Everant.in/index.php/mcsj

Survey Report

Mathematics and Computer Science Journal

A Survey of Relation Classification via Convolutional Neural Network

Rahul soni, Nitu singh
[1]PG Scholar, MIP India
[2]Professor, Department of Computer Science MIP India

ARTICLE INFO | ABSTRACT

Corresponding Author:
Rahul soni

Abstract: *Relation classification is one of the important research issue in the field of natural language processing (NLP) tasks. It is a crucial intermediate step in complex knowledge intensive applications like automatic knowledgebase construction, question answering, textual entailment, search engine etc. Therefore specific algorithms and pre processing approaches are required for relation extraction among various entities. This paper presents review on relation classification and discusses and analyzes the enhancement of convolutional neural network (CNN) as an algorithm for semantic relation classification tasks..*

KEYWORDS:. **Relation classification, convolutional neural network, features, information extraction, layers.**

1. Introduction

Semantic relations may occur between all parts of speech in sentence. Finding relation among phrases of sentence is applicable in all branches of Natural Language Processing. There is need of fastest, accurate and low cost algorithm for relation classification. In today's search engine also relation classification is a leading light and plays significant role in information extraction [1,2].Relation classification is an essential intermediate step in Natural language processing (NLP) tasks such as automatic knowledgebase population [3] , information extraction, question answering [4].The basic goal is to impulse the system so that it behave and communicate like human beings. In this paper, we will focus on classifying relations between entities in unstructured text document.

In general, relation classification is the task of extracting relation among goal entities from raw text data. Relation classification task can be described as follows: Given a sentence S with a pair of goal nominals e1 and e2, and system aims to identify the relationship between e1 and e2 in given text with defined constraints of relation set .Our analysis concern on approaches developed over readily and freely available standard datasets [5, 6].Majority of relation extraction systems are developed for binary relation extraction. Binary relation can be like Capital-of (Delhi ,India),employee-of(David ,IBM). Higher-order relations are also possible. Generally biomedical relations are of higher orders.

The very first approaches are handcrafted approaches [7] that are based in hand built patterns. However such traditional approaches needs expert's knowledge for linguistic rules implementation.In last decade, with latest data manipulation and operational capabilities, the data-driven approach operational over big data is enabled in relation classification. Data driven approach comprises supervised and unsupervised techniques [8].Survey of previous models explains supervised relation classification algorithms are highly performing and most prominent ones. They are classified into feature based and kernel based methods. Feature based methods [9,10,11] uses variety of features like stemming, POS, NER, WordNet, FrameNet etc implemented over classifiers(Maximum-Entropy model, Support Vector machine).Whereas kernel based methods[12,13] utilizes preprocessed input as parse trees implemented over kernel functions for relation classification. But these approaches are costly and crucial to apply in new relation categories; and so Distant Supervision approach [14] is implemented. The other approach is Bootstrapping or Semi-supervised learning approach [15] that initialize with few well defined patterns, then iteratively learn more. All these approaches are dependent on pre-existing NLP tools which hinders the performance of model and uses manually built feature. Such models are not able to use the nobility of big data.

Resurgence of interest in Deep Neural Network (DNN) [16,17] researchers are focused on exploiting Deep Learning for automatic feature learning from data. Since freely available standard dataset provided by Hendrickx [5] with task of Multi-Way classification of relations between pairs of given entities. Zeng et al. [18] and Socher et al. [19] were first to present their DNN model for this relation classification problem. Socher presented Recursive Neural Network (RNN) [20] to learn features via parse trees. Zeng proposed Convolutional Neural Network (CNN) which extracts lexical level features and sentence level features and combination is used for relation classification. Santos et al. [21] used CNN for relation classification by ranking using class embedding and ranking of loss function to deal with artificial class.
Xu et al. [22] exploited dependency path to learn feature for classifying relation in given relation set. Neural network models illustrated above performs better than traditional models. However due to optimization problem CNN is preferred over RNN.
The structure of paper continues is as follows. Section 2 presents the detail of Convolutional Neural Network. Section 3 presents our conclusion of survey.

