Sentiment analysis of movie reviews based on deep learning

Fuqian Zhang$^{1,a}$, Qingtao Zeng$^{2,b,*}$ Likun Lu$^{3,c}$ and Yeli Li$^{4,d}$

$^{1}$School of Information Engineering, Beijing Institute of Graphic Communication, Beijing, 100000, China
$^{2}$ School of Information Engineering, Beijing Institute of Graphic Communication, Beijing, 100000, China
$^{3}$School of Information Engineering, Beijing Institute of Graphic Communication, Beijing, 100000, China
$^{4}$School of Information Engineering, Beijing Institute of Graphic Communication, Beijing, 100000, China

$^a$email: 513987760@qq.com, $^b$email: zengqingtao@bigc.edu.cn, $^c$email: lklu@bigc.edu.cn, $^d$email: liyl@bigc.edu.cn

*Corresponding author’s e-mail: zengqingtao@bigc.edu.cn

Abstract. In recent years, with the rapid development of NLP (Natural Language Processing) and deep learning, public opinion and public opinion on the Internet have decreased a lot compared to the past. Many Internet users have changed from mere "bystanders" to disseminators of Internet information. Movie review sentiment analysis technology is an emerging category in the field of information mining. More and more people have joined the "review army". The quality of a movie is closely related to movie reviews! The manual screening method not only consumes a lot of manpower and material resources, but also is inefficient. Therefore, the use of deep learning-based sentiment analysis has become the current general trend. Based on the principle of word mosaic (word vector) and deep learning, this paper proposes a movie review sentiment analysis technology based on deep learning and machine learning word mosaic. Experiments show that the method used in this article has reached the correct rate of emotional classification of movie reviews. 83.13%, the experimental results prove the practicability and scalability of this method and the effectiveness of this method.

1. Introduction

With the rapid development of communications and the Internet in recent years, the penetration rate of the Internet has greatly increased. More and more viewers have joined the Internet family. The good and bad parts of movies or products can be reflected through comments, so online Movie reviews have become an important factor affecting the reputation of movies. Online movie reviews contain a lot of pictures, language, and text. Some self-media platforms such as Weibo and short video platforms contain a lot of movie reviews and articles. How to use machines to distinguish whether the emotions contained in these comments are positive or negative is an important challenge faced by current natural language processing.

Sentiment analysis is a branch of data mining, which is a sentence of emotional state. Sentiment analysis technology is used in many fields, such as e-commerce, public opinion analysis, market voice, consumer voice. At present, three main methods of emotion dictionary, machine learning and deep
learning (sentiment dictionary and The combination of rules) perform sentiment analysis at different text levels. These two methods require you to manually select features first, and then combine these features with machine learning methods for sentiment classification. The quality of feature selection has a great impact on the classification results.

The sentiment dictionary analysis method requires manual construction of data. The effect is poor, and the applicability and scalability are not strong. The sentiment analysis method based on deep learning is currently the mainstream. This article will start from the text vectorization and deep learning model construction.

Nasukawa[1] et al. proposed the concept of sentiment analysis in 2003, and it caused a great sensation on the Internet platform at that time. Collobert[2] et al. obtained the C&W model through training on the data set. In 2012, Liu[3] classified the branch of sentiment analysis in detail, dividing the text into document-level, sentence-level, and word-level, and divided sentiment analysis methods into methods using sentiment dictionaries and machine learning methods.

2. Materials and Methods

2.1 Introduction to word embedding

Word embedding is a digital representation of a word, which is usually represented by mapping it to a higher-dimensional vector (word vector). Word embedding is an unsupervised learning method that can learn implicit information about words from a corpus. The distributed word vectors obtained through training have the syntactic and semantic relations of words. Mikolov[4][5] et al. used CBOW[6] and Skip-Gram[7] to build a word2vec model. The model uses softmax as the output of the model. However, when there are many categories, softmax will sort all words and then take the maximum value. Compared with the traditional bag-of-words model that uses words as simple text functions, the new word embedding method is a feature extraction method based on deep learning algorithms that implicitly learn the grammatical and semantic information in words. This method avoids the heavy manual burden of artificial/semi-artificial emotional dictionaries and text annotations in sentiment analysis tasks.

The following figure shows the flow chart of the emotional analysis framework based on word vectors embodied in this article.

![Figure 1 Flow chart of sentiment analysis based on word vector](image)

There are some differences between Chinese and English text preprocessing. The goal of text preprocessing is to convert the text into a structured data format. Under normal circumstances, for convenience, a vector space model is usually used to represent text. VSM assumes that the order of words between documents does not affect the representation of the text, and represents the text as a vector of words and their frequencies. To give a simple example, suppose there are currently two texts:

I am a boy, I have an orange.
You are a girl. You have an apple.

