Sentiment Analysis by Double Classification of Takeaway Platform Reviews Based on Deep Learning LSTM Models

Yunzhi Liao

School of Information Engineering, Chang'an University, Taiyuan, 710064, China
Email: Lyz13453171040@163.com

Abstract:

Sentiment analysis has a wide range of applications in the fields of opinion analysis, sentiment dialog, and product reviews. However, the sentiment information expressed in texts under different topics varies greatly; for example, a model that performs well on a movie review set has poor model classification on a social platform review set due to inconsistent recognition of antiphonal phrases, different expression of emoji sentiment, and missing contextual information. In this paper, the authors focus on tens of thousands of latest reviews of Chinese takeout platforms Meituan and Elema, and use the LSTM model in deep learning to double classify the data (positive and negative). This paper analyzes the performance of LSTM models in the field of sentiment analysis of takeout reviews and concludes that domain-specific text sentiment analysis requires specific analysis.

Keywords: Sentiment analysis; Deep learning; LSTM; Takeaway Platform Reviews

I. Introduction

According to statistics, as of January 2021, the number of Internet users worldwide exceeds 4.5 billion, and the number of social media users reaches 4.2 billion, with an unimaginable amount of data generated by users in the form of posts and comments every day. As one of the popular areas of machine learning, sentiment analysis is quite widely used in today's society. An analytical tool that can automatically analyze the sentiment tendency of users' text content can not only help enterprises or individuals understand users' views on events, attitudes or reactions to services, but also provide great help in the fields of heart disease treatment and emotional catharsis.

Currently, there are three main mainstream sentiment analysis methods: (1) dictionary and rule-based methods (2) machine learning-based methods (3) deep learning-based methods. Among them, the sentiment dictionary-based method is more traditional, which relies too much on the establishment of the dictionary, resulting in the inability to update the dictionary every day, as well as poor migrability and other problems. Sentiment analysis methods based on machine learning have some progress compared with dictionary analysis methods, but they also have shortcomings such as slow execution speed and inability to handle contextual information. In order to solve the above problems, deep learning-based sentiment analysis methods are beginning to be studied and applied, and some results have been achieved.

Feng et al. proposed the SCBILSTM model (Figure 1). This model combines a convolutional neural network with a self-attention mechanism and a bi-directional long and short-term memory network, which improves the performance compared to traditional neural network models. However, the effectiveness of this model has not been tested on multiple platforms [1].Li et al. proposed a P-BiLSTM-SA model that takes into account people's emotional personality traits and effectively categorizes comments into personality groups based on the user's personality traits, which significantly improves the performance of the model compared to the model without personality analysis, but it does not take into account the influence of emojis, face characters and mixed language expressions on sentiment [2].Yang et al. proposed a sophisticated deep learning model DBG-GBGCN, which contains sentiment classification results combining deep semantic features and syntactic structural information. The model is more stable in the test results of other public datasets and can handle some of the context-implicit information better. Due to excessive consideration of textual implicit meanings, the established structure appears redundant and has poor computational performance [3]. Lv et al. analyzed customer reviews of tourism brands and used a fine-grained sentiment analysis model to classify tourist reviews of "Fantasia Oriental Myth Theme Park" into seven categories. The study did not address the impact of invalid reviews such as antiphonal reviews, invalid reviews and machine generated reviews [4].

