pre-transformation to vector representation. The transformation from words to vector representations has been described by several authors [38], [39].

We considered the model in [1] as baseline model, and trained it to automatically find relevant information (features or patterns) in a source sentence to predict semantic relations. We also added several modifications to keep some key elements of the baseline model that are consistent and consequent with a RE task. The diagram of the baseline model is shown in Fig. 1.

CNN-based architectures learn semantic information from sentences in the hidden layers during training. Although the extraction of semantic information is not previously known, the convolution layers learn features in the representations of the source domain. In the baseline model, let S be an input sentence that could be represented as $S = \{w_1, w_2, w_3 \dots w_n\}$, where w_i is the i th word in S ; and let V be the vocabulary size of each dataset and V_{xd} , the embedding matrix with a d dimensional vector from pre-trained word embeddings.

Since we aim to compare the modifications and addition of representations to the baseline model in the biomedical domain, we used different kernel sizes and softmax function modifications; afterward, we included multi representations that captured different characteristics from the input. All

of them were gradually changed. Then, for each w_i in S , the distributional and not distributional representation was obtained and concatenated in a vector. As result, a matrix representing S was processed by the model in order to perform the classification.

Table 1. Survey of studies into RE tasks using Deep Learning approaches. Source: Created by the authors.

CNN-based Models	
MODEL	F1
CNN [1]	82.7
CR-CNN [17]	84.1
CNN[16]	82.8
Attention- CNN [24]	85.4
depLCNN + NS [18]	85.6
Multi-Attention CNN [25]	88.0
Selection-Attention + CNN[26]	88.0
RNN-based Models	
MV-RNN	82.4
Entity-Att BLSTM[10]	85.2
Hierarchical Attention BILSTM [27]	84.3
Attention BILSTM[28]	84.0
BILSTM [14]	82.7
Others	
BRNN [29]	86.3
DRNN [30]	86.1
SDP-LSTM [13]	83.7
DepNN [31]	83.6
FCN [32]	83.0
RCNN [33]	96.5
PCNN [34]	84.0

