Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 15, No. 3, 2024

Using the Bert model and the attention mechanism to obtain an accurate sentiment analysis model

Jianxiang Ma

Chengdu University of Technology, Chengdu, China
783905508@qq.Com

Abstract: This study takes the network reviews captured by Ctrip as the object of investigation. 29,025 reviews are preprocessed first, and the training set, verification set and test set are divided according to proportion. The Bert model is used for training and data parameters are retained. Only the subject and affective adverbs were retained and the negative adverbs were artificially weighted. The data set was retrained. The research results supplemented the innovation and application value of the training results, and could prove whether there were commercial malicious competition and other phenomena in bad reviews in network evaluations.

Keywords: Bert model, Attention mechanism, Negative adverb weight, Accuracy vs. Epochs, Precision vs. Epochs, Recall vs. Epochs, F1-score vs. Epochs.

1. Introduction

In recent years, with the development of network technology and the continuous development of all media, information is everywhere and no one knows it. Trading platforms, travel card online shopping and tourism industries have also risen. But at the same time, the emergence of malicious competition between businesses is also worrying. According to the impact of the authenticity of online reviews on the formation of tourist trust published online for the first time by the Travel Research Journal Network in 2019, it is found that more and more users use large review websites as information sources during the survey process. The empirical results show that the authenticity and trust of online reviews play an important role in the context of online tourism. Scandinavian Hotel and Travel magazine, also mentions social media as one of the most important sources of information for tourists. In particular, the opinions obtained from peers through online travel reviews have a great influence in the travel decision-making process(Nowacki&Niezgoda, 2020).Online travel reviews are often outside the control of the destination marketer and are therefore considered more authentic (Munar, 2012).When travelers read online travel reviews and are informed as a result, their impressions of destinations and businesses change (Nowacki&Niezgoda, 2020).Therefore, it is important to understand why online travel reviews are so powerful in influencing the image of a destination. High-rated reviews changed the perceived image more than low-rated reviews. Low rating reviews have a strong impact on emotional image. In high-rating scenarios, specific travel reviews are more likely to enhance the image of a destination than abstract travel reviews. In the low rating scenario, the deterioration of the image of the destination by abstract travel reviews was greater than that by specific travel reviews. Based on this, this study believes that in the highly developed all-media condition, whether malicious evaluation and random evaluation exist in network evaluation is the key factor. The purpose of this study is to reflect the application value, to more accurately analyze the phenomenon of malicious bad reviews and random ratings on network evaluations, and to provide references for the ratings and evaluations of various travel shopping software.

2. Literature review

Long Short-Term Memory (LSTM) is a special recurrent neural network (RNN), which is specially designed to solve the long-term dependence problem in traditional RNN. It effectively captures and remembers key information in long sequences by introducing a gating mechanism, which is suitable for many tasks that need to model time series or language sequences. The Bert model is a pre-trained language model that learns common language representations from text corpora through large-scale unsupervised learning, which can significantly improve a variety of natural language processing tasks. The BERT model adopts a pre-training-tuning method. In the pre-training phase, a large amount of unlabeled text data is used to learn the language representation. This stage uses the Transformer structure, which is a neural network architecture based on the attention mechanism and can capture long-distance dependencies in text.

Are you interested in the analysis of online reviews at present, online goods are sold online? In order to improve personal economic interests, online evaluations have appeared indiscriminate evaluations, scoring indiscriminately, and even for the embarrassing situation between commercial competition. In the comments, 30,000 comments are crawled, and the Bert model is selected for analysis. Coupled with negative adverbs and weights, if the accuracy rate is improved, the feasibility of the model training is proved. It can accurately determine whether there is indiscriminate evaluation or even whether there are false statements in the evaluation, misleading dissemination and deliberate defamation of false statements.

