Selection gate-based networks for semantic relation extraction

Jun Sun and Yan Li

School of Computer and Information Engineering,
Henan University,
Kaifeng, Henan, 475000, China
Email: sunjunjerry@outlook.com
Email: yli@vip.henu.edu.cn

Yatian Shen*

School of Computer Science and Engineering,
Southeast University,
Nanjing, Jiangsu, 210000, China
Email: sy602@126.com
*Corresponding author

Lei Zhang, Wenke Ding, Xianjin Shi and Xiajiong Shen

School of Computer and Information Engineering,
Henan University,
Kaifeng, Henan, 475000, China
Email: zhanglei@henu.edu.cn
Email: xd@henu.edu.cn
Email: shixj@haedu.gov.cn
Email: shenxj@henu.edu.cn

Guilin Qi

School of Computer Science and Engineering,
Southeast University,
Nanjing, Jiangsu, 210000, China
Email: gqi@seu.edu.cn

Jing He

The Corporate and Investment,
Bank Technology,
J.P. Morgan Chase N.A.,
25 Bank St., Canary Wharf,
London, E145JP, UK
Email: jing.x2.he@jpmorgan.com

Abstract: Semantic relatedness between context information and entities, which is one of the most easily accessible features, has been proven to be very useful for detecting the semantic relation held in the text segment. However, some methods fail to take into account important information between entities and contexts. How to effectively choose the closest and the most relevant information to the entity in context words in a sentence is an important task. In this paper, we propose selection gate-based networks (SGate-NN) to model the relatedness of an entity word with its context words, and select the relevant parts of contexts to infer the semantic relation toward the entity. We conduct experiments using the SemEval-2010 Task 8 dataset. Extensive experiments and the results demonstrate that the proposed method is effective for relation classification, which can obtain state-of-the-art classification accuracy.

Keywords: relation extraction; selection gate networks; neural networks.

Reference to this paper should be made as follows: Sun, J., Li, Y., Shen, Y., Zhang, L., Ding, W., Shi, X., Shen, X., Qi, G. and He, J. (2021) 'Selection gate-based networks for semantic relation extraction', *Int. J. Embedded Systems*, Vol. 14, No. 3, pp.211–217.

Biographical notes: Jun Sun received his BS degree in Computer Network from Henna University in Kaifeng, Henan Province in 2017 and MS degree in Computer Science from Henan University in Kaifeng in 2020. His research interest includes natural language processing, deep learning and computer vinces.

Yan Li received his BS degree in Computer Science from Henna University in Kaifeng, Henan province in 2016 and MS degree in Computer Science from Henan University in Kaifeng in 2019. His research interest includes natural language processing, deep learning and machine learning.

Yatian Shen received his MS degree in Data Mining from Henan University in Kaifeng, China in 2009 and PhD degree in Computer Application from Fudan University in Shanghai, China in 2017. Currently, he is working in the Post-Doctoral Research Station of Computer Application at Southeast University in Nanjing, China. His research interest includes natural language processing and deep learning. He is a member of youth working committee of Chinese language information society, a member of China Artificial Intelligence Society, CIKM international conference volunteer in 2012 and WSDM conference volunteer in 2015.

Lei Zhang received his Bachelor's degree from Henan University in 2003, Master's degree from Henan University in 2006, and Doctorate from Harbin Institute of Technology in 2015. The main research directions include machine learning, public opinion monitoring, aviation meteorological big data, and remote sensing satellite big data processing.

Wenke Ding is a PhD student in Software Engineering at Henan University. She received her BS degree in Automation in 2001 and her MS degree in Applied Mathematics in 2008 from Henan University.

Xianjin Shi is a PhD student in Software Engineering at Henan University. He obtained his Bachelor's degree in Chemistry in 1995 and Master's degree in Management Science in 2007 from Henan University. As a Senior Engineer, Academic and Technical Leader of the Education Department of Henan Province, and an expert in educational informatisation of the Education Department of Henan Province, he has been engaged in education informatisation and education big data related work for 25 years.