2. Convolutional Neural Network in Relation Classification

Convolutional neural network are now quite applicable to problems in Natural Language Processing and results state of art model. A CNN is an extended architecture of the feed forward artificial neural networks with multiple layers. CNN can both be supervised and unsupervised, among them supervised CNN is preferred due to better accuracy. Artificial neural network basically has an input layer, a hidden layer, and an output layer. Every hidden layer and output layer nodes connections mimic the behavior of visual cortex of animal and acts as neurons, whereas CNN applies convolutions over the input layer for computing the output. Thus a local connection is created, where every region of the input is connected to the output neurons. CNN are designed for minimal preprocessing effort. Different variations on CNN architecture are proposed by researchers to improve the relation classification performance.

Simplest way to understand convolution is to think a sliding window function applied to a matrix. Sliding window can be named as a filter, feature detector or kernel. To achieve full convolution filter, multiply its values element-wise with the original matrix, then sum them up for each element by sliding the filter over the whole matrix. The first layer in CNN is input layer. It could have single or multiple channels depending upon representation and need or can have separate channels for different word embeddings like Glove and word2vec. The second layer or convolutional layer is composed of feature maps. To move from input layer to feature map the input layer is convolved with filter, then added with a Bias at pooling layer. The pooling layer scrutinizes their input. Applying a max operation at the result of each filter is a common method of pooling. Obtained fields are then passed through a non-linear function (e.g.; sigmoid function, ReLU, Hyperbolic tangent) which is a layer of neurons that exploits activation function. The filters are initialized randomly and updated after every pass of the algorithm. Each filter varies from other but the same filter is used within single feature map.

Finally, after multiple convolutional and max pooling layers, extreme reasoning in the neural network is performed via fully connected layers. Fully connected layer neurons have complete connections to all activations in the previous layer same as traditional Neural Networks and their activation can be evaluated by matrix multiplication. The feature vector generated by max pooling is fed to loss layer. The loss layer identifies the variation in predicted and true labels generated as a penalty of network training, using loss function (Sigmoid cross entropy, Softmax, Euclidian). Thus output layer extracts the relation label of input sentence.

Researchers choose hyper parameters and regularization (dropout strategies, Max-norm regularization) schemes to solve problem of over- fitting and huge learning rate in their architecture to reach state of arts.

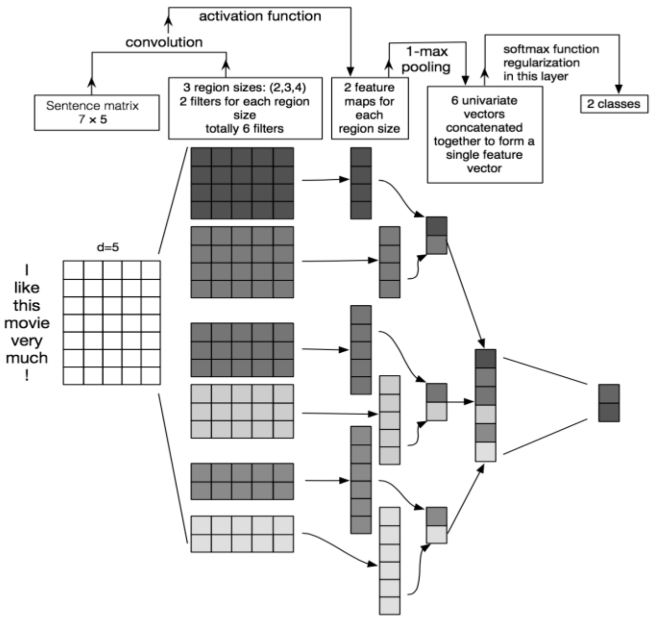

Figure 1: Typical example of Convolutional Neural Network architecture for sentence classification. Image taken from [23].