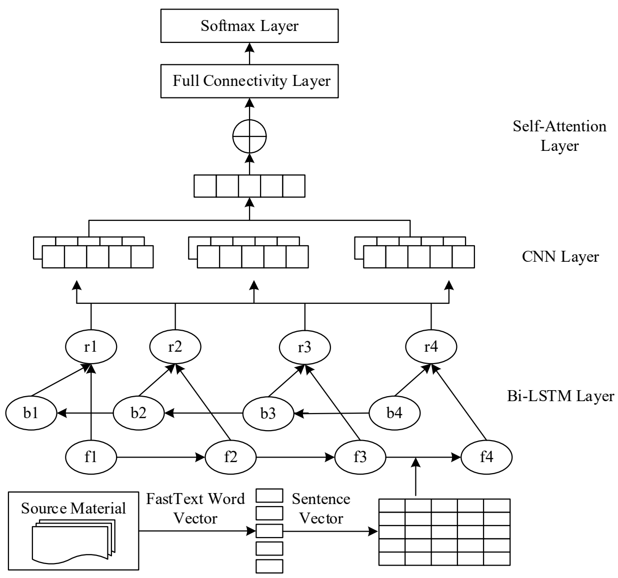

Figure 1

Abroad, Kora et al. proposed a meta-ensemble deep learning method (combining a three-level meta-learner and synthesizing the analysis results of multiple deep learning models), which trained as many as 114 models using the "Arab-Egyptian Corpus", exceeding six basic deep learning models. However, it relies on trained models and suffers from the curse of dimensionality when dealing with high-latitude data [5].Zhang et al. applied sentiment analysis to mobile big data using a hybrid sentiment lexicon approach, and concluded that average word lengths of less than 2.3 are more effective, and that the MoSa system effectively analyzes the sentiment of public opinion [6].D'Aniello et al. based on the fact that sentiment, affect, emotion, and opinion are completely different concepts and proposed KnowMIS-ABSA model with different tools to measure the dimensions of opinions to improve accuracy in the marketing domain [7].Sweidan et al. proposed a sentence-level ontology-XLNet hybrid sentiment analysis classification method for enhanced feature extraction, but in some cases, such as when the context is long, the XLNet's performance drops drastically and the accuracy is lower than models such as BERT [8].

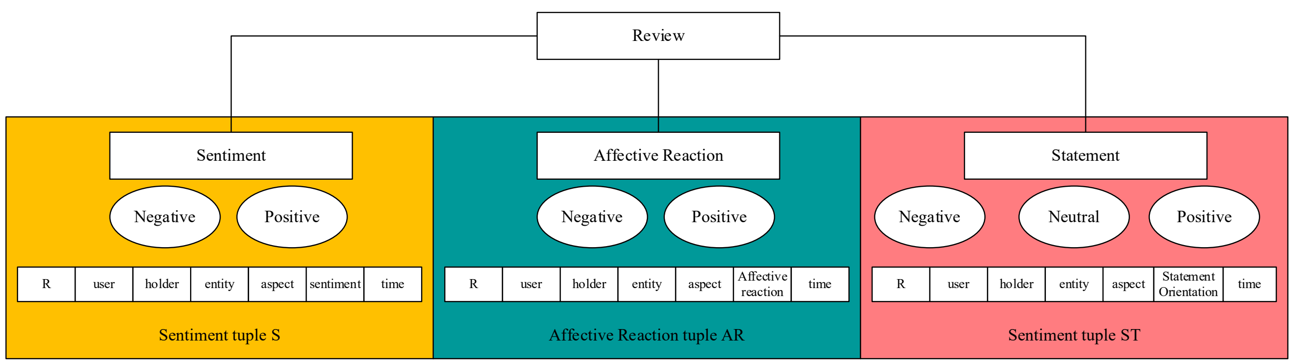

Figure 2

However, most of the current research on sentiment analysis focuses on the improvement of model design, and there is not much scenario-specific text sentiment analysis. This paper uses the traditional LSTM model to perform bi-classification sentiment analysis on tens of thousands of reviews on domestic and international seller platforms (Meituan and Elema) in 2023, and uses cross-validation techniques to train and test the samples several times, continuously reduce the loss function value, test the model performance by accuracy and F1-score, and analyze the causes of classification errors and the direction of the future improvements that can be made.

II. Material and Method

A. Experimental platform

The experimental platform and experimental environment of this paper are shown in Table 1.

Table 1 Experimental Platform

Experimental Environment	Specific Information
Operating system	Microsoft Windows 10 Family Chinese Version(64-bit)
Processor	AMD Ryzen 7 5800H with Radeon Graphics 3.20 GHz
Graphics Card	NVIDIA GeForce RTX 3050 Ti Laptop GPU
Development Language	Python 3.11
Development Tool	Pycharm 2021.3.2

B. Data set

This experiment uses the Chinese takeout reviews data set provided by AiDigital (containing about 12,000 Chinese takeout reviews, of which 8,000 are positive and 4,000 are negative) and the corresponding sentiment labels, as well as the year 2023 review data of Meituan and Elema for training and testing. The reviews obtained by the crawler were manually filtered to obtain 1500 positive and 1300 negative reviews. The text removes comments that contain both positive and negative comments to improve the binary classification performance of the model. For example, “Hot dry noodles don’t have much flavor, too much oil! The bean skin tasted pretty good, with more dried beans and lots of glutinous rice and diced meat.”

C. Model

LSTM models (shown in Fig. 3) have been proposed and applied to deep learning as early as the last century, and contextual contextual considerations and the fact that gradients can vanish or explode with long sequences are well suited to LSTM algorithms.

