

# Research on Emotion Recognition Model of Takeaway Evaluation Text Based on LSTM-CNN

Ruiqing Gao

University of Technology Sydney, Sydney, Australia

---

**Abstract:** Takeaway evaluation is a literal evaluation of the products and services experienced by consumers, which not only provides a realistic basis for the improvement of products and services of merchants, but also affects the purchase decision of consumers in the future. In this paper, NLP(natural language processing) is used for preprocessing in tensorflow environment, and an emotion recognition model of takeaway evaluation text based on LSTM-CNN is established. The model is based on Bi-LSTM and attention mechanism, and uses Word2Vec method to vectorize text vocabulary. Then, the LSTM-CNN serial hybrid model is used to extract the context and local semantics of the text, and the effectiveness of the algorithm is verified by an example. The precision reaches 0.937, the recall rate reaches 0.896,F1 and the F1 value reaches 0.906. Through horizontal comparison, it can be found that the model in this paper is outstanding in this task, and the emotion enhancement model also improves the classification accuracy to some extent.

**Keywords:** LSTM-CNN, Takeaway, Evaluation Text, Emotion Recognition.

---

## 1. Introduction

In the fast-paced Internet era, more and more people use the take-away platform to order food, resulting in a large number of take-away comment data. Customers can choose goods according to the number and content of comments, and businesses can also mine valuable information from positive and negative comments to adjust food, sales and industrial structure [1]. Takeaway evaluation is a literal evaluation of the products and services experienced by consumers, which not only provides a realistic basis for the improvement of products and services of merchants, but also affects the purchase decision of consumers in the future. Takeaway merchant rating refers to the quantitative evaluation method of products or services provided by foreign sellers. Accurate take-away merchant rating can not only help the platform realize the management of survival of the fittest, but also reduce the cost of consumers' choice among a large number of similar merchants, which has important practical guiding significance for platform managers and consumers.

Massive texts come from many users of the Internet, with various forms and no fixed format, so it is difficult to process them by simple automatic means. If you rely on manual processing, there are problems such as excessive workload and poor real-time performance [2]. In order to realize the emotional analysis of massive texts, scholars analyze the emotional tendency of texts through statistics, machine learning and neural networks [3-5]. In this paper, NLP(natural language processing) method is used for preprocessing in tensorflow environment, and an emotion recognition model of takeaway evaluation text based on LSTM-CNN is established to carry out experiments on the takeaway data set. The experimental results show that the LSTM-CNN model has the highest accuracy.

## 2. Brief Introduction of Text Sentiment Analysis Method

Emotional analysis of text, also known as subjective analysis, refers to identifying and mining subjective

information in the original text by NPL, text mining and computer linguistics, so as to get the attitude and opinions of critics on the research object [6]. The analysis method based on emotion dictionary comes from the basic grammar rule text analysis, and the method is relatively simple. Based on the method of machine learning, it is necessary to manually label a large number of corpus before it can be used as a training set, and then extract text features through SVM (Support Vector Machine), random forest and naive Bayes algorithms, and finally build a classifier model, and then use the model to identify the emotional tendency of new texts after analyzing and verifying the model [7-8].

LSTM (Long Short Term Memory) is a complex circulatory neural network with long-term memory. LSTM uses three gates: input gate, forgetting gate and output gate to control information. In most NLP problems, in addition to the data that needs to be input first, the subsequent data is also needed [9]. For example, in the task of aspect-level emotion analysis, the emotional tendency of a target word is analyzed, and the emotional words expressing this aspect may appear in front of or behind the target word.

Because of the excellent performance of deep learning algorithm, many current researches on emotion recognition use deep learning related algorithms to build models. CNN (Convolutional Neural Network) can reduce the difference of input frequency and capture local information, but it does not consider the global characteristics and background. In short, the modeling capabilities of CNN and LSTM are limited [10-11]. On this basis, combining CNN and LSTM to build a network for speech emotion recognition, so as to learn the best description of information.