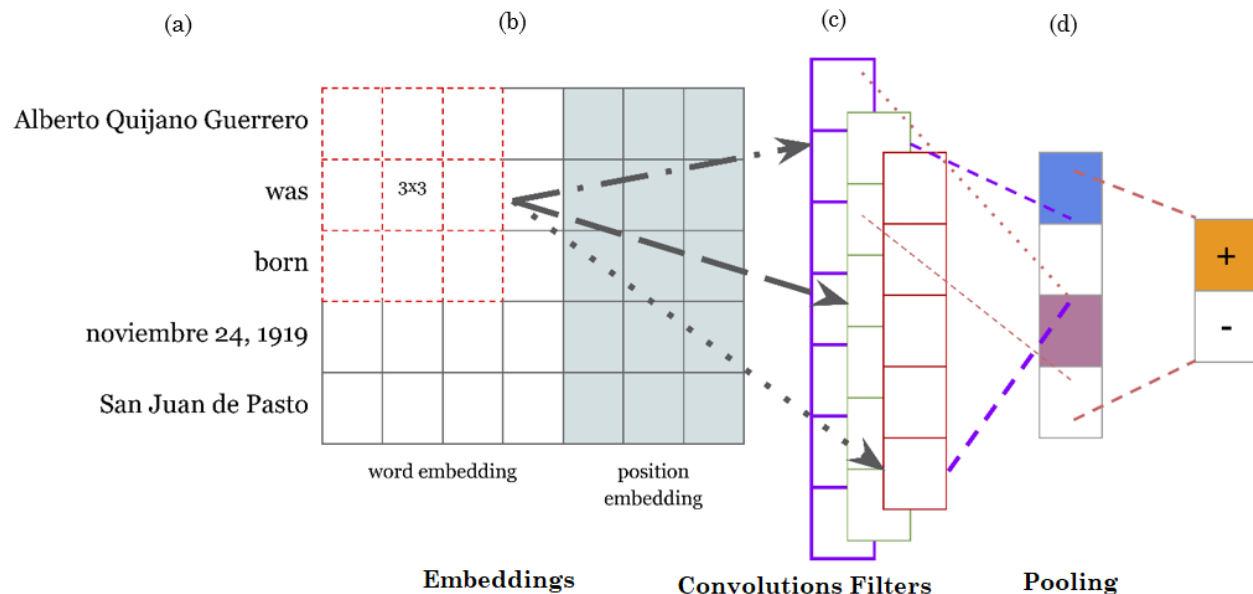


Fig. 1. Diagram of the CNN-based model architecture implemented in this case study. (a) Words from a sample sentence represented as vectors (Input layer). (b) Vector representation where convolutional kernels are applied (Convolutional layer). (c) Vector pooling resulting from convolutional kernel. (d) Max values representative of the relation. Source: Created by the authors.

This is a crucial part of our study in order to extract semantic relations, we needed to change the preprocessing, context length, paddings, kernel sizes, and validate implementation of baseline model.

Moreover, each word vector was supported by its corresponding information.

3.2. Representations

Representations have been effective tools to address the growing interest in DL for NLP tasks. While classical techniques used feature engineering and exploration to provide a more qualitative assessment and analysis of results from the point of view of computational linguistics, DL models learn features automatically.

In this paper, before S is processed by the baseline model, it is transformed in the form of a vector to capture different characteristics of the token; nevertheless, more information could be obtained from sentences to enhance automatic characterization via CNNs.

Multi-representation in DL models must be robust, and they should perform a

satisfactory relation extraction in similar tasks across different domains. We used the following representations to add characteristic elements of sentences.

Word Embedding: It is employed to capture syntactic and semantic meanings of words in distributed representations.

In characterized by their own syntactic and lexical sentence S , every word w_i is represented by a real-valued vector.

These word representations are encoded in an embedding matrix X^d , where V is a fixed-sized vocabulary.

Unfortunately, said word representations usually take a long time to train, and freely available trained word embeddings are commonly implemented [40]. We used pre-trained word2vec [38], Glove [39], and FastText [41] to conduct the experiments.

Position Embedding: In RE tasks, the words close to entities are usually informative and determine the relation between entities. We prove the relative position of words an entities similar to [1].

We used the relative position of both entity pairs. Apparently, it is not possible to capture such structural information only

through semantic and syntactic word features. It is necessary to specify which input tokens are the target nouns in the sentence and where they are placed.

The position characterized by their own syntactic and lexical of entities is a relative distance, which is also mapped to vector representations.

4. EXPERIMENTS

These experiments are intended to show that DL models (baseline) for Relation Extraction (RE) can be adapted to another domain using multi-representation. First, we introduce the datasets and the metrics to evaluate precision: recall and f1-score. Next, we describe the parameters of the baseline model, the evaluation of the multi-representation, its effects, and performance on the data. Finally, we compare the performance of the baseline model with the modified model.

4.1. Datasets in the biomedical domain

In this study, we explored RE tasks focused on the biomedical domain, especially relations such as protein-protein interactions (PPIs), drug-drug interactions (DDIs), and chemical-protein interactions (CPIs). Several scarce resources were utilized to adapt a pre-trained model.

We used three annotated corpora in the biomedical domain. All of them are publicly available and detailed below.

To extract semantic relations regarding Adverse Drug Effects, a subtask of DDI

was applied to the corpus ADE-EX, as follows. The sentence “*we report two cases of pseudoporphyria caused by naproxen and oxaprozin*” contains a semantic relation of the type Adverse Drug Effect between *pseudoporphyria* and *oxaprozin*.

In turn, in BioInfer, we used the Protein-Protein interaction task to find semantic relations in the sentence “*snf11 a new component of the yeast snf-swi complex that interacts with a conserved region of snf2*”, where *snf11* and *snf2* are two named entities that represent proteins.

