

THE EFFECTS OF THE LDA TOPIC MODEL ON SENTIMENT CLASSIFICATION

DU Jing wei

Department of Information Science and Technology, JINAN University,
GuangZhou, CHINA

ABSTRACT

Online reviews are a feedback to the product and play a key role in improving the product to cater to consumers. Online reviews that rely heavily on manual categorization are time consuming and labor intensive. The recurrent neural network in deep learning can process time series data, while the long and short term memory network can process long time sequence data well. This has good experimental verification support in natural language processing, machine translation, speech recognition and language model. The merits of the extracted data features affect the classification results produced by the classification model. The LDA topic model adds a priori a posteriori knowledge to classify the data so that the characteristics of the data can be extracted efficiently. Applied to the classifier can improve accuracy and efficiency. Two-way long-term and short-term memory networks are variants and extensions of cyclic neural networks. The deep learning framework Keras uses Tensorflow as the backend to build a convenient two-way long-term and short-term memory network model, which provides a strong technical support for the experiment. Using the LDA topic model to extract the keywords needed to train the neural network and increase the internal relationship between words can improve the learning efficiency of the model. The experimental results in the same experimental environment are better than the traditional word frequency features.

KEYWORDS

Deep learning, LDA topic model, Two-way long-term and short-term memory cycle network, Natural Language processing, Sentiment classification.

1 INTRODUCTION

With the advancement of the times, we have already entered the era of big data. Data is a clear representation of the characteristics of things. Discovering the hidden features of things through massive data has become a hot topic in current technology. As a sub-task in the direction of natural language processing, sentiment classification plays a vital role in various fields. The advertiser who sell the ads and the goods will receive feedback. The advertisement put by the merchants, the commodities sold and the feedback they get all have the attribute of whether the consumers like them. A better understanding of the market demand and the development of marketing strategies cannot be separated from strong data demonstration. Traditional machine learning algorithms, such as SVM, logistic regression, etc., have been thoroughly defeated by deep learning algorithms in sentiment classification tasks. The LSTM-based emotional classification task generally show good prediction results, but there are also shortcomings.

In most basic LSTM sentiment classification tasks, only the word frequency is sorted and the first N words with higher frequency are extracted as the feature lexicon. Although these words will dig out the internal relationship during the model training, the feature words are extracted. It is somewhat sloppy based on a single traditional word frequency. LDA [1] topic classification can extract words with implicit factors instead of single word frequency extraction words, which can

improve the correlation of feature vectors before training to a certain extent. And reduce redundant vocabulary.

2.OVERVIEW OF THE PRINCIPLES INVOLVED IN THE ALGORITHM

The model algorithms used in the article are LDA topic model, RNN, and algorithms such as Bi-LSTM [3]-[4].

The LDA topic model is a generative model based on Bayesian theory. It assumes that the document topic distribution obeys the multinomial distribution with the parameter θ , the parameter θ obeys the dirichlet distribution with the parameter vector α , and the word distribution of each topic is also a multinomial distribution. This multi-distribution parameter is subject to a dirichlet distribution and updated with the gibbs sample update rule.

Recurrent Neural Network(RNN) can process ordered data, retaining t-1 and related information at the previous moment to affect the output of the current t-time, but for long sequences, it faces the situation of gradient disappearance or gradient explosion, resulting in RNN for long The processing of time series data is less than ideal. For long-term sequence data processing, LSTM (Long Short-Term Memory) long-term and short-term memory network can store relevant information far away from the current state through special design, which has a good model compliance in text processing. A sentence sequence consists of words[5]-[6], and words can be read as a time series data.

Although LSTM solves some of the above problems, it cannot encode information from the back to the front. Language is an art. Sometimes sentences may have irony or puns, and may require post-text information combined with pre-words to judge. BiLSTM (Bi-directional Long Short-Term Memory)[7]-[8] takes into account the context information of the sentence, and can better capture the two-way semantic dependence information, so as to better understand and judge the whole semantics.

3.EXISTING PROBLEMS AND ALGORITHM IMPROVEMENT

In the feature extraction stage, the first 10,000 words with the highest frequency of occurrence of words in the data set are first used as feature words to construct the dictionary V .

In the comments after the stop word is removed, there is still no word about the emotional polarity of the comment, so that the dictionary based on the word frequency structure contains a part of the vocabulary that is not related to the result of the task, so that the original data is converted into a vector [9]. There are quite a few feature dimensions that are not related to emotions, but they cannot specifically identify which ones are redundant in each vector data. Therefore, it is necessary to construct a feature matrix with a large dimension to contain all the dimensions in the feature vector to achieve the expected experimental precision.