Xiajiong Shen received his BS degree in Computer System Architecture from Wuhan University, Wuhan, China in 1986 and PhD degree in Control Theory and Control Engineering from Shanghai University, Shanghai, China in 2006. His research interest includes spatial data analysis, software engineering, distributed/parallel computing, and formal concept analysis. His awards and honours include: first prize of Henan Provincial Science and Technology Progress Award (2014-J-014-R07/15), third prize of Henan Provincial Science and Technology Progress Award (2006-J-298-R01107), academic technology leader of Henan Provincial Department of Education.

Guilin Qi received his Master's degree in Mathematics and Information from the Jiangxi Normal University in 2002, and a Doctorate in Computer Science from Queen's University of Belfast in 2006. Now, he is the Director of the Institute of Cognitive Intelligence at Southeast University. His main research interests include artificial intelligence, data mining, knowledge mining, etc.

Jing He received her BEng in Computer Science from the Henan University in Kaifeng, Henan in 2009 and MSc in Financial Systems Engineering from University College London in 2010. She is working in financial services in derivatives trading technology and interested in market prediction with machine learning algorithms.

1 Introduction

Relation extraction is an important task in natural language processing, which includes various advanced applications such as information extraction, dialogue system, information retrieval, and big data analysis. It is an important challenge in the field of natural language processing, which has gained widespread attention in recent years.

The traditional methods of relation extraction are mainly supervised learning methods (Kambhatla, 2004; Suchanek et al., 2006; Qian et al., 2008; Mooney and Bunescu, 2005; Bunescu and Mooney, 2005), which require many tools of natural language processing. However, these methods have three drawbacks. First, the upstream NLP system produces the extracted features, and these methods can cause the propagation of the errors in existing tools, which discourage the performance of some systems (Bach and Badaskar, 2007). Second, the manual extraction of certain features is a time-consuming and laborious task. Third, the above feature engineering methods do not scale well in the process of relational extraction, which makes it difficult for the project to realise the learning features and parameters effectively.

In recent years, neural network models have increasingly focused on their ability to minimise the effort in feature engineering of NLP tasks (Collobert et al., 2011). Moreover, most researchers have explored different methods of deep learning (Meng et al., 2018; Xiong et al.,

2018) for relation extraction, such as recursive neural network (Socher et al., 2012; Tang et al., 2015; Huang and Shen, 2016), convolutional deep neural network (Zeng et al., 2014; Xu et al., 2015a), long short-term memory neural network (LSTM) (Xu et al., 2015b) and generative adversarial network (GAN) (Zeng et al., 2018).

However, these models often do not take into account the most important information between entities and contexts. How to effectively select the most important information from the contexts is an important issue, which is the most relevant to each entity.

For instance, given the example input, "that coupled with the death and destruction caused by the storm was a very traumatic experience for these residents". With an annotated target entity mentions e_1 = 'death' and e_2 = 'storm', the goal would be to automatically recognise that this sentence expresses a cause-effect relation between e_1 and e_2. Intuitively, contexts such as 'destruction', 'caused' and 'traumatic' are important information in the sentence, which determines the semantic relation between the entities.

Inspired by the idea mentioned above, to choose the important contextual information around entities, we propose a selection gate mechanism to filter a large number of irrelevant contexts, retain information that has important connections with entities and those are important for understanding the semantics of sentences. Similar ideas are also used in text summarisation tasks (Zhou et al., 2017). The selection gate selects the information related to the entities based on the semantic representation of sentences and entities.

We encode the text segments into their distributed representation through a recurrent neural network, specifically, a bidirectional-LSTM (Hochreiter and Schmidhuber, 1997). Then, to choose the important contexts information, which is the most relevant to the two entities, we propose selection gate-based networks (SGate-NN) to model the relatedness of an entity with its contexts, and select the most relevant parts of contexts to infer the semantic relation towards the two entities. Empirical results on the SemEval-2010 Task 8 dataset show that the proposed approach just with minimal feature engineering obtains state-of-the-art classification results.

The main contribution can be summarised as follows:

1 To preserve dense and distributed information about the text, we encode the text segment to its semantic representation through bi-LSTM.

2 To choose the most important context information, which is most relevant to the two entities, we propose selection gate-based networks (SGate-NN) to model the relatedness of entities with its contexts, and select the most relevant parts of contexts to infer the semantic relation towards the entities.

