Multi-label Semantic Relation Classification Between Pair of Nominals

Kartik Dhiwar,
PG Scholar, Department of Computer Science and Engineering, SSGI, SSTC, Bhilai (CG), India.

Abhishek Kumar Dewangan
Professor, Department of Computer Science and Engineering, SSGI, SSTC, Bhilai (CG), India.

Abstract – Relation classification is a keynote in the field of Natural Language Processing (NLP) to mine information from text facing problems of over-reliance on the standard of handcrafted features. Features annotated by specialists and linguistic data derived from linguistic analysis modules is expensive and ends up with the difficulty of error propagation. Relation extraction plays a crucial role in extracting structured data from unstructured sources like raw text. One might want to seek out interactions between medicines to create medical information or extract relationships among people to create a simply searchable knowledgebase. We propose a deep Convolutional Neural Network model for the multi-label text relation classification task without hand crafted features. This model outperforms the best existing model as per our knowledge without depending much on manually engineered features with the small updates in the loss function applied.

Index Terms – Relation Classification, Features, Label, Convolutional Neural Network, Information Extraction.

1. INTRODUCTION

Natural Language Processing tasks are now applicable to deep neural network. Instead of building hand-craft features, Deep Neural Network automatically build features by fitting different domains and auto learning. This paper, demonstrates an advanced convolution neural network, applied to Relation classification task. The relation classification is the task of extracting relation among goal entities from raw text. The relation extraction problem can be split into two paces: detecting existence of relation among pair of entity of interest in the same sentence and categorize the detected relation entity pair into some defined classes. If we are only categorizing the given relation which are notable to express some predefined expected relation it gives relation classification task. Our relation classification problem can be termed as follows: Given a sentence S with a pair of goal entities e1 and e2, and machine learning system aims to identify the relationship between e1 and e2 in given sentence with defined rules of relation set.

For Example in sentence –

“There were apples, pears and oranges in the bowl.”

If pears and bowl are entities the relation is CONTENT-CONTAINER (pears, bowl).

Relation classification is a functional Natural Language Processing (NLP) task acts as an intermediate step in trivial NLP applications of automatic knowledge base construction, construction of thesaurus, information extraction([1][2]) and question-answering [3].Hand crafted features would give results, but human language is natively ambiguous, and is impossible to go with all phrases that mark relationship. A forward move would be to use machine learning methods to detect the relations and make predictions. Since the last decade there has been increasing interest in applying machine learning approaches to this task to get improved results Deep learning is implemented. If we have some labeled training data, such as examples of pairs of people that are in a relationship, we could train a machine learning classifier to automatically learn the patterns for us. This sounds like a considerable concept, but there are typical challenges:

How to disambiguate words that defined to the same entity? Obtain labeled training data for our machine learning model is challenging. How to tackle ambiguous, uncertain or conflicting data?

We perform following task in our proposed model for relation classification.

- We use word embedding for word vector representation.
- We used convolutional neural network for feature extraction.
- We performed max pooling for dimensionality reduction.
- We used k-fold cross validation to make our method scalable.
- We choose hyper parameters and regularization (dropout strategies) to solve problem of over- fitting and huge learning rate in their architecture.
- Our loss function improved the accuracy.
- We used Softmax classifier for relation classification.

The rest part of paper is organized as follows- in section 2 some related work is reviewed. In section 3 our proposed method is presented. Section 4 describes the experiments done and 5 explain data set .Section 6 describes our evaluation methods and 7 describe results. Section 8 contains some discussion and finally section 9 concludes the paper.