Likewise, in the corpus ChemProt, Chemical-Protein interactions are annotated. For example, in the sentence “*Discovery of novel 2-hydroxydiarylamide derivatives as TMPRSS4 inhibitors*”, *2-hydroxydiarylamide* is a chemical and *TMPRSS4* is a protein with a semantic relation to said chemical. For protein-protein interactions (relations), we used the BioInfer dataset [42]. For adverse drug events, we used the ADE corpus [43]; and, for chemical-protein interactions, we used the ChemProt corpus [44]. The detailed information of each dataset is listed in Table 2, Table 3 and Table 4, respectively.

BioInfer, a public resource providing an annotated corpus of biomedical English, is aimed at the development of Information Extraction (IE) systems and their components in the biomedical domain.

The ADE (Adverse Drug Effect) corpus consists of MEDLINE case reports annotated with drugs and conditions (e.g., diseases, signs and symptoms), along with untyped relationships between them.

Table 2. Summary statistics of BioInfer dataset. Source: [42].

Name	Sentences	Positives samples	Total
BioInfer	1100	2534	9666

Table 3. Summary statistics of ADE-EX dataset. Source: [43].

Name	Sentences	Positives samples	Total
ADE	4272	988	3184

Table 4. Summary statistics of ChemProt dataset. Source: [44].

Positives samples			
Relation	Training Set	Dev. set	Test set
Active	786	550	664
Inhibitor	2251	1092	1661
Agonist	170	116	194
Antagonist	234	197	281
Substrate	705	457	643
Negative	12461	8070	11013
Total	16589	10482	14456

ChemProt consists of PubMed abstracts annotated with chemical and protein entities. The relations were annotated with 10 chemical-protein relations. According to the shared task description, only 5 out of 10 semantic relation types would be evaluated.

Other important datasets for our study are SemEval-2010_Task_8 datasets [45] and ACE 2005 from LDC (Linguistic Data Consortium) [46]. Both were used in the baseline model, and their statistics are presented below.

SemEval 2010 task 8 is focused on multi-way classification between pairs of nominals. The task was designed to compare different approaches to semantic relation classification.

ACE-2005 consists of 6 main sources: broadcast news (bn), newswire (nw), broadcast conversation (bc), telephone conversation (cts), weblogs (wl), and usenet (un).

reACE, (Edinburgh Regularized Automatic Content Extraction) consists of

English broadcast news and newswires with several annotated entities, such as organization, person, fvw (facility, vehicle or weapon), and gpl (geographical, political or location), along with relationships between them. Relationships are classified into five types: general-affiliation, organization-affiliation, part-whole, personal-social, and agent-artifact.

4.2. Measures

For the relation classification task, we used the F1-score as our measure for evaluation. The F1-score is defined as the harmonic mean between precision (P) and recall (R), such that, $Precision = TP / (TP + FP)$. Precision is the ratio of correctly predicted positive relations to the total predicted positive relations. In turn, Recall is the ratio $TP / (TP + FN)$. Recall is the intuitive ability of the classifier to find all the positive samples. The F1-score is the weighted average of Precision and Recall.

$F1 \text{ Score} = \frac{2 * (\text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})}$, where TP, FP and FN are true positives, false positives, and false negatives, respectively.

4.3. Hyperparameters and resources

We considered a baseline model as a traditional approach to word representations and a CNN model with several windows without the combination of multi-representation. The benefit of multiple window sizes has been demonstrated; here, we used {3, 4} and {2, 3, 4, 5} to generate features. We tested several word representations with sizes $d=50$ and $d=100$, while the dimensionality of entity position indicators was $d=20$. Other parameters are listed in Table 5.

5. RESULTS AND DISCUSSION

The performance of the datasets SemEval 2010, ADE, BioInfer, Chemprot, and reACE is presented in Table 6 with the F1-score. Overall, we can make two observations: (1) The baseline model is not stable for each dataset. Although the predictions of the baseline model are not sufficient to establish what happened, the performance is meant to provide a neutral benchmark to measure the effects of adaptation changes. (2) Our models did not produce, in terms of performance, results comparable to those reported with the baseline model when biomedical domain datasets were used. ADE and reACE contain an imbalanced class distribution, which is perhaps the reason behind their 99.56 and 25.24 performance, respectively.