3 We conduct experiments using the SemEval-2010 Task 8 dataset. Extensive experiments and the results demonstrate that the proposed selection of gate-based networks (SGate-NN) is effective for relation classification, which can obtain state-of-the-art classification results just with minimal feature engineering.

2 Related work

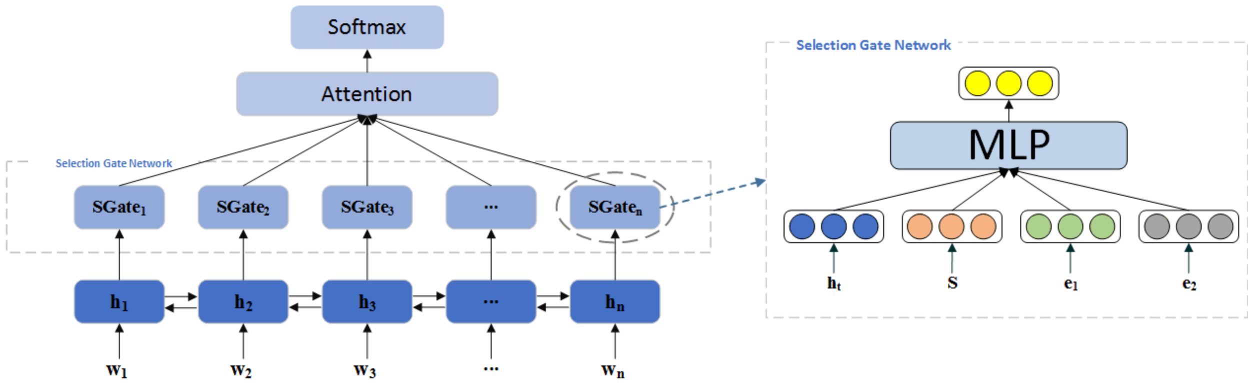
Entity relation extraction, also known as relation extraction, is a key task in the field of natural language processing. Entity relation extraction task is often used as a classification task. There are already a large number of methods that have been used to solve this task and supervised approaches perform well in this task (He et al., 2019). In supervised methods, researchers focus on extracting complex feature representation, usually feature-based methods and kernel-based methods.

Kambhatla (2004) and Suchanek et al. (2006) convert classification clues, such as sequences and parsing trees, to eigenvectors. In the kernel-based methods, such as convolution tree kernel (Qian et al., 2008), sub-sequence kernel (Mooney and Bunescu, 2005) and dependent tree kernel (Bunescu and Mooney, 2005) have been proposed successively. Besides, with the appearance of structural information (Plank and Moschitti, 2013), semantic information is introduced into the kernel methods. However, due to the difficulty of manual annotation, the limited amount of data has led to the emergence of distant-supervised methods (Mintz et al., 2009; Riedel et al., 2010; Hoffmann et al., 2011; Takamatsu et al., 2012).

Some results show that supervision methods have better classification results (Cherichi and Faiz, 2019). However, the performance of supervised methods is highly dependent on the quality of the designed feature. In recent years, with the exploration of deep learning, many researchers have turned their attention to automated feature engineering. In the field of natural language processing, such methods are mainly based on learning the distributed representation of each word (word embedding) (Turian et al., 2010). Socher et al. (2012) proposed a recursive neural network (RNN) for classifying relationships, which determines the semantic relationship by learning the vectors in the syntactic tree paths connecting the two entities. The recurrent neural networks for relational extraction use an explicit weighting of important phrases in a sentence (Hashimoto et al., 2013). The convolution neural network (CNN) is used to extract the features of sentence-level and word level, and the features of these two levels are connected to form the final feature vector (Zeng et al., 2014). Ebrahimi and Dou (2015) reconstructed a recurrent neural network on the dependency path between two tagged entities. In addition to convolution networks, a loss function for data cleaning has also been proposed (Xu et al., 2015a). Xu et al. (2015b) collected heterogeneous information using the shortest dependent path (SDP) between two entities. Huang and Shen (2016) used a word-level attention mechanism to better determine which part of the sentence has the most influence on the two entities of interest.

Different from the previous methods, we propose a semantic-based selection gate model. The model selects the parts that are highly related to entities from some contexts through the selection gate network.