2. RELATED WORK

While our current works emphasis on the supervised method for relation extraction, we tend to consider the supervised systems during this section. Most of work on relation classification has been supervised, generally named as a multi-label or multiclass classification task. The supervised systems (either kernel based or feature-based) some latest systems have implemented the distant supervision method for relation extraction. This method is basically similar to the standard systems in representing relation mentions however make an attempt to get training data automatically investing large knowledge bases of facts and corpus. Feature-based methods depends on features computed from the output of an external lingual pre-processing step [4] [5], whereas kernel methods uses tree kernels, subsequence kernels or dependency tree kernels. With several subtle machine learning approaches being implemented, Deep Neural Network (DNN) [6] [7] has created outstanding accomplishment within the field of Natural Language processing in his technique provides how to automate feature learning virtually from knowledge itself. Since Hendrickx [8] provided an obtainable benchmark (SemEval-2010 task 8 dataset) that focus the tasks of classifying relations between target entities during a given sentence set [9] and [10] gives representative models of deep neural network planning to solve task of relation classification. Xu et al. [11] dig out the information from shortest dependency path between pair of tagged nominals to predict relations in pre-defined relation set and is based on convolutional neural network models. [14] Classifies relation using class ranking in convolutional neural network. [15] Exploits negative sampling in nominals to improve feature extraction. With relation to deep neural network framework, convolutional neural network have raised a typical attention and been with success applied in varied relation extraction and NLP tasks, like participant role labeling, hash tag prediction, question answering and linguistic parsing.

3. PORPOSED MODELLING

This section provides an elaborated explanation of our projected methodology. As stated in previous section the present analysis is concentrated in various aspects of multi-class and multi-label learning like enhancing classification ability of classifiers, by extracting good features and choosing relevant features for the classifiers. Thus, creating only the most relevant and scalable algorithm. We implemented smart options from the previous work and to improve the classification ability by utilizing the position feature and improved loss function among class labels. Our work is closely related to [12].

- Word Embedding
- Feature Extraction (Convolution)
- Pooling
- Classification And Regularization

3.1. Word Embedding

The input of CNN for relation extraction consists of sentences tagged with the two nominals of target interest. As CNNs will solely work with static length inputs, we compute the max distance between entity mentions associated by a relation and select an input width larger than this distance. We insure proper length of every input relation mention by trimming large sentences and padding small length sentence with token. Let n be relation mentions length and y = [y1.y2…..yn] be some relation mention where yi is the i-th word in the mention. Also, let yi1 and yi2 be the two tagged entity mentions of interest. Previous to input the network, each word yi is first transformed into a vector ei by looking up the word embedding table W that is by s pre-trained word embeddings (Glove). Whereas, in order to embed the positions of the two entity nominals and the other words in the relation mention into the illustration, for each word yi, its relative distances to the two entities i to i1 and i to i2 are mapped into real-value vectors di1 and di2 respectively using a position embedding table T. Range of relative distances is from -n + 1 to n - 1 so the position embedding matrix T has size[(2n - 1) x pd] (pd is a hyper parameter shows dimension of the position embedding vectors).Finally, the word embeddings ei and the position embeddings d1 and d2 are combined to represent a single vector yi = [ei, di1 ,di2] to denote the word yi. Thus, the real sentence y can now be viewed as a matrix y of size [(pe + 2pd) x n] where pe is word embedding vector dimension.

y = [y1, y2……yn]

3.2. Feature Extraction(Convolution)

To extract higher level features the matrix y constituting the input relation mention is fed into the convolutional layer. If a widow size w, consider a filter as a weight matrix f = [f1, f2….fw] (column vector fi is of size pe + 2pd). The centre of convolutional layer is acquired by using the convolutional operator on the 2 matrices y and f to generate a score sequence

s = [s1.s2……sn-w+1]

$$s_i = g(\sum_{j=0}^{w-1} \mathbf{f}_{j+1}^{\top} \boldsymbol{Y}_{j+i}^{\top} + b)$$

