Application Research of Deep Learning in Natural Language Processing

Junmin Zhu[1, a], Yiang Gao[2]

[1] Cangzhou Bank, Cangzhou, 061000, China

[2] Tsien Hsue-Shen College, Nanjing University of Science &Technology, Nanjing, 210094, China

[a]email: 541472883@qq.com

Keywords: Deep learning, Natural language processing, Neural network

Abstract: With the development of deep learning and its application in related fields, the performance of machine learning has been significantly improved. The studies of deep learning applied in natural language processing tasks has made many breakthroughs. This paper first introduces the basic concepts of deep learning from the perspective of its structure and application motivation, then focuses on the analysis of the neural network language model, using distributed representation to remove dimension disasters. Finally, this paper introduces the typical applications of deep learning in natural language processing, including syntactic analysis, word meaning learning and sentiment analysis to provide some references for the relevant researchers.

1. Introduction

As the fastest developing research direction in the field of machine learning and artificial intelligence, deep learning has attracted great attention from academia and industry [1]. Deep learning is a general term for a series of machine learning algorithms based on feature self-learning and deep neural network. At present, the research of in-depth learning has made great progress, and has made great breakthroughs in the traditional feature selection and extraction framework. It has exerted more and more important influence on many fields, including natural language processing, biomedical analysis, remote sensing image interpretation, and has achieved revolutionary success in the field of computer vision and speech recognition. At present, how to apply in-depth learning technology to solve the tasks related to natural language processing is a research hotspot of in-depth learning. Natural language processing, as an important research direction in the intersection of computer science and artificial intelligence, integrates the knowledge and achievements of linguistics, computer science, logic, psychology, artificial intelligence and other disciplines. Its main research tasks include part-of-speech tagging, machine translation, named entity recognition, machine question and answer, emotional analysis, automatic summarization, parsing and co-reference resolution. As a highly abstract symbolic system, natural language is difficult to measure the relationship between texts. Relevant research relies heavily on artificial construction features. The advantages of deep learning method lie in its strong discriminant ability and feature self-learning ability, which is very suitable for the characteristics of natural language such as high dimension, label-free and large data. In-depth learning can automatically complete feature extraction and data representation, and advocates learning to extract effective representations of different dimensions and levels, so as to continuously improve the ability of data interpretation. From the perspective of cognitive science, the thinking of deep learning is very consistent with people's learning mechanism [2]. This paper mainly studies deep learning applications in natural language processing.

2. Structure and Motivation of Deep Learning

2.1. Deep Structure

Deep learning is a hotspot in the field of machine learning by building deep neural networks to simulate the mechanism of human brain to interpret and analyze the data of learning images, voice

  DOI: 10.25236/csam.2019.064

and text. The validity of traditional machine learning depends largely on the validity of the data representation and input features designed by human. The role of machine learning method in this process is only to optimize the learning weight so as to output the optimal learning results. Unlike traditional machine learning methods, in-depth learning attempts to automate data representation and feature extraction. Furthermore, in-depth learning emphasizes on extracting effective representations of different levels and dimensions through the learning process, so as to improve the interpretation ability of data at different abstract levels. From the perspective of cognitive science, this idea is very consistent with human learning mechanism. The difference between deep learning and shallow learning is that first, deep learning requires that the model structure must have enough depth, usually requires more than one layer of hidden nodes, and some may even reach multiple layers. This multi-layer non-linear mapping structure is helpful to the approximation of complex functions. Secondly, in-depth learning emphasizes the importance of feature learning. Through unsupervised pre-training algorithm, the features of the input original samples in the original space are mapped to a new feature space layer by layer, and it is possible to use the new features to achieve classification or prediction more easily. In addition, the generative pre-training method also avoids the problem of over-fitting which may occur because of the strong expressive ability of network functions [2].