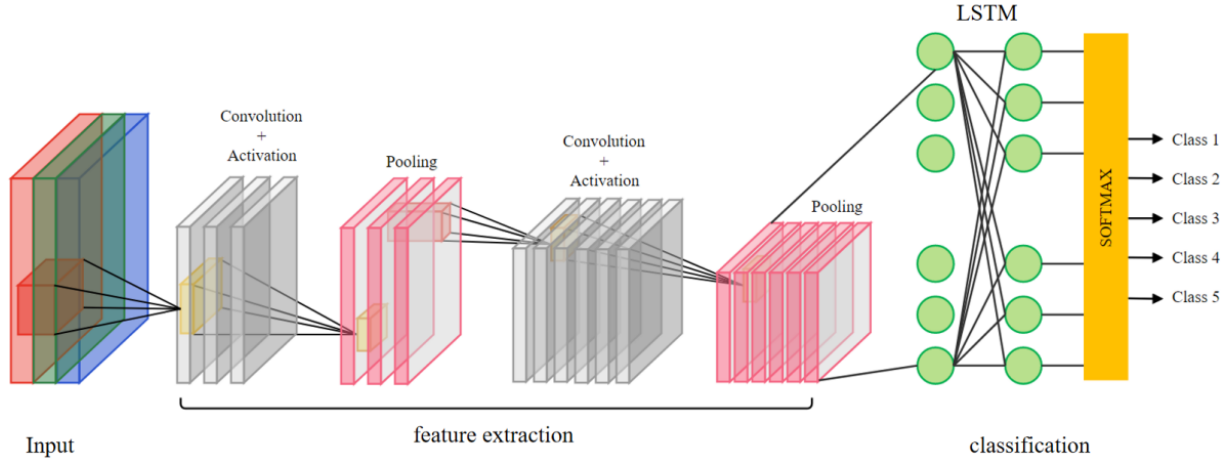
## 3. Emotion Recognition Model of Takeaway Evaluation Text Based on LSTM-CNN

The existing take-away data on the Internet are uneven, with an average length of 20-50. Most take-away comments are short in text length, but there are still a few take-away comments with a long text length ranging from 100 to several

hundred words. In addition to the dishes will affect the emotional tendency of user comments, whether the user's order will affect the emotional tendency. Therefore, we can continue to use the captured information, that is, the different delivery time and ordering time of the user's order, to study whether it will affect the user's emotional tendency. Through the comparative analysis of the user's order emotion, we can provide reference for merchants and take-away enterprises to

make decisions in many ways.

This project intends to study an emotion recognition method of take-away comment text based on LSTM-CNN, in order to solve the problems of poor real-time performance, difficulty in applying to large-scale text, and inability to extract text context and local semantic features at the same time. This structure is shown in Figure 1.



**Figure 1.** Emotion recognition model of takeaway evaluation text based on LSTM-CNN

A attention mechanism based on Bi-LSTM is used, and the vectorization of text vocabulary is realized by Word2Vec. On this basis, using the LSTM-CNN serial hybrid model, the semantic information of the text is extracted from the context and local semantic information respectively. Finally, the text is classified by Softmax classifier.

A large number of comments are freely written by different users, and there is no structured or standardized grammar and pattern, which is highly unstructured. The open source word vector tool Word2Vec can transform text words into high-dimensional real number vectors with certain semantic information by using CBOW(Continuous Bag Of Words) or Skip-gram [12-13]. In this paper, the Skip-gram model of Word2Vec is selected to vectorize the text vocabulary.

Bi-LSTM model based on attention mechanism is a baseline model in this chapter, which is developed by subsequent improvements. The input layer is responsible for converting the input sentence and given aspect into vector representation, and then the Bi LSTM layer constructs aspect related memory matrices from the word vector sequence of the sentence. The attention layer then captures the information that needs to be focused on the given aspect from the memory matrix output by the Bi LSTM layer. The classification layer consists of a multi-layer perceptron and a softmax classifier.

Attention calculates  $p = \{p_1, p_2, \dots, p_n\}$  for each hidden layer vector, and the attention weight can be interpreted as the probability that the correct word should be paid attention to when judging the emotional polarity. The calculation process of attention can be expressed as the following formula:

$$d_i = \tanh(h_i^T, w, t_s) \quad (1)$$

$$p_i = \frac{\exp(d_j)}{\sum_{j=1}^n \exp(d_j)} \quad (2)$$

Where  $t_s$  is that vector representation of the target word and  $w$  is the train parameter matrix.