Figure 1 Relation extraction model structure diagram based on a selection gate network (see online version for colours)



3 Model

In this section, we describe the proposed selection gate-based relation extraction model, as shown in Figure 1. First, the text is input into a bidirectional LSTM network and the hidden representation of each word at each time step in the sentence is output. Then, the selection gate selects the information that is associated with entities based on the semantic representation of the sentence. Finally, we use attention to choose highly relevant semantic information. In the following sections, we will describe these three sections in detail.

3.1 Sentence representation layer

In our model, we use bi-LSTM to encode sentences, the LSTM is defined as:

$$i_t = \sigma\left(W_i \cdot x_t + U_i \cdot h_{t-1} + b_i\right) \tag{1}$$

$$f_t = \sigma\left(W_f \cdot x_t + U_f \cdot h_{t-1} + b_f\right) \tag{2}$$

$$o_t = \sigma\left(W_o \cdot x_t + U_o \cdot h_{t-1} + b_o\right) \tag{3}$$

$$g_t = \tanh\left(W_g \cdot x_t + U_g \cdot h_{t-1} + b_g\right) \tag{4}$$

$$c_t = i_t \otimes g_t + f_t \otimes c_{t-1} \tag{5}$$

$$h_t = o_t \otimes \tanh c_t \tag{6}$$

where i_t, f_t, o_t and c_t are the four components of the LSTM, which are called input gates, output gates, forgetting gates and cell units, respectively. g_t is the extracted feature vector, and h_t is the output of the hidden unit at each time step t.

The bi-LSTM consists of a forward LSTM and a backward LSTM. The forward LSTM encodes the sentence from left to right, and the backward LSTM encodes the sentence from right to left. Then, we get the concatenation of the hidden representations of each word at each time step:

$$\vec{h}_t = LSTM\left(x_t, \vec{h}_{t-1}\right) \tag{7}$$

$$\overleftarrow{h}_t = LSTM\left(x_t, \overleftarrow{h}_{t+1}\right) \tag{8}$$

The hidden layer states obtained by the forward network and the backward network are connected to obtain a representation of the sentence:

$$h_t = \begin{bmatrix} \vec{h}_t \\ \overleftarrow{h}_t \end{bmatrix} \tag{9}$$

3.2 Selection gate layer

Correctly finding the relation between entities and contexts is the key to the task of entity relation extraction. However, not all contexts have an important relation with entities. Therefore, we propose a selection gate mechanism to filter a large number of irrelevant contexts, retain information that has important connections with entities and those that are important for understanding the semantics of sentences.

Specifically, the selection gate layer in our model uses four-vector as inputs, which are the sentence word vector h_t, the sentence representation vector S, and the two entity vectors e_1, e_2. The sentence word vector h_t is the output of the bi-LSTM encoder, representing the meaning of the word x_t and context information. The sentence vector S is used to indicate the meaning of the sentence. For each word x_t, the selection gate network to generate a vector SGate using h_t, S, e_1 and e_2, and then representation h_t' is constructed. We connect the forward hidden state $\vec{h}_n$ and the backward hidden state $\overleftarrow{h}_1$ as the sentence representation S:

$$S = \begin{bmatrix} \vec{h}_n \\ \overleftarrow{h}_1 \end{bmatrix} \tag{10}$$

For each step t, the selection gate calculates the gate vector SGate with the sentence S, the bi-LSTM hidden h_t and the two entity vectors e_1, e_2 as inputs.

$$SGate_t = \sigma\left(W_S h_t + U_S S + V_S e_1 + V_S e_2 + b\right) \tag{11}$$

$$h_t' = h_t \otimes SGate \tag{12}$$

where W_S, U_S and V_S is the weight matrix, b is the bias vector, and σ represents the sigmoid activation function. $\otimes$ is element-wise multiplication. Then we got a new sentence representation sequence $(h_1', h_2', h_3', \ldots, h_n')$.