The high-frequency words in the dictionary constructed on the basis of word frequency may have little or no correlation with emotional polarity but have become part of the dictionary. In order to reduce the appearance of such words in the dictionary, the two-layer LDA theme model is used to extract the features twice [10], and the extracted keywords are used as the data of the subsequent neural network model. These words can be filtered out by prior knowledge. Increase the relevance of the classification. There are some hidden themes behind the first layer of LDA extraction, and then the second screening and classification of these words, the second topic classification of these implicit keywords, divided into two types of topics, the weighting value is more The larger words

are used as the second classification result, reducing the number of words unrelated to positive and negative comments. The specific method is to treat each comment as a separate text and set the number K of hyperparameters. Since the topic extraction is the intermediate link of our classification, the setting of the first K value should not be too small, otherwise the established model is not enough. The characteristics of the description data should not be too large, otherwise redundant data will be generated. After dividing the theme, it is equivalent to a soft dimensionality reduction, and a corpus of size V is extracted. And because the emotional classification task for us is only praise and bad review, it is a two-category task. The LDA algorithm is used for the second time on the reduced-dimensional data. Here, the LDA algorithm needs to be improved, and the topic distribution in the Bayesian model. The service is transformed from a multinomial distribution [2] to a binomial distribution, the conjugate prior distribution of the binomial distribution is changed to a beta distribution, K is set to 2, the subject distribution is subject to a binomial distribution, and the parameters of the binomial distribution p_0, p_1 are obeyed by β . Distribution, for the parameters α_1, α_2 of the β distribution, we think that the positive and negative comments have the same important position, so the initial values are set equal, here take $50/2+1 = 26$, the subject-word features obey the multiple distribution, put this For the KV dimension matrix, we initialize it to 0.005. For the sampling of the subject distribution, we use a random number of 0.5 for sampling training, and for the word distribution we still use Gibbs sampling. According to the Occam razor principle, although the complexity of the model is reduced, the conjugate distribution is more in line with the characteristics of the data model, which increases the interpretability of the prediction.

The second LDA model parameters are defined as follows:

W_{mn} : The n th word of the m th document is obtained by observation.

α : The priori parameter is parameter of the binomial distribution, which is subject to the beta distribution.

γ : The a priori parameter, the subject-word distribution of the conjugate distribution dirichlet distribution parameters.

θ : Subject distribution parameters obtained from the previous analysis, subject to binomial distribution

ϕ : This is an implicit variable, the word distribution obeys the polynomial distribution.

t_{mn} :A subject under the binomial distribution.

Based on the word bag model, the joint distribution probabilities of all variables are:

$$p(w_{mn}, t_{mn}, \theta, \phi | \alpha, \gamma) = \prod_{n=1}^{N_m} p(w_{mn} | t_{mn}, \phi) p(t_{mn} | \theta) p(\theta | \alpha) p(\phi | \gamma)$$

4.EXPERIMENTAL VERIFICATION.

The experiment used the movie review of the public data set IMDb (Internet Movie Database) as a data source for sentiment analysis. Use word2vec as a word embedding technique.

Experimental group: Using the feature words for the words with the highest frequency of 10,000 words to form a word list, with 256 dimensions as the representation dimension of the word, each comment takes $n=\{100,200,300,400\}$ words, the excess part will be truncated, using LSTM and

Bi-LSTM [11]-[16] is used as a classifier. Among the 500 million movie reviews, the first 300 million are used as training data sets and the last 20,000 are used as test data sets.

Control group: A dictionary built using the LDA topic model. The first topic classification setting K is 500 topics, in 30,000 comments, we believe that subject to a multi-distribution, through Gibbs sampling, extracting V words under 500 topics, using the improved LDA. The algorithm performs the second topic feature classification, and extracts the extracted words with higher weights to construct a dictionary and input the neural network for training. In order to compare with the dictionary constructed by word frequency, the n and neural network classification models taken by each comment word in this group of experiments correspond to the experimental group respectively.

5. COMPARATIVE ANALYSIS OF EXPERIMENTAL RESULTS

5.1. LSTM experiment results

In the experiment of LSTM [17] as a classifier[19], the results show that the dictionary based on word frequency construction achieves a prediction accuracy of about 88% through five rounds of iteration under the condition of $m=500$, and each iteration takes about 1800 seconds, which is performed on the CPU used. It takes about 2 hours and 30 minutes for a training session.