3. Conclusion

This survey paper explores the use of various techniques for Multi-Way Classification of Semantic Relations between Pairs of Nominal's. While it seems that among all approaches, CNN has more advantages but progress yet to be done. For one, constantly arising question is choosing the CNN architecture and how it will be trained. Investigate the best technique for multi way entity relation classification will be our future work.

References

[1] F. Wu, and D.S. Weld," Open information extraction using Wikipedia," In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 2010.

[2] R. Grishman, "Information Extraction," in IEEE Intelligent Systems, vol. 30, no. 5, pp. 8-15, Sept.-Oct. 2015.

[3] J. Heng and R. Grishman,"Knowledge base population: successful approaches and challenges," In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1 (HLT '11), Vol. 1, Stroudsburg, PA, USA, pp. 1148-1158, 2011.

[4] M.Lyyer, et al. ,"A neural network for factoid question answering over paragraphs," In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014.

[5] Hendrickx, I., et al.," Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals," In Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions, 2009.

[6] C. Walker,et al. ," ACE 2005 Multilingual Training Corpus," LDC 2006T06. DVD. Philadelphia: Linguistic Data Consortium, 2006.

[7] M. Berland, and E. Charniak,"Finding parts in very large corpora," Proceedings of Association for Computational Linguistics, pp. 57-64, 1999.

[8] Quan, Changqin, M. Wang, and F. Ren, "An Unsupervised Text Mining Method for Relation Extraction from Biomedical Literature," Ed. Gajendra P. S. Raghava. PLoS ONE 9.7 (2014): e102039. PMC. Web. 16 April 2016.

[9] 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, 2004.

[10] Y. S. Chan and D. Roth,"Exploiting Background Knowledge for Relation Extraction," In Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010), Beijing, China, pp. 152- 160, August 2010.

[11] T.H. Nguyen and R. Grishman, " Employing Word Representations and Regularization for Domain Adaptation of Relation Extraction," In Proceedings of ACL 2014, Baltimore, Maryland, USA, pp. 68-74, 2014.

[12] Zelenko, D., C. Aone, and A. Richardella, "Kernel methods for relation extraction," The Journal of Machine Learning Research, pp. 1083-1106, March 2003.

[13] M. Zhang, et al. ," A composite kernel to extract relations between entities with both flat and structured features," In Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics,2006.

[14] M. Mintz, et al.," Distant supervision for relation extraction without labeled data," In 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: Volume 2-Volume 2, 2009.

[15] X. Liu and N. Yu, "Multi-Type Web Relation Extraction Based on Bootstrapping," Information Engineering (ICIE), WASE International Conference on, Beidaihe, Hebei, pp. 24-27, 2010.

[16] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," arXiv: 1404.2188, 2014.

[17] R. Collobert, et al.," Natural language processing (almost) from scratch". The Journal of MachineLearning Research, pp. 2493-2537, 2012.

[18] D. Zeng, et al. ,"Relation classification via convolutional deep neural network," In Proceedings of COLING,2014.

[19] R. Socher, et al.," Semantic compositionality through recursive matrix-vector spaces," In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. 2012.

[20] R. Socher, B. Huval, Christopher D. Manning, and Y. Ng Andrew, "Semantic compositionality through recursive matrix-vector spaces," In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Jeju Island, Korea, pp. 1201–1211, July 2012.

[21] Dos Santos, C.N., B. Xiang, and B. Zhou,"Classifying relations by ranking with convolutional neural networks," In Proceedings of 53rd Annual Meeting of the Association for Computational Linguistics. 2015.

[22] K. Xu, et al.," Semantic Relation Classification via Convolutional Neural Networks with Simple Negative Sampling," 2015.

[23] Y. Zhang & B. Wallace, "A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification,"2015.