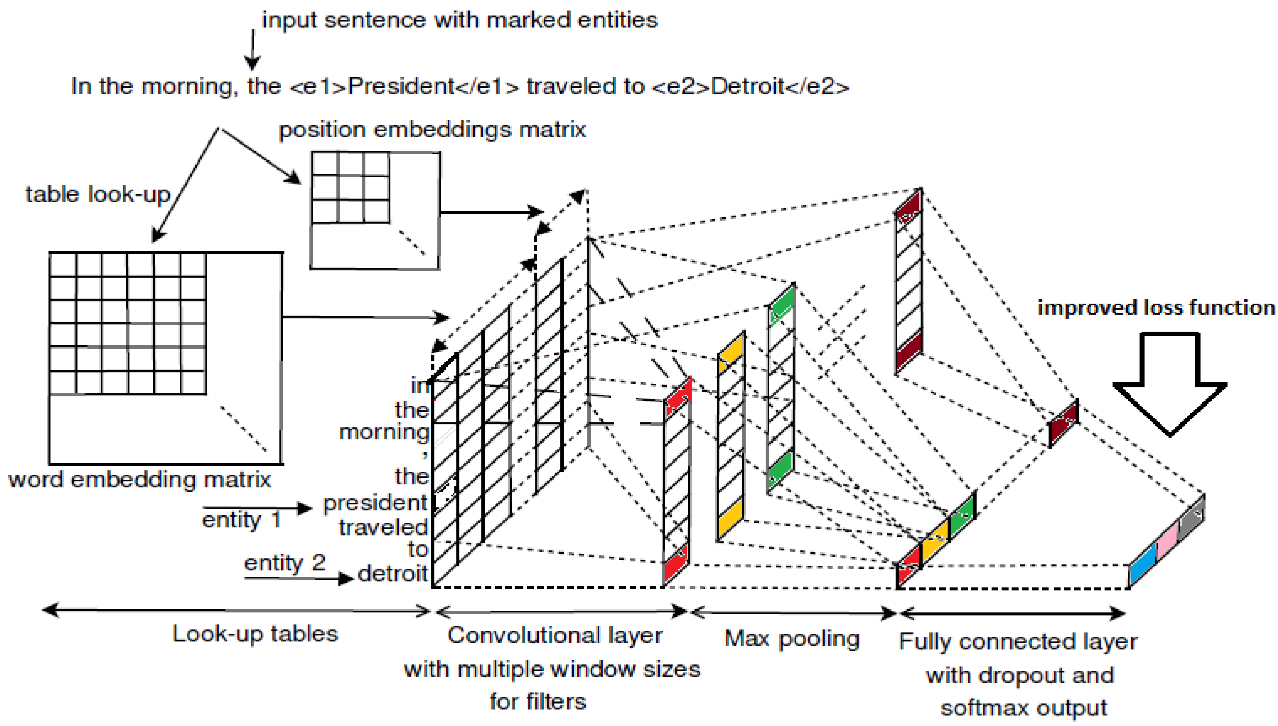

Figure 1 Proposed Convolutional Neural Network for Relation Classification.

Where g is some non-linear function and b is a bias term. This process can be repeated for different filter and window size which increases n-gram coverage of model. To do Relation Extraction we use n-grams associated with relative positions of its words. For measuring n-gram at position i belongs to the corresponding hidden class the filter f of some hidden class of n-gram and scores si are concatenated.

3.3. Pooling

Performing max-pooling over the output of a specific filter size .This is mandatorily a feature vector, where the last dimension corresponds to our features. Every filter and its score are max pooled.

pf = max{s}

3.4. Classification And Regularization

Finally, pooled score of each filter are linked to a solely feature vector z =[p1,p2……pm] to represent the relation mention, where pi is max pooling score for filter I and m represent number of filters.After multiple convolutional and max pooling layers, intense reasoning within the neural network is implemented via fully connected layers. Fully connected layer neurons have complete connections to every activation in the previous layer same as traditional Neural Networks and their activation can be estimated by matrix multiplication. The feature vector generated by max pooling is fed to loss layer. The loss layer evaluates the variation in true and predicted labels generated as a penalty of network training, using loss function Softmax classifier. Thus output layer extracts the relation label of input sentence.

Regularization dropout strategies followed to solve problem of over- fitting and huge learning rate in architecture to reach state of art. Overall, the parameters in the proposed CNN are the word embedding matrix W, the position embedding matrix T, the m filter matrices, the weight matrix C for the fully connected layer. The gradients are computed using back-propagation while training is done via stochastic gradient descent with shuffled mini-batches and the AdaDelta update rule.