Li Lei et al. [6] proposed a fusion object recognition and Bi-LSTM model sentiment analysis model, which mainly identifies key evaluation objects through CNN, fuses evaluation objects with text information, and analyzes the final sentiment category through Bi-LSTM model. However, the most important part of the evaluation statement is the emotional description part, which expresses the user 's point of view. Liu et al. [5] used a dictionary-based sentiment analysis model to analyze online reviews of Chinese tourists in Australian destinations. This technology requires a lot of manual intervention, and the final analysis results rely too much on the quality and rules of the dictionary. With the rapid growth of information, it is impossible to complete the task only by manual processing.Law et al. [9] used long short-term memory(LSTM) neural network to predict tourism flow, and proved that the performance of LSTM method is better than other traditional methods. However, LSTM only has the ability to remember the forward information, but not the backward sequence. Grabner et al. [10] proposed a sentiment analysis dictionary for the classification of tourist reviews ; adhi et al. [8] designed a sentiment analysis model based on naive Bayesian classifier and semantic extension method. Experiments show that this method can improve the accuracy of sentiment analysis. However, machine learning-based methods rely too much on the quality of labeled data and require complex feature engineering.

3. Introduction to relevant models

3.1. Bert model

BERT 's Masked Language Model (MLM) mission

One of the core innovations of BERT is Masked Language Model(MLM). is used to predict obscured words through context. In the MLM task, some words in the input sequence are replaced with special markers. For these obscured words, the goal of the model is to predict their original words through other words in the previous and subsequent texts, assuming that the input sequence is, where xi is replaced by. The model predicts through the following steps:

Coding context: Through Transformer Encoder, the context representation of each word is generated, where hi contains the context information of xi.

Prediction of obscured words: Through a fully connected layer and softmax layer, the model generates the probability distribution of each word on the vocabulary:

$$P(x_i \mid X) = softmax(W_o h_i + b_o)$$

Among them, and are trainable weights and biases. The model is trained by maximizing the correct prediction probability of these masked words.

3.2. Attention mechanism

Attention mechanism calculation process

Import Query, Key, Value:

Stage 1: Calculate the correlation or similarity between Query and Key(common method dot product, cosine similarity, MLP network) to obtain the attention score;

Dot product: Similarity Query, Key Query Key

Cosine similarity: Similarity (Query, Key i_i) $= \frac{Query-Key_i}{Query\cdot Key_i}$

MLP network: Similarity (Query, Key $_i$) $= MLP($ Query, Key $_i$)

Stage 2: Scale the attention score(divided by the root number of the dimension), and then the softmax function. On the one hand, it can be normalized, and the original calculated score is organized into a probability distribution with the sum of all element weights of 1 ; on the other hand, the weight of important elements can be more prominent through the internal mechanism of softmax. Generally, the following formula is used to calculate: $a_i = Softmax(Sim_i) = \frac{e^{Sim_i}}{\sum_{j=1}^{L_x} e^{Sim_j}}$

In the third stage, the Value value is weighted and summed according to the weight coefficient to obtain the Attention Value(at this time, V has some attention information, more important information is more concerned, and unimportant information is ignored):

$$Data\ presentationAttention(Query, Source) = \sum_{i=1}^{L_x} a_i \cdot Value_i$$

4. Experiment and analysis

4.1. Data presentation

This experiment took Ctrip, a popular tourist attraction in Harbin this year, Ice and Snow World as an example, and captured 29,025 comments on the Internet. Among them, 26,035 were positive and 2990 were negative, of which 409 were removed from the negative review, and the remaining 2582 were removed.