Our hypothesis is that, as sample relations are extracted from an unbalanced corpus, our baseline model is more sensitive and a significant performance gap is produced. Our model, on the ADE corpus, achieved a high F1-score; however, the variation between the lowest and

highest values of F1 in other datasets does not guarantee a superior performance.

This is perhaps not surprising since in-domain datasets contain short fragments of texts with scarce grammatical information, from which convolutions can capture relevant biomedical information and achieve a high F1. Moreover, after comparing CNN performance, we hypothesize that general and biomedical domain have equally or similarly difficult for RE task, when there is a difference across domain and domain models.

With supervised domain training and the model needs to capture knowledge and learn automatically features from the target domain. While we suspect there is still room for improvement, without utilizing domain specific information, the datasets may differ in ways we cannot account for with our reasoning.

Our baseline model exhibits two main aspects: First, unbalanced corpora have a negative impact on the F1-score.

More importantly, using a corpus with a balanced proportion of positive to negative relations can result in a better performance. Second, there are a number element in a deep learning model implementation, which makes exact replication of the results difficult, particularly performance results, but we compared our modified model from CNN base model (baseline), and the performance was quite similar. We believe our performance can be attributed to (1) vector representation and (2) class imbalance from the dataset. Thus, in-domain word embeddings and position embedding combinations are better for our model than out-of-domain word embeddings, although they cannot achieve results comparable to those of the baseline.

We presented a multi-step reasoning to train a model for other domains in cases in which data with other distribution and classes is available and the task is the same.

Table 5. Best hyper parameter configuration in our mode
Source: Created by the authors.

Parameter	Range	Selected
Dropout rate	0.2 – 0.5	0.5
Weight decay	0 - 1e-10	1e-5
Optimizer	-	Adam
Learning rate	[0.1, 0.01, 0.001]	0.001
Decay learning rate	true, false	True
Number of epochs	20, 50, 80, 100	100
Batch size	[16, 30, 50, 128]	50

Table 6. Weighted F1-scores of the baseline and the four dataset variations
(baseline score marked in boldface). Source: Created by the authors.

Dataset	F1
SemEval2010	82.76
ADE	99.56
BioInfer	38.50
reACE	25.24
ChemProt	59.80

We also showed that our reasoning for model adaptation did not achieve a performance similar to that of the baseline model. Therefore, we carried out corpus-based exploration to address the adaptation of a deep learning model. We evaluated the DL model on different datasets. We also tested the deep learning model to extract semantic relations between entities implementing a similar experimental setup to that in the study by Zeng et. al [1].

After training and testing, our DL model should have learned how to extract semantic relationships due to the automatic learning of similar in-domain and out-of-domain features. However, our results confirm the need

for a balanced dataset and additional information about

the in-domain task. The proportion of positive and negative relations and the number of annotated data in the samples are different in each dataset (ADE, BioInfer, reACE, Chemprot, and SemEval).

Nevertheless, the problem of class imbalance between datasets has been reported in the literature [47].

We observed that the model is sensitive to word representations, which plays a significant role in model training. There are several embedding representations: position embedding (which represents the relative positions of entities and words in the sentence), medical and biological embedding

(which contains specific information) in-domain word embedding (which includes methods that can generate domain-sensitive word embeddings).

Future studies can consider a similar reasoning, exploring, with combinations, different word representations (static, contextualized, with domain knowledge, and others).

6. CONCLUSIONS

In this paper, we proposed a DL model adapted to a new domain, more specifically, RE task for biomedical domain. We used an architecture to transfer the RE task from the generic domain to a biomedical one. After pre-processing the dataset, we obtained experimental results on several benchmark datasets. Nevertheless, we cannot confirm any advantage of the proposed model because it did not achieve a similar performance on different biomedical datasets or results comparable to those of SemEval 2010, which reached an F1-score of 82.76 (Baseline). Even though reACE, BioInfer, Chemprot, and ADE exhibited F1-scores of 25.24, 38.50, 59.80, and 99.56, respectively, these outputs cannot be rejected. We also analyzed the error and discuss the reasons behind our results.