Weighting the output of hidden layer with attention weight:

$$z_s = \sum_{i=1}^n p_i h_i \quad (3)$$

Finally, the output of the attention layer passes through the full connection layer to get the final emotion classification result.

In the task of text classification of takeaway evaluation, because takeaway evaluation text has the characteristics of short length, compact structure and clear meaning, CNN can be used to classify emotions. Through the calculation of different kernel functions in convolution layer, multiple feature maps can be obtained, and then different features in the text can be mined.

The basic unit of the convolution layer is N-gram, and if the window size is  $d$ , the word vectors of consecutive  $d$  words can form an N-gram vector  $c_i$ , as shown in Formula (4).

$$c_i = v_i \oplus v_{i+1} \oplus \dots \oplus v_{i+d-1} \quad (4)$$

Where  $\oplus$  is the join operation, which is used to join the word vectors from head to tail.

The purpose of convolution layer is to extract the semantic features of sentences, and each convolution kernel should extract some features. In this paper, the number of convolution kernels is set to 128. For each sentence matrix  $z$

output by embedding layer, convolution operation is performed:

$$S = f(wz + b) \quad (5)$$

Where:  $S$  represents the feature matrix extracted by convolution operation; Weight matrix  $w$  and bias vector  $b$  are the learning parameters of the network.

Then, the extracted features are transferred to the library, and all the features are merged in the library, thus simplifying the representation of features. On this basis, a clustering algorithm based on K-Max function is proposed, and the semantic information represented by each filter is expressed by selecting the first  $k$  maxima of each filter. In the convolution layer, the phrase is convolved, pooled and feature extracted to obtain generalized binary and ternary feature vectors. On this basis, the two types of feature vectors are fused to form a least squares model.

Finally, input to the full connection layer for emotional analysis:

$$p(y|w^T, w) = \text{soft max}(w_f q + b_f) \quad (6)$$

Where  $w_f, b_f$  is a parameter that can be learned.

## 4. Experimental Analysis

The experimental take-away data set comes from the network, and it is the take-away review data of a take-away platform. From it, 7 000 pieces are randomly selected as training set, 1500 pieces as verification set and 1 500 pieces as test set, and the training set, test set and verification set are not intersected. In order to avoid over-matching, a leakage layer is added to the output layer of LSTM, and the leakage probability is 0.5. On this basis, L1 regularization and L2 regularization are adopted, and the parameter is set to 1. The initial value of the learning rate is set at 0.001 and 0.9. In the learning stage, an "early termination" mechanism is introduced, that is, once the performance of the test sample can not be further improved, the learning is terminated.

This experiment is written in Python, and the data of the input model is represented by a 400-dimensional word vector trained by Word2Vec. The threshold for judging a long text is set to 100, and the number of keyword extraction is 20. The accuracy and loss changes of the training set during the experiment are shown in Figure 2.

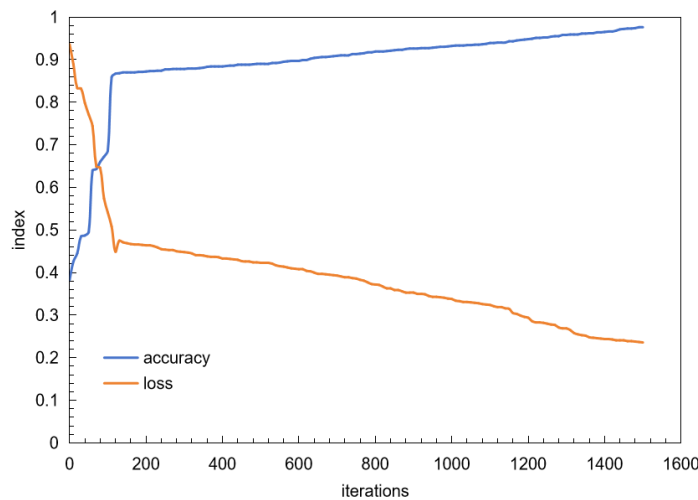


Figure 2. Accuracy and loss change of training set

In order to verify the validity of this model, precision, recall rate and F1 value are used for evaluation, and the

experimental results are compared with the evaluation results of different models, as shown in Table 1.