4.2. parameter setting

In this experiment, the unjustified criticism of black or urban man-made negative factors in the comments, and even rose to the critical attack question of region and character quality question 5, was evaluated as 1 temporarily, and the normal supported difference rating was 0. First, the model was trained on the first thousand comments. In the second training, irrelevant words and punctuation were screened again, and only emotional words were retained. The model was trained with all 2582 comments. The accuracy of the two training results is compared. If the accuracy increases, it represents the feasibility of the experiment. The Bert model is used for training, where x is the text and y is the dataset

4.3. Experimental result analysis

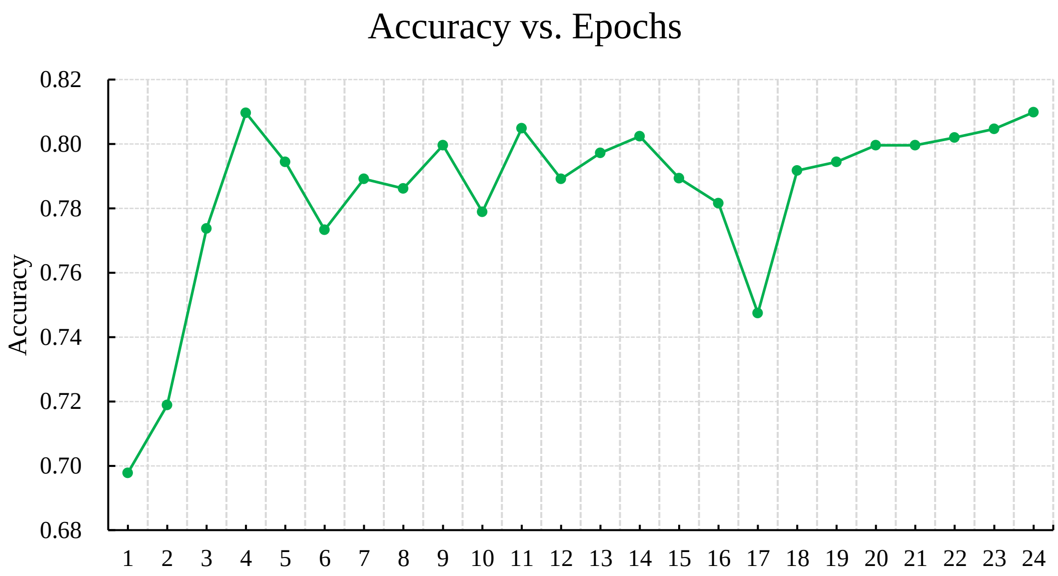
The experimental results show that the accuracy of the data set has reached 85% through Bert model training. The feasibility of the model training is fully proved. For example, the model is dealing with "very poor experience, because of the weather, the Ferris wheel stopped, the snow project basically stopped, and everyone queued up on the big slide and couldn't play for 6-7 hours." When you can pass "very poor", "basic", "not up".And other emotion-related information words can effectively identify the emotional tendency of the comment. The experimental model shows good performance, and the accuracy can reach 82%.

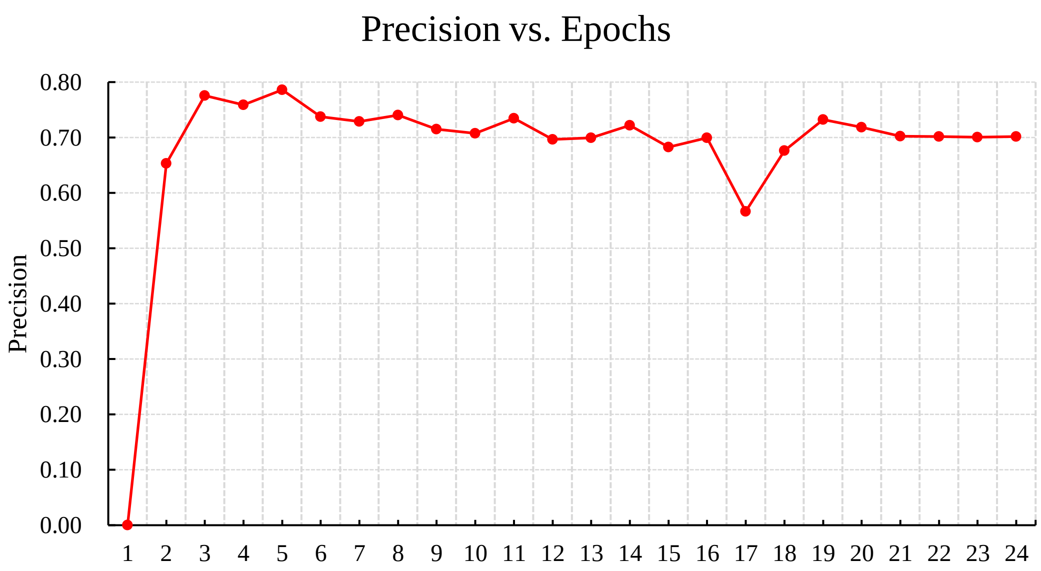
```
100%|██████████| 193/193 [00:37<00:00,  5.10it/s]
100%|██████████| 49/49 [00:03<00:00, 12.56it/s]
Epoch 14: Train Loss: 0.0905 | Eval Loss: 0.4884 | Accuracy: 0.8269 | Precision: 0.74
76 | Recall: 0.6525 | F1-Score: 0.6968
```