Through the training of the theme model by LDA, the obtained model finally extracts 10000 weighted words from 50,000 words and builds a dictionary, and when $m=200$, the same network structure is used for training. After five rounds of iteration, the prediction accuracy of about 88% is also achieved, and each iteration is only about 480s (the same CUP and experimental conditions). The five rounds of iteration takes about 40 minutes, which is 4 times more efficient than when using the word frequency dictionary. . However, when the feature vector constructed by the word frequency dictionary is $n=200$, the experimental precision is only 60%.

In the case where the other experimental conditions are the same, using the dictionary model constructed by word frequency, the experimental effect of the comment length $n=500$ is equivalent to the effect of the LDA extracted word dictionary model of $n=200$, and the above-mentioned LDA theme can be verified from the result. The dictionary composed of the feature words extracted by the model can be more close to the emotional polarity, reduce the redundant vocabulary and improve the efficiency of the training model.

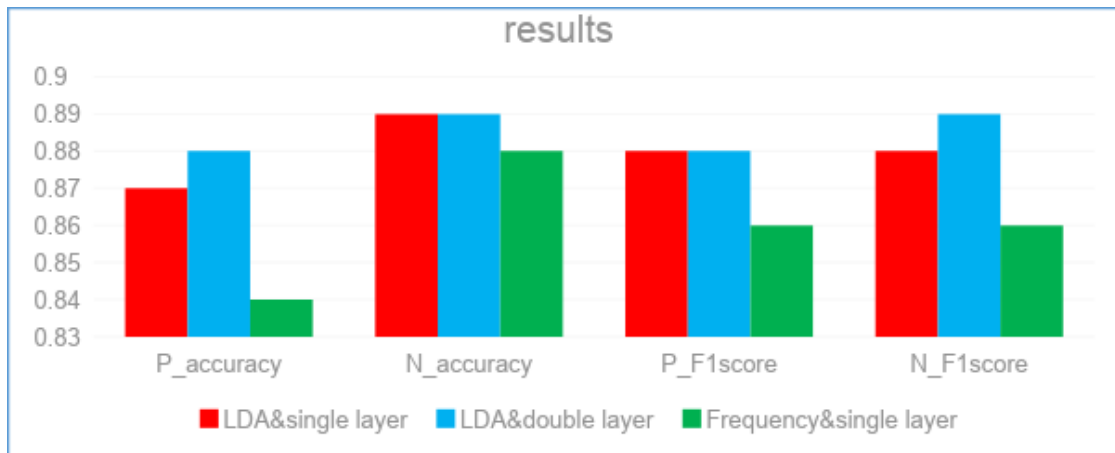
5.2. BI-LSTM RESULTS

In the experiment of the Bi-LSTM network classification model [18], one layer and two layers of Bi-LSTM are used, and each layer is compared with different nodes. Here, the dimensions of the words are uniformly taken as $n=256$.

Through the observation [20] of the results in the table, the comprehensive accuracy rate and the value of F1 are considered. The classification results of LDA extracted feature words of single layer and double layer are compared with the classification results characterized by word frequency, and constructed by LDA feature words. When the dictionary is $m=200$, the two-layer stacked Bi-LSTM, the 64-node model and the dictionary constructed with LDA feature words are better than the same under the same conditions when the $m=300$ is used, the single-layer Bi-LSTM and 64-node models are better than the classification result of word frequency characteristic words.

The experimental results are shown in Figure 1:

Figure 1. Different results comparison



6. EXPERIMENTAL SUMMARY

This experiment mainly discusses the advantages and disadvantages of the two methods when quantifying words in feature engineering, and makes targeted improvement to specific classification tasks based on theoretical knowledge, and demonstrates through experimental results. Through experimental comparison, it is found that the data is softly reduced by using LDA as the prototype, and the data content of 50,000 comments and several words of each comment is transformed into a matrix of (k, m), k is much less than 50,000, and then Feature extraction is performed on the (k, m) matrix for specific task requirements. The results show that the classification result of the feature extraction words after adding prior knowledge has certain advantages over the single word frequency characteristic words.

There are also deficiencies in this paper. The selection of the number of topics K for the first classification and the alpha value of the dirichlet distribution do not explore the optimal values in more detail. In this experiment, only LSTM and Bi-LSTM which perform well in classification effect are selected for experimental verification, and the use of multiple classifiers to verify the validity of double-layer LDA feature words can more effectively confirm its validity.

