

The Multimedia Sentiment Model Based on Online Homestay Reviews

Wenguang Song^{a, b}

^a School of Computer Science, Yangtze University, Jingzhou, Hubei 430100, China

^b School of Computer Science, University of Regina, Regina, Saskatchewan S4S 0A2, Canada

Hanyu Li*

School of Computer Science, Yangtze University, Jingzhou, Hubei 430100, China

E-mail: 201871340@yangtzeu.edu.cn

Qian Yu

School of Computer Science, University of Regina, Regina, Saskatchewan S4S 0A2, Canada

Wan Li, Bingxin Zhang, Qiujuan Zhang

School of Computer Science, Yangtze University, Jingzhou, Hubei 430100, China

Zhigang Liu

Control Technology Institute, Wuxi Institute of Technology, Wuxi, Jiangsu 214121, China

Received: 17 July 2020; Accepted: 26 July 2020; Published: 08 August 2020

Abstract: Aiming at the fact that traditional sentiment analysis is based on text, without considering the factors such as special symbols and emoticon images, which can't fully extract the user's emotions, this paper proposes a sentiment analysis method of online homestay reviews based on image-text fusion. For text datasets, first use Word2vec to build a topic clustering model, then find the corresponding topic attribute dictionary through the topic center words, use Bayesian classifier is used for sentiment analysis, compared with SVM and decision tree methods, to evaluate the effect; For the picture dataset, Convolutional Neural Network (CNN) model is initialized by parameter migration, and image sentiment classification model is obtained by fine-tuning training of CNN model after parameter migration; Finally, the fusion method is designed to calculate the emotional probability of image-text, then judge the emotional polarity and compare it with the user's score. The accuracy rate is 88.6%, which is higher than that of text sentiment analysis model or image sentiment analysis model. The experimental results show that the sentiment analysis of image-text fusion has better classification effect on image-text reviews and more effectively avoid the problem of inconsistency between user ratings and the emotion expressed in comments.

Index Terms: homestay; online review; image-text fusion; user score; CNN.

1. Introduction

Under the background of sharing economy, homestay has emerged in this era of artificial intelligence with a kind of 'freedom, exploration, individuality. The natural geographical advantages, thematic design and personalized service have more satisfied the consumers' accommodation demand [1]. Through the analysis of user behavior to better serve businesses and consumers. On the one hand, users' feedback is an important factor influencing consumers' decisions. On the other hand, merchants can also obtain valuable information from these feedbacks to improve deficiencies and improve service quality. The feedback information includes user scores and comments. This paper starts from the phenomenon that the user's score is inconsistent with the opinion expressed by the comment. Considering the diversity of the content of the review, it not only contains text, but also pictures, videos, audio, Web links etc., but mostly takes image-text as the main content. There are many traditional text-based sentiment analysis methods, and some people also have conducted sentiment analysis on images [2], however, users' sentiment polarity can't be mined only by text or images. Some studies have proposed to combine the features of images with the features of texts for sentiment analysis, which greatly improves the accuracy of sentiment classification.

Under the above background, this paper proposes a multimedia sentiment analysis method for online homestay reviews. It is expected that the method of model fusion can more accurately judge the emotional tendency and identify the inconsistency between user's score and emotional expression compared with user's score.

2. Materials and Methods

Sentiment analysis, also known as emotional classification or opinion mining, conducts research on a certain thing, analyzes the implicit emotional state, and judges the user's attitude, viewpoints and opinions [3]. The research object of this paper is online homestay reviews, which mainly includes user score, text and images. Among them, the user's scores and comments are inconsistent, and the emotional expressions between the images and texts are inconsistent, which have a great influence on the overall ranking of the homestay. The specific scores and comments are inconsistent as shown in Table 1. In this paper, emotional analysis is carried out on text review and image review respectively, and two classification models are obtained. Then, a method is designed to fuse the two models to get the fused model. Finally, three models are used to predict the emotion probability value. The accuracy of the three models is analyzed by comparing the emotional probability value with the user's score. The framework, as shown in Fig.1, consists of three modules: image sentiment analysis model, text sentiment analysis model and sentiment classification.

Table 1. User comments and ratings are inconsistent

User Score	Very good(On behalf of the full marks)
Review	The landlord is very nice, the surrounding area is very busy, the transportation is convenient, the living is very comfortable, very satisfied with an accommodation experience.
Addition	Before checking in, the landlord asked for praise, but the room was very inconsistent, the service attitude was not good, the air conditioner was bad, the key was that the price was not very cheap. When I went there, I found it for a long time, in short, it was very bad.