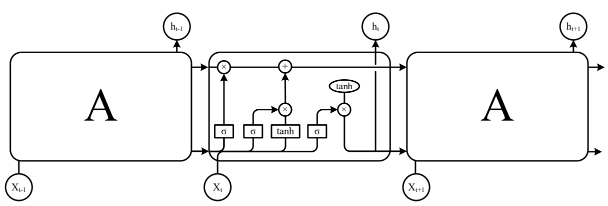

Figure 3

The repetition module in LSTM contains four interacting layers (three sigmoid layers and one tanh layer) and interacts in a very specific way. Three elaborate structures called “gates” (forgetting gates, input gates, and output gates) are used to remove or add information in the cell state. They consist of a sigmoid neural net layer and a point multiplication operation. In the schematic, each line from the output of one node to the input of the other nodes conveys a complete vector, i.e., a processing matrix representing the sentence.

The text uses a 2-layer LSTM model and the output of this model is used as input to a fully connected linear layer which, due to binary sentiment analysis, outputs a result of dimension 1 followed by a Dropout. if the final result is greater than or equal to 0.5, it outputs a positive value; otherwise it outputs a negative value. In this paper, binary cross entropy is used as a loss function to represent the convergence of the results.

D. Experimental procedure

To ensure that the model performs better in takeout review sentiment analysis, this paper used the Chinese takeout review data set from AiDigital for initial training. Then, retraining and testing were performed using positive and negative sentiment-attitude reviews from Meituan and “Elema” for the year 2023. Data cleaning was performed on the review texts by loading deactivated words and using a cleaning function. In this paper, utf-8 encoding is considered. After that, sentences are segmented using jieba and represented by word2vec algorithm using high dimensional vectors (word embedding) and real vectors are used to represent sentences by putting words with similar meanings together using matrices. The inputs in matrix form are then encoded as lower dimensional one-dimensional vectors, which are used as input parameters for the LSTM model..

The text was crawled to obtain 1500 positive and 1300 negative comments and then divided into 5 disjoint sets, each set was divided into training and testing sets in a ratio of 5:1. The model was trained in 5 iterations of 200 comments each.

Ultimately these data were divided into five sessions for training and testing. The data from the 5 tests were also averaged to produce the final test results (accuracy and F1-score) for this experiment, as shown below.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} \tag{1}$$

$$\text{F1} - \text{score} = \frac{2TP}{2TP + FN + FP} \tag{2}$$

● TP : True Positive FN : False Negative FP : False Positive TN : True Negative

In addition, more than 300 hotel reviews in the ChnSentiCorp_htl_all data set and more than 300 movie reviews in the dmsc_v2 data set were selected for comparative experiments in this paper, in order to analyze the performance of the trained LSTM models in sentiment analysis of different topics.

III. Results

A. Results of takeaway reviews on LSTM

In this paper, the model was first constructed by using data set from AiDigital to converge the analytical accuracy to more than 70 percent, after which the performance of each epoch is shown in Table 2. The final cross-entropy loss function value image during the five training sessions is shown in Figure 4.

Table 2 Experimental Results

Times	Accuracy(%)	F1(%)
1	78.47	75.65
2	85.22	82.95
3	89.13	86.11

4	90.75	88.63
5	91.06	89.47
Average	86.93	84.56

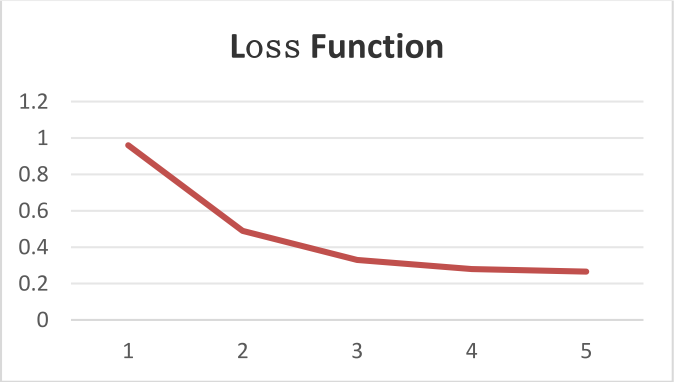

Figure 4

As can be seen from the above results, after 5-fold cross validation, the loss function has converged to a lower stable value. However, during the training process, due to the complexity of the model structure, the iterative computation is more time-consuming, while the performance is average compared to some of the latest hybrid sentiment classification models, and the final F1-score is difficult to reach a lower value.