Table 1. Performance comparison between the proposed model and the hybrid model

model	Precision	Recall	F1
LSTM	0.774	0.78	0.714
CNN	0.789	0.769	0.725
Bi-LSTM	0.812	0.797	0.81
LSTM+ attention mechanism	0.835	0.805	0.826
The proposed model	0.937	0.896	0.906

It can be seen that the accuracy of Bi-LSTM model is higher than that of simple LSTM and CNN model, while the accuracy of LSTM+ attention mechanism model is higher than that of Bi-LSTM model. The accuracy of this model is the highest, with precision, recall and F1 being 0.937, 0.896

and 0.906 respectively. Through horizontal comparison, it can be found that the model in this paper is outstanding in this task, and the emotion enhancement model also improves the classification accuracy to some extent.

## 5. Conclusions

Massive texts come from many users of the Internet, with various forms and no fixed format, so it is difficult to process them by simple automatic means. If you rely on manual processing, there are problems such as excessive workload and poor real-time performance. In this paper, the word embedding NLP is used for preprocessing in tensorflow environment, and an emotion recognition model of takeaway evaluation text based on LSTM-CNN is established. The experimental results show that the accuracy of this model is the highest, and the precision, recall and F1 are 0.937, 0.896 and 0.906 respectively. Through horizontal comparison, it can be found that the model in this paper is outstanding in this task, and the emotion enhancement model also improves the classification accuracy to some extent.

## References

- [1] Bernhard, K. , Suzana, I. , Mathias, K. , Stefan, F. , & Helmut, P. (2018). Deep learning for affective computing: text-based emotion recognition in decision support. *Decision support systems*, 115(10), 24-35.
- [2] Zhang, G. , Ananiadou, S. , & Odbal, S. (2022). Examining and mitigating gender bias in text emotion detection task. *Neurocomputing*(7), 493.
- [3] Ian, W. , John, M. C. , Vladimir, A. , & Paul, B. (2018). A comparison of emotion annotation approaches for text. *Information*, 9(5), 117.
- [4] Xu, Q. , Zhang, C. , & Sun, B. (2020). Emotion recognition model based on the dempster–shafer evidence theory. *Journal of Electronic Imaging*, 29(2), 1.
- [5] Sourina, O. , Li, L. , & Pan, Z. (2012). Emotion-based interaction. *Journal on Multimodal User Interfaces*, 5(1-2), p.1.
- [6] Lee, S. Y. M. , Chen, Y. , Huang, C. R. , & Li, S. (2013). Detecting emotion causes with a linguistic rule-based approach. *Computational Intelligence*, 29(3), 390-416.
- [7] Xing, W. U. , Hai-Tao, L. , & Shao-Jian, Z. (2015). Sentiment analysis for chinese text based on emotion degree lexicon and cognitive theories. *journal of shanghai jiaotong university*, 20(1), 6.
- [8] Bandhakavi, A. , Wiratunga, N. , Massie, S. , & Padmanabhan, D. (2017). Lexicon generation for emotion detection from text. *Intelligent Systems, IEEE*, 32(1), 102-108.
- [9] Yamamoto, Y. , Niitsuma, M. , & Yamashita, Y. (2016). Automatic recognition of negative emotion in speech using support vector machine. *The Journal of the Acoustical Society of America*, 140(4), 3400-3400.
- [10] Pais, A. L. , Moga, A. S. , & Buiu, C. (2010). An integrated framework for emotion recognition and expression using robot artists. *ICIC Express Letters*, 1(2), 169-174.
- [11] Cai, L. , Hu, Y. , Dong, J. , & Zhou, S. (2019). Audio-textual emotion recognition based on improved neural networks. *Mathematical Problems in Engineering*, 2019(6), 1-9.
- [12] Huang, H. , Liu, Q. , & Wu, L. (2014). Improving emotion recognition from imbalanced corpus with a novel re-sampling method. *ICIC Express Letters*, 8(7), 1845-1850.
- [13] Xie, B. , Sidulova, M. , & Park, C. H. (2021). Robust multimodal emotion recognition from conversation with transformer-based crossmodality fusion. *Sensors*, 21(14), 4913.