3.3 Attention layer

The input of the attention layer is the new sentence representation sequence $(h_1', h_2', h_3', \ldots, h_n')$. The goal is to assign a different importance score to each representation at each time step in the sequence. The definition of the attention mechanism is as follows:

$$e_i = v^T \tanh(Wh_t') \tag{13}$$

$$\alpha_i = \frac{\exp(e_i)}{\sum_{i=1}^{n} \exp(e_i)} \tag{14}$$

$$c_i = \sum_{i=1}^{n} \alpha_i h_t' \tag{15}$$

Normalise the importance score e_t and then get the context vector c_i by weighting the sum. We input it into the linear layer whose output length is the class number and add the softmax layer to output the probability of the semantic relation classification. The calculation method of the softmax function is as follows, where C is the number of semantic relation categories:

$$\text{softmax}(c_i) = \frac{\exp(c_i)}{\sum_{i'}^{C} \exp(c_i')} \tag{16}$$

3.4 Model training

We train the model in an end-to-end manner. The loss function is the cross-entropy error of the semantic relation classification:

$$loss = -\sum_{t \in N} \sum_{c=1}^{C} P_c(t) \log(P_c(t)) \tag{17}$$

where N is the training data, C is the number of semantic categories, t means a sentence, $P_c(t)$ is the probability of predicting t as class c given by the softmax layer, $P_c(t)$ indicates whether class c is the correct semantic category.

4 Experiments

In this section, we mainly introduce the dataset used in the experiments, the evaluation metric and compare the performance with other methods.

4.1 Dataset

We do our experiments on the SemEval-2010 Task 8 dataset (Hendrickx et al., 2009). This dataset is public and contains a total of 10,017 annotation examples, including 8,000 training instances and 2,718 test instances. The data has nine directional relation classes, and the other has no directional classes. The data in SemEval-2010 Task 8 focuses on the semantic relationship between named pairs. For example, thief and screwdriver are in an INSTRUMENT-AGENCY relation in 'a thief who tried to steal the truck broke the ignition with a screwdriver'. In the experiment, we do not distinguish the direction of the relation, using 10 kinds of tags. To compare with the previous research results, we used the macro-averaged F1-score value as the evaluation criterion in our experiment.

4.2 Parameter settings

In this section, we studied the effects of different parameters in our proposed method: word embedding size, learning rate, batch size and bi-LSTM hidden size. For the initialisation of the word embedding used in our model, we use the publicly available word2vec vectors, which are 100 billion words trained from Google News. The vector has a dimension of 300 and is trained using the skip-gram model (Mikolov et al., 2013). The words that do not appear in the pre-trained word set are randomly initialised. Other parameters are initialised by a uniformly distributed random sample in [–0.1, 0.1]. The final hyperparameters are shown in Table 1.

Table 1 Hyperparameters of our model.

Mini batch size	5
Word embedding size	300
Bi-LSTM hidden size	192
Learning rate	0.001
Iterative times	600

4.3 Analysis and comparison of results

To prove the effectiveness of our approach, we choose six methods as competitors to be compared with our method in Table 2.

A competitor, as described in Hendrickx et al. (2009), uses artificial manual features and SVM as the classifier. Gormley et al. (2015) combine two handcrafted features and uses the word embedding as input. Socher et al. (2012) assign a matrix to each word and modifies the meaning of some words, not just adds word embedding during the recursive process. Zeng et al. (2014) and Xu et al. (2015a) use convolution neural networks model to extract the features of the text. Zeng et al. (2014) learn a more robust relation representation from the shortest dependent path. Hochreiter and Schmidhuber (1997) also consider the shortest dependent path, but it uses a neural network structure of long-short-term memory (LSTM) to get a distributed representation of the sentence.

Table 2 Comparison of the proposed method with existing methods in the SemEval-2010 Task 8 dataset

Model	*Texture sets*	*F1*
SVM	POS, stemming, syntactic patterns, WordNet	78.8
Socher et al. (2012)	POS, NER, WordNet	82.4
Zeng et al. (2014)	Position embeddings, WordNet	82.7
Gormley et al. (2015)	Dependency parsing, NE	83.0
Xu et al. (2015a)	Word embeddings, position embeddings	84.1
Xu et al. (2015b)	POS embeddings, WordNet	83.7
Ours	Word2Vec	86.9

The model proposed by us is called selection gate-based networks (SGate-NN), which models the relatedness of entities with their contexts, and selects the most relevant contexts to infer the semantic relation towards the entities. The experiments demonstrate that this model is very important for semantic classification, our proposed SGate-NN model yields F1-score of 86.9%, whereas the previous best model achieved only F1-score of 84.1% (Xu et al., 2015a).

