RESEARCH ARTICLE OPEN ACCESS

The Application of Deep Learning in Natural Language Processing

Yongxing Peng *
*(School of Information Science and Technology, Jinan University, Guangzhou, China)

--************************----------------------------------

Abstract:

In recent years, due to the improvement of computing performance and neural network breakthrough, depth learning in the field of picture processing and natural language processing has made a great breakthrough. This paper mainly introduces the related research work of the combination of depth learning in natural language processing, and expounds and analyzes the combination effect and advantages and disadvantages of different depth neural network results in different sub-tasks.

Keywords **—Deep Learning, Natural Language Processing**

--************************----------------------------------

I. INTRODUCTION

The concept of deep learning originated from the study of artificial neural networks, and multiple hidden layers of multi-layer perceptrons are a good example of the depth learning model. For neural networks, depth refers to the number of non-linear combinations of levels in the function of network learning. The current neural network learning algorithm is mostly for the lower level of network structure, this network is called shallow structure neural network, such as an input layer, a hidden layer and an output layer of the neural network. In contrast, a network with a higher level of nonlinear computations is called a neural network of depth structures, such as an input layer, three hidden layers, and an output layer. The use of deep neural network for natural language processing has the following advantages: In the network to express the complex objective function of the ability of the depth of the neural network to achieve a high variable function and other complex high-dimensional function of the representation, which can be high latitude information to express the natural language information. In the aspect of information sharing, the multiple levels of extraction features obtained by the depth learning can be reused in similar tasks, which is equivalent to providing unsupervised data for the task solution and obtaining more useful information. As the natural language processing of the various sub-tasks vary widely, the traditional method of dealing with the sub-tasks, the need for different sub-tasks to take different methods of feature representation, can not multi-task sharing resources. The depth of learning methods just have the ability to share this information, you can use a common approach to each sub-task.

II. ENTITY RECOGNITION

Entity identification is the prerequisite for other information extraction tasks to identify the boundaries of entities in the text. Collobert et al. [1] proposed a generalized depth learning framework for naming entity recognition, word-like labeling and other tasks. Yanjun Qi et al [2] on the framework Collobert has been improved and extended, no language word spacing for Chinese, Japanese, etc., is proposed based on the depth learning framework characters. The framework consists of two parts, the first part is the character embedded layer (word vector) and the sentence level of training and output character tags.

The first part mainly uses N-dimensional vector to express the meaning of the characters, the goal is to make similar characters in the vector space is also similar, non-similar characters in the vector space distance as far as possible. First construct the character set with the characters in the corpus, and then randomly initialize a vector that follows the normal distribution for each character. And then use a sliding window for each character and its neighbors to extract the characteristics of the character, where the assumption that the characteristics of each character can be described by its neighbors. Finally, according to the training error, reverse the correction of each character representation vector. The above vector training is mainly for tagged corpus, for many non-tagged corpus, can be learned through a semi-supervised network model. First, each sentence in the expected library is taken as a positive sample, the characters in the character set are randomly replaced by the characters in the positive sample, and then the training is carried out using the network results of the supervised training. The objective function is to make the interval between the positive and negative samples the largest The Finally, there are supervised and semi-supervised training links to get the last word vector.

The second part is the distinction between the sentence-level annotation, mainly for each character role tagging, can be seen as a multi-classification problem with the neural network. However, taking into account the transfer dependencies between labels, you need to build a role transfer matrix. For example: the name of the beginning of the label B-PER rarely followed by O tags, usually B-PER tags followed by E-PER tags. So the use of neural networks to predict the role of characters at the same time also need to use Viterbi algorithm to obtain the role of conversion matrix. The character role is marked as follows:

TABLE I

AN EXAMPLE OF INFORMATION EXTRACTION TAGGING ON A SAMPLE SEQUENCE OF CHINESETEXT. THREE CHARACTER-BASED IE TASKS ARE INCLUDED, (1) WS: WORD SEGMENTATION;(2) POS: part-of-speech tagging and (3) NER: name entity recognition.