4. EXPERIMENTS

4.1 Hyper parameters and Resources

For experiments, we use: tanh for the non-linear activation function, 100 filters for every window size and position embedding vectors with dimensionality of pd = 300, the dropout rate = 0:5, the mini-batch size of 30.Finally, we utilize the pre-trained word embeddings Glove [13] which have dimensionality of Pe = 300 and are trained on 100 billion words of Wikipedia using the continuous bag-of-words architecture. These embeddings are publicly available. Vectors for the words not included in the pre-trained embeddings are initialized randomly. Besides the word embeddings Glove, the model does not use any other NLP toolkits or resources.

5. DATASET AND EVALUATION METHODS

5.1. Dataset

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

5.2. Evaluation Methods

Confusion matrix- To assess the accuracy of our relation classification we create a confusion matrix. In which classification results are compared to real information. The strength of a confusion matrix is that it identifies the nature of the classification errors, as well as their quantities. Each cell [*i*,*j*] indicates how often label *j* was predicted when the correct label was *i*. Thus, the diagonal entries indicate labels that tion.The metrics we used are summarised below. Yi and Zi are given and predicted label sets, respectively and n is the number of instances.

5.2.1. Accuracy (A)

Accuracy is defined as average proportion of the predicted correct labels to the total number (predicted and actual) of labels.

$$Accuracy,\, A = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i \cap Z_i|}{|Y_i \cup Z_i|}$$

5.2.2. Recall (R)

Recall is the proportion of predicted correct labels to the total number of predicted labels, averaged over all instances.

$$Recall,\, R = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i \cap Z_i|}{|Y_i|}$$

5.2.3. F1-Measure (F)

F1-Measure is the harmonic mean of precision and recall, as followed from single-label classification.

$$F1-Measure,\, F = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i \cap Z_i|}{|Y_i| + |Z_i|}$$

6. RESULTS AND DISCUSSIONS

	C-0	C-1	C-2	C-3	C-4	C-5	C-6	C-7	C-8	C-9	sum	acc	rec	f
Cause-Effect- 0	278	0	0	0	13	1	0	1	19	11	323	87.4	86.1	86.7
Component-Whole- 1	0	244	7	1	4	16	7	11	41	5	336	84.7	72.6	78.2
Content-Container- 2	0	3	172	6	0	0	1	1	16	1	200	76.8	86.0	81.1
Entity-Destination- 3	1	1	17	240	1	0	0	0	24	3	287	86.0	83.6	84.8
Entity-Origin- 4	5	1	1	1	206	3	1	2	17	6	243	74.9	84.8	79.5
Instrument-Agency- 5	3	5	1	5	0	113	1	1	26	12	167	66.9	67.7	67.3
Member-Collection- 6	1	6	2	1	1	1	185	0	26	3	226	83.7	81.9	82.8
Message-Topic- 7	0	3	0	0	0	1	0	189	33	4	230	80.4	82.2	81.3
Other- 8	24	22	24	23	41	25	23	25	217	44	468	48.1	46.4	47.2
Product-Producer- 9	6	3	0	2	9	9	3	5	32	168	237	65.4	70.9	68.0
sum	318	288	224	279	275	169	221	235	451	257	2717	84.1	79.8	84.1