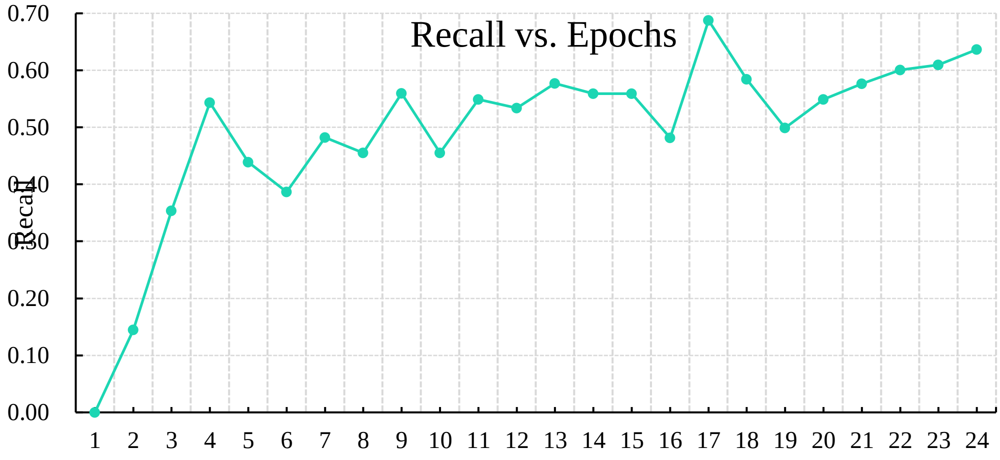
Fig. 1 Model training change rate

Table1. Model training results index

epoch	train_loss	eval_loss	accuracy	precision	recall	f1
1	0.147443655	0.585372205	0.697916667	0	0	0
2	0.143168555	0.545813539	0.71875	0.653846154	0.146551724	0.23943662
3	0.137891437	0.539196274	0.7734375	0.773584906	0.353448276	0.485207101
4	0.132104338	0.469410849	0.809895833	0.759036145	0.543103448	0.633165829
5	0.127322951	0.50426894	0.794270833	0.784615385	0.439655172	0.563535912
6	0.123176634	0.530458136	0.7734375	0.737704918	0.387931034	0.508474576
7	0.1196846	0.47454104	0.7890625	0.727272727	0.482758621	0.580310881
8	0.115482341	0.48064127	0.786458333	0.736111111	0.456896552	0.563829787
9	0.111966239	0.518839924	0.799479167	0.714285714	0.560344828	0.628019324
10	0.109254634	0.566069111	0.778645833	0.706666667	0.456896552	0.554973822
11	0.10753519	0.488655026	0.8046875	0.735632184	0.551724138	0.630541872
12	0.104262715	0.544369482	0.7890625	0.696629213	0.534482759	0.604878049
13	0.101123935	0.539753553	0.796875	0.697916667	0.577586207	0.632075472