Characters	克	林	顿	总	统	前	往	中	东
WS	B	I	E	B	E	B	E	B	E
POS	B-NR	I-NR	E-NR	B-NN	E-NN	B-VV	E-VV	B-NR	E-NR
NER	B-PER	I-PER	E-PER	O	O	O	O	B-LOC	E-LOC

III. RELATIONSHIP EXTRACTION

Relationship extraction is a sub-task of information extraction, used to identify the semantic relations between entities in the text. Chen Yu et al. [3] proposed a feature-based entity relational extraction method using the DBN (deep beliefnets) model, which consists of a multi-tier unsupervised RBM (restricted Boltzmann machine) network and a supervised BP (back-propagation network composed of neural network classifier. The relationship between the DBN method and SVM and the neural network (NN) is compared with the ACE04 corpus. The experimental results show that DBN is superior to these two traditional machine learning methods, and validates the effectiveness of DBN in relation to extraction.

The DBN model combines the advantages of unsupervised learning and supervised learning. It is a neural network with strong classification ability for high dimension sparse eigenvectors. It consists of several unsupervised Restricted Boltzmann Machine (RBM) networks and a supervisor of the back-propagation network (back-propagation, referred to as BP) composition. The unsupervised RBM is mainly used to extract the abstract level of the text, to ensure that the original feature vector is mapped to different feature space, as much as possible to retain the feature information. RBM is composed of a visual layer v (input layer) and a hidden layer h, the joint vector h in the hidden layer and the joint probability distribution of the feature vectors v in the visible layer satisfy:

$$p(v, h) \propto \exp(-E(v, h)) = e^{h^T W v + b^T v + c^T h}$$

Where $E(v, h)$ is the mathematical expectation of the eigenvector v and the eigenvector h, and the absolute value represents the amount of information of the eigenvector v, which represents the eigenvector h. Therefore, the training of RBM is to find the parameter $\theta = (W, b, c)$ which maximizes the joint probability $p(v, h)$, so that when changing the dimension of the input feature, extracting the higher level feature while retaining more original features information. The BP network layer is a supervised classifier. At the last level of the DBN, the feature vector extracted by the front RBM is classified and compared with the correct result, and the overall DBN is fine-tuned.

Yan Xu et al. [7] proposed a new type of neural network SDP-LSTM for relational classification

(SDP: the shortest dependency path between two entities), and classify the SemEval2010 relational classification task data set. The model is characterized by:

- The use of the text of the shortest access path as input, eliminating the impact of irrelevant information;
- The use of multi-channel LSTM network can be integrated on the different information on the path, for example, from different channels can be input word vector, part of speech, grammatical relations and other characteristics;
- Use a custom exit strategy to avoid overfitting.

Shu Zhang et al. [8] proposed a two-way neural network Bidirectional Long Short-Term Memory Networks for relational classification, but also on the SemEval2010 relationship classification task data collection classification. The main features are: word vector, part of speech, named entity, WordNet upper word. This paper compares the traditional machine learning methods such as SVM to illustrate the validity of the model in relation classification.

IV. PRONOUN RESOLUTION

In order to solve the problem that the digestion has been the core problem in natural language processing, Xi Xuefeng et al. [6] proposed a method of digestion based on semantic features using the Deep Learning learning mechanism of DBN (deep belief nets) model. In the process of using the classification method, McCarthy et al. [4] turned the question of judgment to the classification problem, and the classifier was used to judge whether there was a relationship between the finger and each candidate.

The model first defines the feature set of the digestion, and has 24 feature sets and 17 feature sets in English. Each feature is transformed into one-hot code. And then use the depth of the RBM depth network of RBI, respectively, to refer to the relationship between the classification, the network structure is the standard DBN, and Chen Yu et al [3] the same. Finally, the effect of the digestion is compared with that of the existing benchmark system, and there is a certain gap in the overall performance. The main reason for the analysis is that the number of DBN network training nodes has a great influence on the training quality, and the experiment is limited to the performance of the computing platform.