Table 2 shows the macro average F1 measurements of these competing methods, as well as the resources, features, and classifiers used by each method. Based on these results, the following conclusions are drawn:

1 When using a traditional feature set, the richer the feature, the better the performance. A large number of semantic features can improve the semantic generalisation of data, but the quality of traditional features depends on human ingenuity and existing NLP knowledge. Selecting the best feature set is relatively difficult manually.

2 The model (Socher et al., 2012) includes the feature learning process, which does not require the very intensive work of the feature engineer. However, they rely on the syntax tree used in the recursive process. Errors in grammar analysis inhibit the ability of these methods to learn high-quality features. Socher et al. (2012) can effectively capture the combination of meanings and achieve high performance to a certain extent.

3 Zeng et al. (2014) mainly considers the context information and position information of entities, and then extracts word-level features and sentence-level features through convolution operations. Position coding is also another way of feature extraction. Zeng et al. (2014) obtains a substantial increase of approximately 82.7% F1.

4 Gormley et al. (2015) considers traditional and embedded features. This method associates text embedding with any language structure and expresses them with hand-crafted features. The relationship between manual features and neural network features is also considered.

5 Xu et al. (2015a) can learn more powerful relation representation from the shortest dependent path through convolutional neural networks. Xu et al. (2015b) can learn relation representation from the shortest dependent path through long short-term memory networks (LSTM). The above two models demonstrate the validity of the shortest dependent path in the semantic relation classification task.

6 Our method achieves the best result about 86.9%, and this is the best performance among all of the compared methods. The performance demonstrates the effectiveness of selection gate-based networks (SGate-NN), which can model the relatedness of an entity with its contexts, and selects the most relevant and the most important parts of contexts to infer the semantic relation towards the entities.

5 Conclusions

In this paper, we propose selection gate-based networks (SGate-NN) to model the relatedness of the entity with its contexts and select the relevant parts of contexts to infer the semantic relation towards the entities. Experiments on the SemEval-2010 Task 8 benchmark dataset show that our model achieves better performance than previous neural network models and our model can achieve competitive performance.

In future work, we will focus on exploring a better neural network structure about feature extraction in relation extraction. Meanwhile, because end-to-end relation extraction is also an important problem, we will seek better methods for completing entity and relation extraction jointly.

References

Bach, N. and Badaskar, S. (2007) ‘A review of relation extraction’, *Literature Review for Language and Statistics II.*

Bunescu, R.C. and Mooney, R.J. (2005) ‘A shortest path dependency kernel relation extraction’, *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pp.724–731.

Cherichi, S. and Faiz, R. (2019) ‘Upgrading event and pattern detection to big data’, *International Journal of Computational Science and Engineering*, Vol. 18, No. 4, pp.404–412.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K. and Kuksa, P. (2011) ‘Natural language processing (almost) from scratch’, *The Journal of Machine Learning Research*, Vol. 12, pp.2493–2537, JMLR.org.

Ebrahimi, J. and Dou, D. (2015) ‘Chain based RNN for relation classification’, *Proceedings of the Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL: Association for Computational Linguistics*, pp.1244–1249.

Gormley, M.R., Yu, M. and Dredze, M. (2015) *Improved Relation Extraction with Feature-Rich Compositional Embedding Models*, arXiv preprint arXiv:1505.02419.

Hashimoto, K., Miwa, M., Tsuruoka, Y. and Chikayama, T. (2013) 'Simple customization of recursive neural networks for semantic relation classification', *EMNLP*, pp.1372–1376.

He, Z., Zhongcheng, Z., Liang, G., Jiuming, H. and Yan, Z. (2019) 'Chinese entity attributes extraction based on bidirectional LSTM networks', *International Journal of Computational Science and Engineering*, Vol. 18, No. 1, pp.65–71.

Hendrickx, I., Kim, S.N., Kozareva, Z., Nakov, P. and Ó Séaghdha, D., Padó, S., Pennacchiotti, M., Romano, L. and Szpakowicz, S. (2009) 'Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals', *Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions*, pp.94–99, Association for Computational Linguistics.

Hochreiter, S. and Schmidhuber, J. (1997) 'Long short-term memory', *Neural Computation*, Vol. 9, No. 8, pp.1735–1780, MIT Press.