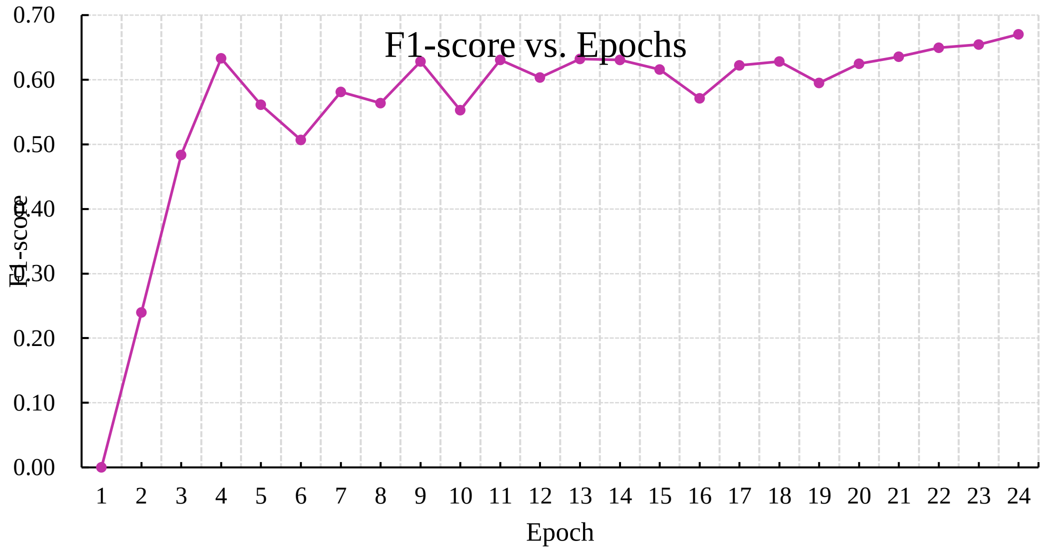


Figure 2. Each index changes with epochs

The four graphs above show four different evaluation metrics (Accuracy, Precision, Recall, F1-score) Variable with the training period (Epochs) curve. Each subparagraph corresponds to an indicator, which is used to show the performance change of the model under different training cycles.

Accuracy vs. Epochs (Accuracy varies with Epochs)

Chart content: The accuracy rate gradually increases with the progress of training, significantly improving in the early stage of training, and then showing some fluctuations in the middle and late stages.

Eventually it tends to stabilize between 0.78 and 0.80.

Unscramble: A rise in accuracy indicates that the model is gradually learning and adapting to the data, but medium-term fluctuations can mean that the model may be overfitting or underfitting at some points. The eventual stabilization means that the model has learned most of the information in the data.

Precision vs. Epochs (Accuracy varies with Epochs)

The chart content: the accuracy rate rises rapidly in the early stage and remains at a relatively stable level in the medium term, with occasional declines and fluctuations in the magnitude.

Explanation: The stability of the accuracy rate shows that the model can effectively distinguish positive and negative samples in most training cycles, but it is possible at the 17 th Epoch.

Some abnormal conditions were encountered, resulting in a temporary decrease in the accuracy rate.

Recall vs. Epochs (recall rate varies with Epochs)

Chart content: The recall rate is gradually increasing, indicating that the model is increasingly identifying positive samples during training. However, the fluctuation of the recall rate is also more obvious. Especially at the 16 th Epoch, there is a significant peak.

Interpretation: The rising trend of the recall rate shows that the model is gradually increasing in capturing the positive samples, but the fluctuation may indicate the performance of the model on some Epochs. It is not stable, especially at the 16 th Epoch when the model may briefly identify too many positive samples.

F1-score vs. Epochs (F1-score changes with Epochs)

Chart content: F1-score rises rapidly and remains stable in the next Epochs, indicating that the model has achieved good performance between precision and recall.

Balance. Despite some fluctuations, the overall trend is upward.