We can abstract these two corpora into models, and then turn them into the following vector form.

|     | i  | m  | an | boy | have | Orange | you | are | girl | apple |
|-----|----|----|----|-----|------|--------|-----|-----|------|-------|
| Doc1| 2  | 0  | 1  | 1   | 1    | 1      | 1   | 0   | 0    | 0     |
| Doc2| 0  | 0  | 1  | 0   | 0    | 0      | 0   | 2   | 1    | 1     |

First use this model to establish an n-dimensional vector for each word, then correspond to each statement of the movie, use the vector corresponding to the word to establish a relative comment vector, and finally use machine learning or deep learning to train the data set we collected.

2.2 Neural network model

Sentiment analysis, general equipment, ego attention, systematic god economic network model entry network, parallel non-linear child layer combination and more model learning model ability.

Depth Shine Neural Network Ability to study in depth Depth non-linear display ability. This is the main cause of successful acquisition of each individual area in deep learning theory. As stated above, there is a certain limit in terms of the number of imports that are displayed and the number of imports that are displayed. Ability to display Japanese texts in a high-level model, a model of non-linearity, a model of non-linearity, and a bidirectional GRU network structure. General non-linear child layers Bi-directional GRU school Academic number Imported multi-level ego attention. This is a systematic calculation process that requires attention to the ego.

![Figure 2 Multi-head self-attention mechanism](image)

At the end of the model, the fully connected layer is used to input the feature information learned by the model network layer into the fully connected layer, and the text emotion and attitude classification is completed through the softmax function. It is to map all the output values of the neuron to the interval \((0, 1)\) to get the probability of each emotion category.

\[
S(x_i) = \frac{e^{x_i}}{\sum_{k=1}^{n} e^{x_k}}
\]
At the same time, in order to prevent the network model from over-fitting, the dropout layer was also used in the model we constructed. Dropout was first proposed by Hinton et al. [8] in 2012 and proved its effect.

The principle of Dropout is shown in the figure. During the learning process of the neural network, some neurons are temporarily deleted from the network with a certain probability and do not participate in training. In the training process, the selection of which neurons are "dropout" is random, so the network is different during training in each batch. By using dropout in the neural network model, the learned parameters will not be overly dependent on some local features, which can enhance the generalization ability of the model and speed up the training of the model.

![Neural network without dropout](image1)

![Neural network with dropout](image2)

**Figure 3 Drop neural network comparison**

### 3. Experiments and results

#### 3.1 Experiment 1

This article collected a total of 6 data sets, each of which contains foreign languages and Chinese.

| Data set | Positive | Negative | Training data set | Test data set | Positive/negative (test) | Positive/negative (test) | Average word count |
|----------|----------|----------|------------------|---------------|--------------------------|--------------------------|--------------------|
| Data set 2 | 6984 | 7541 | 9009 | 4220 | 5001 | 1663/2232 | 65/97 |
| Data set 3 | 5234 | 7433 | 9241 | 4113 | 4198 | 800/2423 | 12/20 |
| Data set 4 | 6972 | 7322 | 9243 | 3974 | 3839 | 1799/2296 | 6/19 |
| Data set 5 | 14231 | 11223 | 15232 | 3281 | 8239 | 3401/3744 | 52/76 |
| Data set 6 | 5981 | 6999 | 9238 | 4005 | 3968 | 1655/2499 | 36/62 |
| Data set 7 | 12331 | 11563 | 14999 | 5239 | 5321 | 3266/3411 | 8/17 |
The purpose of Experiment 1 is to find out the influence of the dimensionality analysis results of word vectors selected for emotion. In Experiment 1, the independent variable is the dimensionality of the word vector. The word vector \((\text{wod}_1, \text{wod}_2, \ldots, \text{wod}_{n-1}, \text{wod}_n)\) is in which the dimension of the word vector is represented as \(n\). The dimension of the word vector is chosen as 150, 250, 350 and 450 respectively. Three combined methods are used to classify emotions for each data set of 6. As can be seen from Table 2, in each model, the dimension of the word vector has a relationship with the data set, and is closely related to the test accuracy of the data set. It belongs to the larger dimension of 350 or 450 word vectors. As the average review time becomes shorter, the dimensional performance of small word vectors is improved.