7. CONCLUSION

In the past ten years, various technologies of deep learning and artificial intelligence have developed rapidly, while language and words have become the medium of human communication, and natural language processing technology has received extensive attention and research. As a sub-task of natural language processing, sentiment analysis has become an important technical means for contemporary enterprises to plan production plans and develop functional strategies.

LDA model can be used for clustering regardless of language and being soft dimensionality reduction, which can be applied to feature engineering to improve classification efficiency. But the theory based on the word bag model (BOW) is always deficient.

Although the classification model used in this experiment has already had a good effect on the emotional classification result, there is still room for improvement. Finding better loss function and activation function, and more concise and powerful word vector expression technology can further improve the accuracy of the model. For the Bi-LSTM model, the author thinks that although it is called two-way, it is always a superposition of two LSTM models. It can't be called a two-way network in the true sense, and there is still room for improvement..

REFERENCES

- [1]Heinrich G. Parameter estimation for text analysis[R]. Technical report, 2005.
- [2]Mimno, David, and Andrew McCallum. "Topic models conditioned on arbitrary features with dirichlet-multinomial regression." arXiv preprint arXiv:1206.3278 (2012).
- [3]Liu G, Guo J. Bidirectional LSTM with attention mechanism and convolutional layer for text classification[J]. Neurocomputing, 2019.
- [4]Huang T, Shen G, Deng Z H. Leap-LSTM: Enhancing Long Short-Term Memory for Text Categorization[J]. arXiv preprint arXiv:1905.11558, 2019.
- [5]Jeffrey Pennington, Richard Socher, Christopher D. Manning,GloVe: Global Vectors forWord Representation,Computer Science Department, Stanford University, Stanford, CA 94305
- [6]Tomas Mikolov,Distributed Representations of Words and Phrases and their
- [7]Zhu Y, Gao X, Zhang W, et al. A Bi-Directional LSTM-CNN Model with Attention for Aspect-Level Text Classification[J]. Future Internet, 2018, 10(12): 116.
- [8]Sun C, Liu Y, Jia C, et al. Recognizing Text Entailment via Bidirectional LSTM Model with Inner-Attention[C]//International Conference on Intelligent Computing. Springer, Cham, 2017: 448-457. Compositionality,arXiv:1310.4546 [cs.CL] Wed, 16 Oct 2013
- [9]Xin Rong ronxin @umich.edu,word2vec Parameter Learning Explained,arXiv:1411.2738v4 [cs.CL] 5 Jun 2016
- [10]Hesam Amoualian,Wei Lu,Eric Gaussier et,al,Topical Coherence in LDA-based Models through Induced Segmentation,Meeting of the Association for Computational Linguistics.2018.4
- [11]Yue Zhang, Qi Liu and Linfeng Song,Sentence-State LSTM for Text Representation,arXiv:1805.02474v1 [cs.CL] 7 May 2018
- [12]Xiaodong Liu, Pengcheng He, Weizhu Chen, Jianfeng Gao1,Multi-Task Deep Neural Networks for Natural Language Understanding,rep,arXiv:1901.11504v1 [cs.CL] 31 Jan 2019
- [13]Sergey Edunov, Alexei Baevski, Michael Auli,Pre-trained Language Model Representations for Language Generation,arXiv:1903.09722v1 [cs.CL] 22 Mar 2019
- [14]Rojas-Barahona, Lina Maria,Deep learning for sentiment analysis,Language and Linguistics Compass,ALC,2012
- [15]Jiangfeng Zeng, Xiao Ma, Ke Zhou:Enhancing Attention-Based LSTM With Position Context for Aspect-Level Sentiment Classification. IEEE Access 7: 20462-20471 (2019)
- [16]Shervin Minaee, Elham Azimi, AmirAli Abdolrashidi:Deep-Sentiment: Sentiment Analysis Using Ensemble of CNN and Bi-LSTM Models. CoRRabs/1904.04206 (2019)
- [17]Subarno Pal, Soumadip Ghosh, Amitava Nag:Sentiment Analysis in the Light of LSTM Recurrent NeuralNetworks. IJSE 9(1): 33-39 (2018)
- [18]Huanhuan Yuan, Yongli Wang, Xia Feng, Shurong Sun:Sentiment Analysis Based on Weighted Word2vec and Att-LSTM. CSAI/ICIMT 2018: 420-424
- [19]Dipanjan Starkar,Text Analylics With Python,

[20]Cameron Davidson-Pilon, BAYESIAN METHODS FOR Hackers probabilistic and Bayesian inference.

Du jing wei

Male, master, major : natural language processing,sentiment analysis