V. EMOTIONAL ANALYSIS

Emotional analysis is the process of analyzing, processing, summarizing and reasoning the subjective text with emotional color, mainly related to the problem of text classification. Yoon Kim [9] proposed the use of convolution neural network CNN to classify the text of the sentence level, through experiments to compare several different feature input in the standard CNN network structure classification effect.

First of all, we need to vectorize the text, the text contrast to the three vector vector initialization form. One is to randomly initialize the word vector, and then adjust the word vector in the process of network training; the second is to use Wikipedia expected library pre-training to get the general word vector to initialize, and then through the training to fine-tune; third is directly using the common word vector , No longer adjust. Then, the initialized sentence convolution, mainly through different sizes of sliding window on the window of the words abstract, extract text features, similar to the traditional language model N-Gram, but the expression is more compact than the N-Gram The Convolution will be through the pool layer to different sliding window to get the characteristics of fusion, the main method is to retain the sliding window to get the maximum value of the feature. The output of the different lengths obtained by the different sliding windows is obtained by pooling the same size of the output for classification, while the pooling reduces the dimension of the feature but retains the most significant information, the model will lose the word position that is less important information Retain the most important information, such as words that express emotions. Finally, the characteristics of the pool after the softmax function classification, the output classification results. Because the depth of the neural network after many characteristics of the abstract, easy to fit, generalization ability is not strong, so in the final output layer before the need to add an exit strategy is to randomly prohibit the input of some neurons, thereby reducing the current training text Unique information to improve the generalization ability.

Finally, the experimental results show that the randomly initialized word vector can be used for text classification and other tasks after adjustment through network training, and its effect is very close to the effect of artificial definition. In addition, the introduction of generic word vectors raises the sorting effect, indicating that the information in the generic forecast library is also important in specific areas of text analysis and contains information that is missing in specific areas.

VI. SUMMARY

As a research field of machine learning has been paid more and more attention in recent years, and many scholars have studied extensively in depth. It has made some breakthrough achievements in the field of natural language processing research, but it is not very mature to combine with deep study in some tasks because of the complexity of natural language processing. In addition, in the practical application of natural language processing, and the combination of deep learning there is still a lot of room for research.

REFERENCES

[1] Collobert R, Weston J, Bottou, L&#, et al. Natural Language Processing (Almost) from Scratch[J]. Journal of Machine Learning Research, 2011, 12(1):2493-2537.

[2] Qi Y, Das S G, Collobert R, et al. Deep Learning for Character-Based Information Extraction[M]// Advances in Information Retrieval. 2014:668-674.

[3] Journal of Software, 2012, 23 (10): 2572-2585. CHEN Yu, ZHENG De-Quan, ZHAO Tie-Jun.Technology Extraction of Chinese Entities Based on Deep Belief Nets [J].

[4] Mccarthy J F, Lehnert W G. Using Decision Trees for Coreference Resolution[C]// Fourteenth International Joint Conference on Artificial Intelligence. 1999:1050--1055.

[5] Ma L, Lu Z, Shang L, et al. Convolutional Neural Networks for Sentence Classification[J]. Eprint Arxiv, 2014.

[6] X Xi-feng, ZHOU Guo-dong.Integration of Pronouns Based on Deep Learning [J]. Journal of Peking University: Natural Science Edition, 2014, 50 (1): 100-110.

[7] Yan X, Mou L, Li G, et al. Classifying Relations via Long Short Term Memory Networks along Shortest Dependency Path[J]. Computer Science, 2015.

[8] Zhang S, Zheng D, Hu X, et al. Bidirectional Long Short-Term Memory Networks for Relation Classification[J]. 2015.

[9] Kim Y. Convolutional Neural Networks for Sentence Classification[J]. Eprint Arxiv, 2014.