### Table 3 Combination method

| vector dimension | method     | Data set 2 | Data set 3 | Data set 4 | Data set 5 | Data set 6 |
|------------------|------------|------------|------------|------------|------------|------------|
| 150              | W2V+SVM    | 84.23%     | 78.99%     | 76.66%     | 85.21%     | 82.23%     |
| 250              | W2V+SVM    | 86.12%     | 79.32%     | 77.13%     | 85.69%     | 83.56%     |
| 350              | W2V+SVM    | 85.52%     | 80.11%     | 77.64%     | 85.89%     | 83.55%     |
| 450              | W2V+SVM    | 85.12%     | 80.14%     | 78.11%     | 86.87%     | 84.54%     |

The purpose of Experiment 1 is to find out the influence of the dimensionality analysis results of word vectors selected for emotion. In Experiment 1, the independent variable is the dimensionality of the word vector. The word vector \((\text{wod}_1, \text{wod}_2, \ldots, \text{wod}_{n-1}, \text{wod}_n)\) is in which the dimension of the word vector is represented as \(n\). The dimension of the word vector is chosen as 150, 250, 350 and 450 respectively. Three combined methods are used to classify emotions for each data set of 6. As can be seen from Table 2, in each model, the dimension of the word vector has a relationship with the data set, and is closely related to the test accuracy of the data set. It belongs to the larger dimension of 350 or 450 word vectors. As the average review time becomes shorter, the dimensional performance of small word vectors is improved.

### 3.2 experiment 2

The purpose of the second experiment is to study the classification effect of sentiment analysis methods based on word vectors in texts with different average word lengths. In Experiment 2, corpora with different average annotation lengths were used as independent variables. The average length of each corpus is 56, 42, 12, and 5.

The data in Table 3 reflects the accuracy of the data set when the average length of the corpus is different. As shown in Table 3, we can see from this table that this method has an effective effect on the average length of sentences. The highest accuracy of this (W2V + SVM) combination method in Data Set 2 is 86.33%. With the shortening of the average review time, from 67 words per review to 11 words per review, the accuracy of the test set dropped from 86.33% to 76.23%. Using the short text review corpus, the accuracy of the test set reached 80.65%.

### Table 4 Corpus accuracy rate

| Number of dimensions | method     | Data set 2 | Data set 3 | Data set 4 | Data set 5 |
|----------------------|------------|------------|------------|------------|------------|
| Average sentence length | 56        | 42         | 12         | 5          |
| 100 W2V+SVM          | 85.23%     | 82.89%     | 79.34%     | 76.23%     |
| 200 W2V+SVM          | 86.33%     | 83.24%     | 80.12%     | 78.32%     |
| 300 W2V+SVM          | 85.13%     | 84.22%     | 80.22%     | 77.24%     |
| 400 W2V+SVM          | 85.77%     | 83.23%     | 79.69%     | 77.11%     |

### 4. Conclusion and Outlook

The word embedding and deep learning methods used in this article have a good effect in the emotional analysis of movie reviews. Sentiment analysis based on word embedding has high accuracy and scalability, and shows strong adaptability in front of a large number of texts and sentences, and highlights strong performance in shorter sentences. From the experimental results, it can also be seen that the dimension of the vector is important for sentiment analysis. The accuracy has a certain impact. The larger the dimension of the vector, the higher the accuracy of emotion classification, but for shorter corpus, the lower the dimension of the vector, the higher the accuracy of emotion classification. This article also has some shortcomings. In the calculation of sentiment analysis, the role of some special symbols, such as text symbols or emojis, is not considered. In the future, I will further expand the data set, more granular the corpus samples taken, and increase the diversity of sentences and improve the accuracy.
Acknowledgments
(1) Major Science and Technology Projects of Guangdong Province in 2019, No. 190826175545233
(2) Beijing science and technology innovation service capability construction project (PXM2016_014223_000025)
(3) BIGC Project (Ec202007)

References
[1] Nasukawa T, Yi J. Sentiment analysis: Capturing favorability using natural language processing[C]. Proceedings of the 2nd international conference on Knowledge capture. ACM, 2003: 70-77.
[2] COLLOBERT R, WESTON J, KARLEN M, et al. Natural Language Processing (Almost) from Scratch[J]. Journal of Ma-chine Learning Research, 2011(1): 2493-2537.
[3] Liu B. Sentiment analysis and opinion mining[J]. Synthesis Lectures on Human Language Technologies, 2012, 5(1): 1-167.
[4] Mikolov T, Chen K, Corrado G, et al. Estimation of Word Representations in Vector Space[C]. International Conference on Learning Representations. 2013:101-109.
[5] Mikolov T, Sutskever I, Chen K, et al. Distributed Representations of Words and Phrases and their Compositionality[J]. Advances in Neural Information Processing Systems, 2013, 26: 3111-3119.
[6] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]// Advances in Neural Information Processing Systems, 2013, 3111~3119.
[7] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[C]// Proceedings of the International Conference on Learning Representations (ICLR 2013), arXiv:1301.3781.
[8] Paccanaro A, Hinton G E. Learning distributed representations of concepts using linear relational embedding[J]. IEEE Transactions on Knowledge and Data Engineering, 2001, 13(2): 232-244.