2.2. Application Motivation

The clear hierarchical structure of human perception system shows that it greatly reduces the amount of data processed by visual system and retains useful structural information of objects. In order to extract the natural image, video, voice and other structurally rich data with potentially complex structural rules, in-depth learning can acquire its essential features. In addition, the cognitive process itself is a deep structure. Human beings organize ideas and concepts in a hierarchical manner [3]. Engineers decompose solutions into concepts and processes at different levels. From the observable sensory data, interpretation or recognition of relevant features gradually get rid of the limitations of selecting artificial features, and gradually become an important idea in the process of in-depth learning. In a sense, all the learning methods that can have the function of automated learning can be included in the scope of in-depth learning. Conversely, linguistic representation is also a sparse representation, using part of all vocabulary to represent corresponding concepts, such as describing a scene, requiring only a small part of vocabulary, similar to the way the brain extracts data. The hierarchical concept of in-depth learning is in line with human cognitive learning process. From the perspective of cognitive science, human cognitive learning process is carried out in layers, and layered structure is the basic requirement of cognitive learning. For example, in the process of solving complex problems, engineers must decompose tasks into several smaller sub-tasks, and sub-tasks and total tasks are also at different levels of cognitive abstraction. Finally, neurobiology studies show that there is a hierarchical structure in the human brain, which provides further evidence for the effectiveness of in-depth learning from the perspective of bionics. There are a series of layered areas in the cerebral cortex [4].

3. Neural Network Language Model

The goal of statistical language model is to learn the joint probability of word sequences in a language. The most difficult problem is the dimension disaster problem. Bengio et al. proposed that dimension disasters can be eliminated by learning a distributed representation for each word, which enables the model to understand the number of sentences with adjacent semantics at the relevant exponential level. At the same time, the model learns the probability function of the distributed representation of each word and the distributed representation of word sequence. If the tested sentences are composed of words in the training set, a high probability can be obtained.

Statistical language model can be expressed in the form of conditional probability. Given all the previous words, the probability of the next word can be obtained.

$$\hat{P}(W_l^T) = \prod_T \hat{P}(W_t|W_l^{t-1})$$

In that formula, W_t is the tth word and W_i^j is a subsequence $W_i^j = (w_i, w_{i+}, \dots, w_{j-,} w_j)$ This method has been applied to many fields of natural language processing, such as language translation, information retrieval and so on.

Neural network language model estimates the probability of grammar model by using neural network. The training set of the model is a word sequence of $w_i, \dots, w_{T,} w_t \in V$. The vocabulary V is a large and finite set. The purpose is to learn a good model.

$$\mathrm{f}(w_t, \dots, w_{t-n+1}) = \hat{P}(W_t | W_l^{t-1})$$

The output of function g is a vector whose probability of estimation of the first element is:

$$\mathrm{f}(\mathrm{i}, w_t, \dots, w_{t-n+1}) = \mathrm{g}(\mathrm{i}, \mathrm{C}(w_{t-}), \dots, \mathrm{C}(\mathrm{w}_{t-n+1}))$$

Function f is a combination of these two mappings and g. In the context, all words are shared. Each of these two parts is related to some parameters. The parameter of mapping c is the eigenvector itself, and g is composed of a forward feedback neural network or a recursive neural network.

4. Typical Applications of Deep Learning in Natural Language Processing

4.1. Syntactic Analysis

The main task of parsing is to automatically identify the syntactic units contained in sentences and the relationships among them, that is, the structure of sentences. Usually, given a sentence as input and using the grammatical features of language as the main knowledge source, Henderson constructs a phrase structure tree. Henderson proposes a Left-corner parser, which successfully applies the neural network to large-scale parsing for the first time. Henderson then trains the parser based on synchronous network; Titov and others use SVM to improve a generation model. Syntax parsers are used for parsing tasks in different fields. On the basis of feature learning, they also seek to improve the system further. Collobert proposes a fast discriminant algorithm for natural language parsing based on deep loop graph transfer network. This method uses fewer text features, and achieves performance indicators comparable to the best discriminant analyzer and benchmark analyzer at that time, but has greater advantages in computing speed. At the same time, Costa and others also try to use recurrent neural network model to solve the problem of ranking candidate additional phrases in incremental parsers. Their work reveals for the first time the possibility of using recurrent neural network model to obtain enough information to correct the results of parsing. But they only tested on several sentence subsets. Menchetti et al. used Collins parser to generate candidate syntax tree and used recurrent neural network model to achieve reordering. Similar to their work, Socher et al. proposed a CVG (Compositional Vector Grammar) model for syntactic structure prediction. This model combines PCFG (Probability Context Free Grammars) with recurrent neural network model, and makes full use of phrase grammar and semantic information. Legrand et al. proposed a bottom-up parsing method based on simple neural network model. Its main advantages are simple structure, low computational cost and fast analysis speed.