Interpretation: F1-score, as the harmonic mean of precision and recall, comprehensively reflects the overall performance of the model. The figure shows the model in most training weeks.

During the training period, a relatively balanced state was achieved, but there were still some fluctuations in the middle and late training periods, which may be due to the influence of data characteristics or model parameter adjustment. Ringing.

5. Summary

During the design period, I read a lot of English literature, and my literature reading ability has been greatly improved. Some of the literature comes from senior students and some from academic websites, and my retrieval ability has also been improved. In literature reading, I got familiar with the development of information extraction field, especially open information extraction field, understood and practiced the most cutting-edge trends in academia and industry, and practiced the latest methods in the field in experiments.

In this survey, it is found that compared with traditional machine emotion learning method, Bert judges the parts of speech of emotion-related words from the perspective of speech of speech and calculates the weight of each position in the evaluation statement with it as the core word, so as to integrate it into the actual context information of Bert evaluation statements, highlight the importance of emotion-related words, and thus improve the effect of emotion analysis. Parallel structures for Bert models and parts-of-speech attention analysis allow for more accurate analysis

The purpose of this paper is to automatically classify discussion topics by neural network model, accurately judge whether the network evaluation is positively correlated with the actual evaluation, and ensure that there is no malicious evaluation or random evaluation. Customs clearance analysis of "good", "bad", "too long" and other keywords to judge the satisfaction of tourists for scenic spots. By comparing the number of positive and negative correlation evaluations, we can get the reasons for the success and popularity of the Ice and Snow World and the existing shortcomings. The neural network model can better assist the staff to analyze the reasons for success and the shortcomings to be improved.

References

[1] PEISW,WANGLL. Text sentiment analysis based on attention mechanism[J]Computer Engineering and Science, 2019, 41 (2):343-353.

[2] ASGHAR M Z, KUNDI F M, AHMAD S, et al. T-SAF:Twitter senti· 46 · Issue 2 ment analysis framework using a hybrid classification scheme[J]. Ex- pert Systems, 2018, 35 (1):e12233.

[3] AFZAAL M, USMAN M, FONG A C M, et al. Multiaspect-based opinion c lassification model for tourist reviews[J]. Expert Systems, 2019, 36 (2):1-24.

[4] TAO Y G. Social media data-based sentiment analysis of tourists'air quality perceptions[J]. Sustainability, 2019, 11 (18)1-23.

[5] LIU Y, HUANG K X, BAO J G, et al. Listen to the voices from home:an analysis of chinese tourists′ sentiments regarding Austra- lian destinations[J].Tourism Management, 2019, 71:337-347.

[6] LI W J, QI F, TANG M, et al. Bidirectional LSTM with self-atten- tion mechanism and multi-channel features for sentimentclassifica- tion[J]. Neurocomputing, 2020, 387:63-77.

[7] HAO J X, FU Y, HSU C H, et al. Introducing news media sentiment analytics to residents'attitudes research[J]. Journal of Travel Re- search, 2020, 59 (8):1353-1369.

[8] MUHAMAD S A, MUHAMMAD Z N, USADA E. Pengaruh seman- tic expansion pada naïve Bayes classifier untuk analisis sentimen to- koh masyarakat[J]. Journal RESTI (Rekayasa Sistem dan Teknologi Informasi), 2019, 3 (2):141-147.

[9] GUAN P F, LI B A, LYU X Q, et al. Attention enhanced bi-direc- tional LSTM for sentiment analysis[J]. Journal of Chinese Informa- tion Processing, 2019, 33 (2):105-111. Guan Pengfei, Li Baoan, Lv Xueqiang, et al. Attention-enhanced bidirection al LSTM sentiment analysis [J]. Chinese Journal of Information, 2019,33(2): 1 05-111.

[10] KIM Y. Convolutional neural networks for sentence classification [C]//Doha:Proceedings of the 2014 Conference on Empirical Meth- ods in Natural Language Processing (EMNLP), 2014.