By examining the miss-classified texts, this paper found that the model performs poorly in detecting antiphonal language and analyzing comments with insignificant emotional attitudes, e.g., “The rider delivered it so fast, it was delivered in only two hours.” , “The rice is a bit cold, the meat given in the dish is quite a lot, it’s just that the package is a bit skimpy,” and so on.

IX. Results of the analysis of film reviews and social media comments

This paper uses trained LSTM models for binary sentiment analysis against movie and hotel reviews, and the results are compared with the final results of the takeaway data set as in Table 3.

Table 3 Comparative results

Field	Accuracy(%)	F1(%)
Takeaway Comments	86.93	84.56
Hotel Comments	81.14	76.45
Film Reviews	80.05	74.94

From the above table, it can be seen that under the classification of the trained LSTM model, the sentiment analysis of takeaway reviews improves by 5.82%, 6.91%, 8.11%, and 9.62% in accuracy and F1-score compared to hotel reviews and movie reviews respectively, mainly because the model takes into account the specific sentiment characteristics of takeaway reviews. For example, “the battle scenes are so fast and detailed that it is boring to watch” is incorrectly categorized as positive, “the place is a bit remote, the room environment is not bad, and the price-performance ratio is not bad.” The model is wrongly labeled as negative. Therefore, the model trained by choosing specific topic data can improve the accuracy of sentiment analysis.

IV. Conclusion

In this paper, the sentiment analysis of Meituan and Elema takeout reviews in 2023 is performed, and the performance of analyzing positive and negative sentiment attitudes is gradually improved in iterative training. The LSTM deep learning model used not only considers the sentiment impact of words before and after the sentence, but also takes advantage of the fact that LSTM can solve the long sequence gradient vanishing and explosion mechanism to improve the accuracy and execution efficiency of the prediction, which is validated with the experimental results. Experimental results are also used to validate the results. In addition, this paper also performs sentiment analysis on movie reviews and social media reviews, and their performance is significantly lower than that of takeaway reviews, concluding that domain-specific text sentiment analysis needs to be analyzed specifically.

This paper did not optimize the model, resulting in less analytical results than a structurally complex model, and the data tested only included takeaway reviews from Meituan and Elema in 2023. In addition, this paper did not consider reviews that contain both positive and negative feedback, nor did it consider facial emoticons, which are not a small percentage in practical applications. More fine-grained sentiment analysis, more comprehensive data set testing, and better performing models can be considered in the future.

Reference

Feng Yuanyuan,Liu Kejian,Li Weihao. Multi-personality microblog sentiment classification based on BiLSTM+Self-Attention[J]. Journal of Xihua University(Natural Science Edition),2022,41(01):67-76.
Li Lihua,Hu Xiaolong. Text sentiment analysis based on deep learning[J]. Journal of Hubei University(Natural Science Edition),2020,42(02):142-149.
YANG Chunxia,XU Ben,CHEN Qigang et al. An aspect-level sentiment analysis model fusing deep BiGRU and global graph convolution[J]. Small Microcomputer Systems,2023,44(01):132-139.DOI:10.20009/j.cnki.21-1106/TP.2021-0386.
Lv Jiaxin, Ni Yuan, Han Pengfei. Measurement of customer fit value of cultural tourism brands - based on a fine-grained sentiment analysis model[J]. Investment and Entrepreneurship,2023,34(01):162-164.
Kora, R., Mohammed, A. An enhanced approach for sentiment analysis based on meta-ensemble deep learning. Soc. Netw. Anal. Min. 13, 38 (2023). https://doi.org/10.1007/s13278-023-01043-6
Zhang Y , Ren W , Zhu T ,et al.MoSa: A Modeling and Sentiment Analysis System for Mobile Application Big Data[J]. Symmetry, 2019, 11(1).DOI:10.3390/sym11010115.
D’Aniello G , Gaeta M , La Rocca I .KnowMIS-ABSA: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis[J].Artificial Intelligence Review:1-32[2023-12-16].DOI:10.1007/s10462-021-10134-9.
Sweidan A H , Bendary N E , Feel H A .Sentence-Level Aspect-Based Sentiment Analysis for Classifying Adverse Drug Reactions (ADRs) Using Hybrid Ontology-XLNet Transfer Learning[J].IEEE Access, 2021.DOI:10.1109/ACCESS.2021.3091394.