© 2016 IJSRSET | Volume 2 | Issue 2 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

A Review of Relation Classification with Convolutional Neural Network

Kartik Dhiwar*[1], Abhishek Kumar Dewangan[2]
Department of Computer Science and Engineering, SSGI, SSTC, Bhilai, Chhattisgarh India

ABSTRACT

Relation classification is one of the important research issues in the field of Natural Language Processing (NLP). It is a crucial intermediate step in complex knowledge intensive applications like automatic knowledgebase construction, question answering, textual entailment, search engine etc. Recently neural network has given state of art results in various relation extraction tasks without depending much on manually engineered features. In this paper we present brief review on different model that has been proposed for relation classification and compare their results.

Keywords: Relation Classification, Convolutional Neural Network, Features, Information Extraction.

I. INTRODUCTION

The relation classification is the task of extracting relation among goal entities from raw text. Relation classification problem can be narrated as follows: Given a sentence S with a pair of goal nominal e1 and e2, and system aims to identify the relationship between e1 and e2 in given text with defined constraints of relation set. (Example - The glass contained juice from raw tomato."ENTITY-ORIGIN (juice, tomato)).In relation classification its needed to categorize the given relation nominals that are identified to express some already presumed relation. Relation classification is significant in real time applications like machine translation, document summarization, construction of thesaurus, information extraction [1, 2] , auto knowledgebase population [3] and question answering [4]. It also facilitates language modelling and word sense disambiguation. The basic goal is to impulse the system so that it behaves and communicates like human beings. In this paper we present a review on different proposed model for relation classification and compare their results.

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 rule based or handcrafted approaches that are based in hand built patterns. However such traditional approach needs expert's knowledge for linguistic rules implementation. Rules can be defined like Initial patterns are selected by approaching all sentences in a corpus. Then, occurrences of the pattern are founded for each pattern ending with -ness, -ing, -ity then checked for similarity.

In Hearst rule [7] for IS-A relation extraction consider example sentence "Agar is a substance prepared from a mixture of red algae, such as Gelidium, for laboratory or industrial use" What is Gelidium and how we identify? If X and Y are entities:

X *and other* Y -Mosque, church *and other* religious places
X *or other* Y- pen, pencil *or other* stationary
Y *such as* X- red algae, *such as* Gelidium
such Y *as* X – *such* writer *as* Gandhi, Nehru
Y *including* X –states *including* Delhi, Kerala
Y, *especially* X- currency *especially* Dollar and Euro

Rule based method have high precision and can be designed for particular domain. In cons, Rule based method have low-recall. Manually hand-building patterns for every relation which is hard to write and maintain as there are large number of relation. They had less accuracy.

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 (SVM)).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 initializes with few well defined patterns, and then iteratively learns 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.

With the 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 gives details of dataset, section 4 presents the discussion and section 5 presents our conclusion of survey.

II. METHODS AND MATERIAL

1. 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 art.

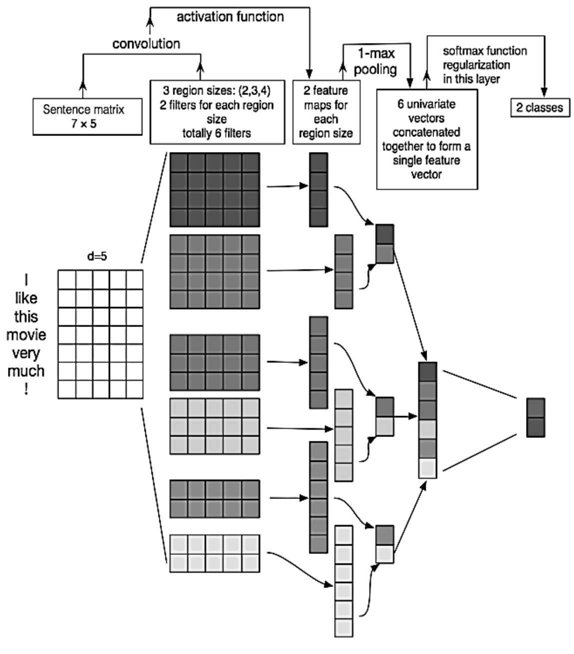

Figure 1. Typical example of Convolutional Neural Network architecture for sentence classification. Image taken from [23].

III. DATASET

We review the model developed over readily and freely available standard datasets of Semeval-2010 task 8: Multi-way classification of semantic relations between pair of nominals and Automatic Content Extraction (ACE) corpus. Semeval-2010 multi way relation classification task is to extract relation among marked entities and classify them among nine groups of relations (Cause-Effect), (Instrument-Agency), (Product-Producer), (Content-Container), (Entity-Origin), (Entity-Destination), (Component-Whole), (Member-Collection), (Message-Topic) and other if given relations not suitable. The dataset consists of Training data that has 8,000 examples of nine relations and other relation. And the testing dataset that has 2,717 examples of nine relations and other relation.

The ACE 2005 corpus is provided by the National Institute for Standards and Technology (NIST). ACE corpus built of around 520 annotated texts collected from various TV news programs, newspapers, and news reports. For relation classification task entities and relations are bounded to 5 types each (PERSON, ORGANIZATION, GEO-POLITICAL ENTITY (GPE), LOCATION, and FACILITY), and (Role, Part, At, Near and Social) respectively.

IV. DISCUSSION

To compare the capability of models to learn feature automatically and their performance we use the F1 scores of model applied over SemEval-2010 Task 8 [5].The SVM approach implemented with various manual built features gives F1 score of 82.2 which is state of art over traditional rule based methods. Various features are implemented but exact performance comparison is very difficult to find. Then for automated feature learning Neural Network is impulse in field. Recursive Neural Network (RNN) was first Neural Network applied giving accepted results. A variation of RNN, MVRNN (Matrix-Vector RNN) with syntactic parsing tree plus other feature obtains F1 score of 82.4. After wards deep neural network approaches like CNN and Deep learning CNN (depLCNN) are applied giving state of art results. The best work is Zeng's [18] CNN which obtained F1score of 82.7.CR-CNN used class ranking for relation classification and achieved F1 score of 84.1.This comparison proves that CNN with word embedding performs best for relation classification.

TABLE I : COMPARE RELATION CLASSIFICATION ON [5]

Method	Feature Sets	F1-score
SVM	16 types of features	82.2
RNN	word embedding +POS, NER, WordNet	74.8 77.6
MVRNN	word embedding, syntactic parsing tree +POS, NER, WordNet	79.1 82.4
CNN	word embedding, position feature +WordNet, words around nominal	78.9 82.7
CR-CNN	word embedding, position feature	84.1
SDP-LSTM	word embedding, Dependency parsing +POS, WordNet, grammar relation	82.4 83.7
depLCNN	word embedding, Dependency parsing +WordNet, words around nominal	81.9 83.7

V. CONCLUSION

This survey paper explores the use of various techniques for Multi-Way Classification of Semantic Relations between Pairs of Nominal's. Deep neural network (DNN) is of big concern in Relation classification. While it seems that among all DNN approaches, CNN has more advantages but progress yet to be done. For one, constantly arising question is choosing the architecture and how it will be trained. Investigate the best technique for multi way entity relation classification will be our future work and we will design our own CNN architecture for further improved results in relation classification.

VI. 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 Machine Learning 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.