Finally, our study explored different representations and results to avoid the duplicity of research efforts in the development of future systems.

7. REFERENCES

- [1] D. Zeng, K. Liu, S. Lai, G. Zhou, and J. Zhao, "Relation Classification via Convolutional Deep Neural Network," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, Dublin, 2014, pp. 2335–2344. Available: [URL](#)
- [2] Y. Lin, S. Shen, Z. Liu, H. Luan, and M. Sun, "Neural Relation Extraction with Selective Attention over Instances," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, Berlín, 2016, vol. 1, pp. 2124–2133. <http://dx.doi.org/10.18653/v1/P16-1200>
- [3] X. Ren et al., "Cotype: Joint extraction of typed entities and relations with knowledge bases," in *Proceedings of the 26th International Conference on World Wide Web*, Perth, 2017, pp. 1015–1024. <http://doi.org/10.1145/3038912.3052708>
- [4] K. Toutanova, D. Chen, P. Pantel, H. Poon, P. Choudhury, and M. Gamon, "Representing Text for Joint Embedding of Text and Knowledge Bases," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, 2015, pp. 1499–1509. <http://dx.doi.org/10.18653/v1/D15-1174>
- [5] N. Konstantinova, "Review of relation extraction methods: What is new out there?" in *International Conference on Analysis of Images, Social Networks and Texts*, Switzerland 2014, pp. 15–28. http://doi.org/10.1007/978-3-319-12580-0_2
- [6] N. Kambhatla, "Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations," in *Proceedings of the ACL 2004 on Interactive poster and demonstration sessions -*, Barcelona, 2004, pp. 1 - 4. <https://doi.org/10.3115/1219044.1219066>
- [7] R. C. Bunescu and R. J. Mooney, "A shortest path dependency kernel for relation extraction," in *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing - HLT '05*, Vancouver, 2005, pp. 724–731. Available: [URL](#)
- [8] R. J. Mooney and R. C. Bunescu, "Subsequence kernels for relation extraction," in *Advances in neural information processing systems*, 2006, pp. 171–178. Available: [URL](#)
- [9] M. Banko, M. J. Cafarella, S. Soderland, M. Broadhead, and O. Etzioni, "Open information extraction from the web," in *IJCAI*, 2007, vol. 7, pp. 2670–2676. Available: [URL](#)
- [10] R. Socher, B. Huval, C. D. Manning, and A. Y. Ng, "Semantic compositionality through recursive matrix-vector spaces," *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural*