Hoffmann, R., Zhang, C., Ling, X., Zettlemoyer, L. and Weld, D.S. (2011) 'Knowledge-based weak supervision for information extraction of overlapping relations', *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1, pp.541–550, Association for Computational Linguistics.

Huang, X. and Shen, Y. (2016) 'Attention-based convolutional neural network for semantic relation extraction', *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pp.2526–2536.

Kambhatla, N. (2004) 'Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations', *Proceedings of the ACL 2004 on Interactive Poster and Demonstration Sessions*, pp.22–es.

Meng, R., Rice, S.G., Wang, J. and Sun, X. (2018) 'A fusion steganographic algorithm based on faster R-CNN', *Computers, Materials & Continua*, Vol. 55, No. 1, p.1.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S. and Dean, J. (2013) 'Distributed representations of words and phrases and their compositionality', *Advances in Neural Information Processing Systems*, pp.3111–3119.

Mintz, M., Bills, S., Snow, R. and Jurafsky, D. (2009) 'Distant supervision for relation extraction without labeled data', *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*, Vol. 2, pp.1003–1011.

Bunescu, R.C. and Mooney, R.J. (2005) 'Subsequence kernels for relation extraction', *2005 Annual Conference on Neural Information Processing Systems, NIPS 2005*, pp.171–178.

Plank, B. and Moschitti, A. (2013) 'Embedding semantic similarity in tree kernels for domain adaptation of relation extraction', *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, Vol. 1, pp.1498–1507.

Qian, L., Zhou, G., Kong, F., Zhu, Q. and Qian, P. (2008) 'Exploiting constituent dependencies for tree kernel-based semantic relation extraction', *Proceedings of the 22nd International Conference on Computational Linguistics*, Vol. 1, pp.697–704.

Riedel, S., Yao, L. and McCallum, A. (2010) 'Modeling relations and their mentions without labeled text', *Machine Learning and Knowledge Discovery in Databases*, pp.148–163, Springer.

Socher, R., Huval, B., Manning, C.D. and Ng, A.Y. (2012) '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*, pp.1201–1211, Association for Computational Linguistics.

Suchanek, F.M., Ifrim, G. and Weikum, G. (2006) 'Combining linguistic and statistical analysis to extract relations from web documents', *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.712–717.

Takamatsu, S., Sato, I. and Nakagawa, H. (2012) 'Reducing wrong labels in distant supervision for relation extraction', *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers*, Vol. 1, pp.721–729, Association for Computational Linguistics.

Tang, D., Qin, B., Feng, X. and Liu, T. (2015) *Effective LSTMs for Target-Dependent Sentiment Classification*, arXiv preprint arXiv:1512.01100.

Turian, J., Ratinov, L. and Bengio, Y. (2010) 'Word representations: a simple and general method for semi-supervised learning', *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp.384–394, Association for Computational Linguistics.

Xiong, Z., Shen, Q., Wang, Y. and Zhu, C. (2018) 'Paragraph vector representation based on word to vector and CNN learning', *Computers, Materials & Continua*, Vol. 55, No. 1, p.1.

Xu, K., Feng, Y., Huang, S. and Zhao, D. (2015a) *Semantic Relation Classification via Convolutional Neural Networks with Simple Negative Sampling*, arXiv preprint arXiv:1506.07650.

Xu, Y., Mou, L., Li, G., Chen, Y., Peng, H. and Jin, Z. (2015b) 'Classifying relations via long short term memory networks along shortest dependency paths', *Proceedings of Conference on Empirical Methods in Natural Language Processing*, to appear.

Zeng, D., Dai, Y., Li, F., Sherratt, R.S. and Wang, J. (2018) 'Adversarial learning for distant supervised relation extraction', *Computers, Materials & Continua*, Vol. 55, No. 1, pp.121–136.

Zeng, D., Liu, K., Lai, S., Zhou, G. and Zhao, J. (2014) 'Relation classification via convolutional deep neural network', *Proceedings of COLING*, pp.2335–2344.

Zhou, Q., Yang, N., Wei, F. et al. (2017) *Selective Encoding for Abstractive Sentence Summarization*, arXiv preprint arXiv:1704.07073.