4.2. Word Meaning Learning

Meaning representation based on unsupervised learning mechanism is widely used in natural language processing, such as input of some learning algorithms or feature representation of special words. In terms of learning word meanings, global context can provide more useful information. Huang et al. proposed a new deep neural network model for word meaning learning based on Collobert and Weston. The model combines local and global text context information to learn hidden words that can better express word meanings. By learning the polysemy representation of each word, it can better explain synonym ambiguity. Meaning; furthermore, on the basis of polysemy of words based on multiple word vectors, the improved model makes the word vector contain more abundant semantic information. Experiments show that compared with other vectors, Huang's method is closest to the manual annotation semantic similarity. Socher et al. mentioned the concept of deep understanding of language. They believe that the vector space model of a single word is in lexical information science. They proposed a deep recursive neural network model, which can represent semantics by learning the combination vectors of phrases and sentences.

Sentences can be sentences of any syntactic type and length. The model is applied to the syntactic tree. Each node is assigned a vector and a matrix; the vector obtains the ontological semantics of the elements; the matrix captures the change information of adjacent words and phrases. The model has achieved remarkable performance in three different experiments, namely, the prediction of the emotional distribution of adverb, adjective combination pairs, the emotional classification of film review markers, and the emotional relationship classification, such as causal or subject information between nouns.

4.3. Sentiment Analysis

Recursive machine learning algorithm can be used to predict the distribution of emotional tags at sentence level. We use Collobert & Weston's word vector representation, then we use the recursive model based on Autoencoders to learn, and describe how to learn phrase representation, phrase structure and emotional distribution. The main goal of the model is to find the vector representation of variable length phrases by unsupervised or semi-supervised methods, and to be used in subsequent related tasks. In affective analysis tasks, the performance of multi-word phrase learning vector representation is usually better than that of corpus representation without any predefined affective dictionary. Experiments show that the methods mentioned in this paper are more accurate than other emotional analysis methods. Zhou et al. proposed a semi-supervised learning algorithm called Active Deep Network Work to solve emotional classification problems. Firstly, they used unsupervised learning algorithm to train RBM on labeled and unlabeled data sets, then built ADN, and fine-tuned the structure by supervised learning method based on gradient descent algorithm. A semi-supervised learning framework is trained by tagged comment data, which is integrated with ADN structure, and a unified model for semi-supervised classification task is realized. Experiments show that the model has outstanding performance on emotional classification datasets. The improvement of RBM performance in ADN is partly due to the improvement of the scale of unlabeled training data, which makes a large number of unlabeled comments abundant. Data opens up the use of space. Glorot et al. proposed an unsupervised learning method to learn how to extract meaningful information representation from online comment data, and applied it to the construction of emotional classifier. The test performance on the class comment benchmark data of Amazon products is remarkable. Socher et al. proposed a deep learning model based on RAE and applied it to sentence-level emotions. Annotation prediction. The model uses word vector space to construct input training data and RAE to realize semi-supervised learning. Experiments show that the accuracy of the model is better than that of similar benchmark systems.

5. Conclusion

Although in-depth learning is still in its early stage of development, it has had a far-reaching impact on many modes, such as image, sound and language. There are still a lot of work to be studied, especially in the field of natural language processing. We need to give full consideration to the training corpus so that the system can guarantee robustness and versatility and achieve better results in different fields.

References

[1] Lin Yiou, Lei Hang, Li Xiaoyu, et al. Deep Learning in NLP: Methods and Applications[J]. Journal of University of Electronic Science and Technology of China, 2017, 46(6): 913-919.

[2] Xi Xuefeng, Zhou Guodong. A Survey on Deep Learning for Natural Language Processing [J].Acta Automatica Sinica, 2016, 42(10): 1445-1465.

[3] Zhu Baobao, Zhang Na. Natural language processing based on deep learning [J]. Wireless Internet Technology, 2017(10): 25-26.

[4] Wang Qiansheng, Yu Nan, Zhang Meishan,et al. N3LDG: A Lightweight Neural Network Library for Natural Language Processing [J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2019, 55(1): 113-119.