- Language Processing and Computational Natural Language Learning*, Jeju Island, 2012, pp. 1201–1211. Available: [URL](http://doi.org/10.18653/v1/d15-1062)
- [11] D. Zhang and D. Wang, “Relation Classification: CNN or RNN?,” in *Natural Language Understanding and Intelligent Applications*, Springer, Kunming, 2016, pp. 665–675.
https://doi.org/10.1007/978-3-319-50496-4_60
- [12] S. Lim and J. Kang, “Chemical–gene relation extraction using recursive neural network,” *Database*, vol. 2018, Jun. 2018.
<https://doi.org/10.1093/database/bay060>
- [13] Y. Xu, L. Mou, G. Li, Y. Chen, H. Peng, and Z. Jin, “Classifying relations via long short term memory networks along shortest dependency paths,” in *proceedings of the 2015 conference on empirical methods in natural language processing*, Lisboa, 2015, pp. 1785–1794.
<http://doi.org/10.18653/v1/d15-1206>
- [14] S. Zhang, D. Zheng, X. Hu, and M. Yang, “Bidirectional long short-term memory networks for relation classification,” in *Proceedings of the 29th Pacific Asia conference on language, information and computation*, Shanghai, 2015, pp. 73–78. Available: [URL](http://doi.org/10.18653/v1/d15-1206)
- [15] R. Zhang, F. Meng, Y. Zhou, and B. Liu, “Relation classification via recurrent neural network with attention and tensor layers,” *Big Data Min. Anal.*, vol. 1, no. 3, pp. 234–244, Sep. 2018.
<https://doi.org/10.26599/BDMA.2018.9020022>
- [16] T. H. Nguyen and R. Grishman, “Relation Extraction: Perspective from Convolutional Neural Networks,” in *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, Denver, 2015, pp. 39–48.
<https://doi.org/10.3115/v1/W15-1506>
- [17] C. dos Santos, B. Xiang, and B. Zhou, “Classifying Relations by Ranking with Convolutional Neural Networks,” 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)*, Beijing, 2015, pp. 626–634.
<https://doi.org/10.3115/v1/P15-1061>
- [18] K. Xu, Y. Feng, S. Huang, and D. Zhao, “Semantic relation classification via convolutional neural networks with simple negative sampling,” in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, 2015, pp. 536–540.
<http://doi.org/10.18653/v1/d15-1062>
- [19] A. Airola, S. Pyysalo, J. Björne, T. Pahikkala, F. Ginter, and T. Salakoski, “All-paths graph kernel for protein-protein interaction extraction with evaluation of cross-corpus learning,” *BMC Bioinformatics*, vol. 9, no. S 2, pp. 1–12, Nov. 2008.
<https://doi.org/10.1186/1471-2105-9-S11-S2>
- [20] S. Kim, J. Yoon, J. Yang, and S. Park, “Walk-weighted subsequence kernels for protein-protein interaction extraction,” *BMC Bioinformatics*, vol. 11, no. 107, pp. 112–119, Feb. 2010.
<https://doi.org/10.1186/1471-2105-11-107>
- [21] I. Segura-Bedmar, P. Martinez, and C. de Pablo-Sánchez, “Using a shallow linguistic kernel for drug–drug interaction extraction,” *J. Biomed. Inform.*, vol. 44, no. 5, pp. 789–804, Oct. 2011.
<https://doi.org/10.1016/j.jbi.2011.04.005>
- [22] Y. Zhang, H. Lin, Z. Yang, J. Wang, and Y. Li, “A single kernel-based approach to extract drug-drug interactions from biomedical literature,” *PLoS One*, vol. 7, no. 11, pp. e48901, Nov. 2012.
<https://doi.org/10.1371/journal.pone.0048901>
- [23] K. Hashimoto, M. Miwa, Y. Tsuruoka, and T. Chikayama, “Simple customization of recursive neural networks for semantic relation classification,” in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, 2013, pp. 1372–1376. Available: [URL](http://doi.org/10.18653/v1/P16-1123)
- [24] Y. Shen and X. Huang, “Attention-based convolutional neural network for semantic relation extraction,” in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, Osaka, 2016, pp. 2526–2536. Available: [URL](http://doi.org/10.18653/v1/P16-1123)
- [25] L. Wang, Z. Cao, G. de Melo, and Z. Liu, “Relation classification via multi-level attention cnns,” in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Berlin, 2016, vol. 1, pp. 1298–1307.
<http://doi.org/10.18653/v1/P16-1123>
- [26] J. Lee, S. Seo, and Y. S. Choi, “Semantic Relation Classification via Bidirectional LSTM Networks with Entity-aware Attention using Latent Entity Typing,” *Symmetry*, vol. 11, no. 6, Jun. 2019.
<https://doi.org/10.3390/sym11060785>
- [27] M. Xiao and C. Liu, “Semantic relation classification via hierarchical recurrent neural network with attention,” in *Proceedings of COLING 2016, the 26th*