Figure 2. Confusion matrix of best performing experiment result

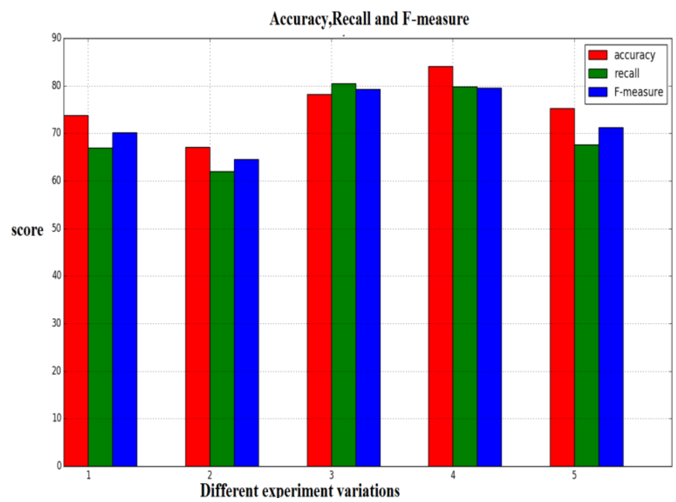

Figure 3 Resultant Comparison Graph of Accuracy, Recall and F-measure

Figure 4. Graph of Loss value

Our detailed review and survey discussion on relation classification can be found at [17].To compare the capability of models to learn feature automatically and their performance we use the accuracy and 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.Our-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.

Classifier	Feature Set	F1
SVM (Rink and Harabagiu, 2010)	POS, prefixes, morphological, WordNet, dependency parse, Levin classes, ProBank, FrameNet, NomLex-Plus, Google n-gram, paraphrases, TextRunner	82.2
RNN (Socher et al., 2012)	word embeddings	74.8
	word embeddings, POS, NER, WordNet	77.6
MVRNN (Socher et al., 2012)	word embeddings	79.1
	word embeddings, POS, NER, WordNet	82.4
CNN+Softmax (Zeng et al., 2014)	word embeddings	69.7
	word embeddings, word position embeddings, word pair, words around word pair, WordNet	82.7
FCM (Yu et al., 2014)	word embeddings	80.6
	word embeddings, dependency parse, NER	83.0
OUR-CNN	word embeddings	82.8
	word embeddings, word position embeddings	**84.1**

Table I. Comparison with present methods

7. CONCLUSION

This paper explores the use of Deep Neural Network techniques for Multi-Way Classification of Semantic Relations between Pairs of Nominal's. Deep neural network (DNN) is very relevant for Relation classification. While it seems that among all DNN approaches, CNN has more advantages but more advancement can be done. For one, perpetually arising question is selecting the model architecture and how it will be trained. Further improvement can be done for addressing wrong labels. Our eexperimental results show that the proposed method provides significant improvements with respect to comparable methods.

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] 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.
[4] 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.
[5] R. Grishman, "Information Extraction," in IEEE Intelligent Systems, vol. 30, no. 5, pp. 8-15, Sept.-Oct. 2015.
[6] 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.
[7] 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.
[8] Zelenko, D., C. Aone, and A. Richardella, "Kernel methods for relation extraction," The Journal of Machine Learning Research, pp. 1083-1106, March 2003.
[9] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," arXiv: 1404.2188, 2014.
[10] R. Collobert, et al.," Natural language processing (almost) from scratch". The Journal of Machine Learning Research, pp. 2493-2537, 2012.
[11] 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.
[12] D. Zeng, et al.," Relation classification via convolutional deep neural network," In Proceedings of COLING, 2014.
[13] 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.
[14] Xu, Y., et al. Classifying relations via long short term memory networks along shortest dependency paths. in Proceedings of Conference on Empirical Methods in Natural Language Processing. 2015.
[15] Nguyen, Thien Huu and Ralph Grishman. "Relation Extraction: Perspective from Convolutional Neural Networks." (2015).
[16] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation.
[17] 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.
[18] K. Xu, et al.," Semantic Relation Classification via Convolutional Neural Networks with Simple Negative Sampling," 2015.
[19] Y. Zhang & B. Wallace, "A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification,"2015.
[20] Dhiwar K. & Dewangan A.K., "A Review of Relation Classification with Convolutional Neural Network ", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Print ISSN: 2395-1990, Online ISSN: 2394-4099, Volume 2, Issue 2, pp.1167-1171, March-April-2016.

Author

Kartik Dhiwar is a Post Graduate research scholar in Department of Computer Science and Engineering, SSGI, SSTC, Bhilai (CG), India. Author obtained his bachelor degree in Computer Science and Engineering. My research interest includes NLP and Deep learning.