- International Conference on Computational Linguistics: Technical Papers*, Osaka, 2016, pp. 1254–1263. Available: [URL](#)
- [28] P. Zhou et al., “Attention-based bidirectional long short-term memory networks for relation classification,” in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Berlin, 2016, pp. 207–212. <http://doi.org/10.18653/v1/p16-2034>
- [29] R. Cai, X. Zhang, and H. Wang, “Bidirectional recurrent convolutional neural network for relation classification,” in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Berlin, 2016, pp. 756–765. <http://doi.org/10.18653/v1/p16-1072>
- [30] Y. Xu et al., “Improved relation classification by deep recurrent neural networks with data augmentation,” *ArXiv Prepr.*, Oct. 2016. Available: [URL](#)
- [31] Y. Liu, F. Wei, S. Li, H. Ji, M. Zhou, and H. Wang, “A dependency-based neural network for relation classification,” *ArXiv Prepr.*, pp.1-10, Jul. 2015. Available: [URL](#)
- [32] M. Yu, M. Gormley, and M. Dredze, “Factor-based compositional embedding models,” in *NIPS Workshop on Learning Semantics*, 2014, pp. 95-101. Available: [URL](#)
- [33] S. Lai, L. Xu, K. Liu, and J. Zhao, “Recurrent convolutional neural networks for text classification,” in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, Austin, 2015, pp. 2267-2273. Available: [URL](#)
- [34] D. Zeng, K. Liu, Y. Chen, and J. Zhao, “Distant supervision for relation extraction via piecewise convolutional neural networks,” in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, 2015, pp. 1753–1762. <http://dx.doi.org/10.18653/v1/D15-1203>
- [35] S. Pawar, G. K. Palshikar, and P. Bhattacharyya, “Relation Extraction: A Survey,” *ArXiv Prepr.* ArXiv171205191, Dec. 2017. Available: [URL](#)
- [36] J. Legrand et al., “PGxCorpus: A Manually Annotated Corpus for Pharmacogenomics,” *bioRxiv*, Jan. 2019. <https://doi.org/10.1101/534388>
- [37] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998. <http://doi.org/10.1109/5.726791>
- [38] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” in *Advances in neural information processing systems*, Lake Tahoe, Nevada, 2013, pp. 3111–3119. Available: [URL](#)
- [39] J. Pennington, R. Socher, and C. Manning, “Glove: Global Vectors for Word Representation,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, 2014, pp. 1532–1543. <http://dx.doi.org/10.3115/v1/D14-1162>
- [40] J. Turian, L. Ratinov, and Y. Bengio, “Word representations: a simple and general method for semi-supervised learning,” in *Proceedings of the 48th annual meeting of the association for computational linguistics*, Uppsala, 2010, pp. 384–394. Available: [URL](#)
- [41] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, “Enriching word vectors with subword information,” *Trans. Assoc. Comput. Linguist.*, vol. 5, pp. 135–146, Jun. 2017. http://dx.doi.org/10.1162/tacl_a_00051
- [42] S. Pyysalo et al., “BioInfer: a corpus for information extraction in the biomedical domain,” *BMC Bioinformatics*, vol. 8, no. 50, Feb. 2007. <https://doi.org/10.1186/1471-2105-8-50>
- [43] H. Gurulingappa, A. M. Rajput, A. Roberts, J. Fluck, M. Hofmann-Apitius, and L. Toldo, “Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports,” *J. Biomed. Inform.*, vol. 45, no. 5, pp. 885–892, Oct. 2012. <http://dx.doi.org/10.1016/j.jbi.2012.04.008>
- [44] J. Kringelum, S. K. Kjaerulff, S. Brunak, O. Lund, T. I. Oprea, and O. Taboureau, “ChemProt-3.0: a global chemical biology diseases mapping,” *Database*, Feb. 2016. <https://doi.org/10.1093/database/bav123>
- [45] I. Hendrickx et al., “Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals” in *Proceedings of the Workshop on Semantic Evaluations*, Uppsala, 2010, pp. 33–38. Available: [URL](#)
- [46] B. Hachey, C. Grover, and R. Tobin, “Datasets for generic relation extraction” *Nat. Lang. Eng.*, vol. 18, no. 1, pp. 21–59, Jan. 2012. <https://doi.org/10.1017/s1351324911000106>
- [47] T. Ming Harry Hsu, W. Yu Chen, C.-A. Hou, Y.-H. Hubert Tsai, Y.-R. Yeh, and Y.-C. Frank Wang, “Unsupervised domain

adaptation with imbalanced cross-domain data,” in *Proceedings of the IEEE International Conference on Computer*

Vision, Santiago de Chile, 2015, pp. 4121–4129. <http://doi.org/10.1109/iccv.2015.469>