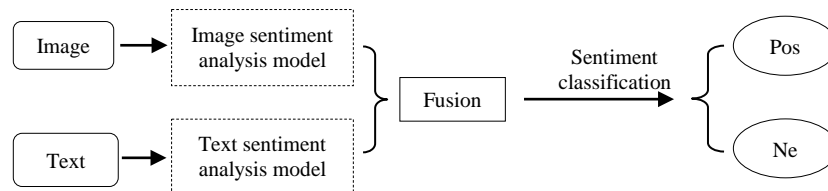


Fig.1. Framework for Emotional Classification of Graphic and Textual Comments.

2.1. Text Sentiment Analysis

At present, there are two main methods for text sentiment analysis: one is based on sentiment dictionary, the other is based on machine learning. The sentiment analysis method based on the sentiment dictionary is to assign weights to words according to the emotional intensity and emotional tendency (the weight range is [-7, 7]), negative numbers represent negative words, non-negative numbers represent positive words, positive and negative degrees are reflected in the numerical value, and then calculate the emotional score. This method is not difficult, but first of all, the coverage of sentiment dictionary is not complete, which will lead to low accuracy and poor robustness [4], although the dictionary can be expanded by high-frequency words in statistical corpus, this will lead to difficulty in updating; Secondly, The same emotional word is expressed differently in different scenes[5].

The sentiment analysis methods based on machine learning are divided into supervised and unsupervised. Supervised methods need to label samples manually, then extract features, input them into the constructed classification model for training, and finally predict on the test set. Unsupervised methods does not require manual labeling of samples, but requires prior knowledge to improve the classification effect [6]. This paper uses the supervised labeling samples in machine learning to conduct sentiment classification experiments. The framework is shown in Fig.2 below.

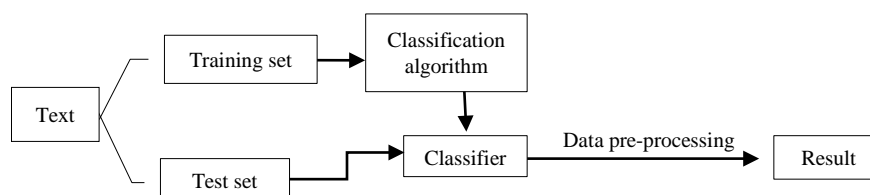


Fig.2. Text Emotion Analysis Model Framework.

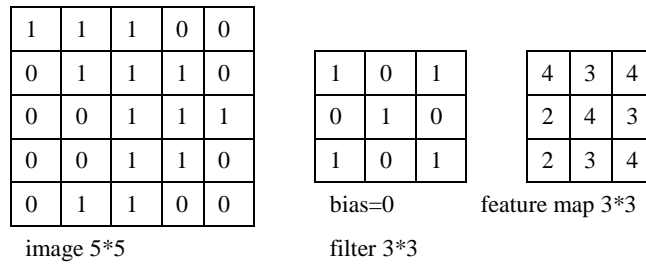
2.2. Image Sentiment Analysis

A picture contains a lot of information, which can reflect the user's emotional tendencies. Reference [7] proposes a Cross-modality Consistent Regression (CCR), which trains paragraph vector model and multi-modal regression model of text sentiment analysis on the basis of fine-tuning convolution neural network in order to achieve consistency among different modal features; Reference [8] proposed an emotional analysis method based on Simple Multiple Kernel Learning (SimpleMKL). The main idea is to use text information to improve the ability of image classification through SimpleMKL. Once an image classifier is obtained, it can directly predict other unmarked images without text. The idea of reference [9] is to construct CNN model by combining image features and text features. The research shows that the sentiment analysis combined with text and image has better classification effect than text-based sentiment analysis.

In view of the above analysis, this paper establishes an multimedia sentiment analysis model based on CNN, as shown in Fig.3. The composition and function of the model are as follows:

① 3 convolutional layers: the local features of the image are obtained by convolution to reduce the amount of parameters.

Suppose there is a 5 * 5 image, and a 3 * 3 convolution kernel is used for convolution to obtain a 3 * 3 matrix feature map, as shown below:



$x_{i,j}$ is used to represent row i and column j of the image; Each weight of filter is numbered, $w_{m,n}$ is used to represent the weight of row m and column n and w_b is used to represent the bias of filter; Each element of the Feature Map is numbered, $a_{i,j}$ is used to represent the element of row i and column j of the feature map, f is used to represent the activation function, and the relu activation function is used in the experiment. Then, formula (1) is used to calculate the convolution:

$$a_{i,j} = f\left(\sum_{m=0}^2 \sum_{n=0}^2 w_{m,n} x_{i+m,j+n} + w_b\right) \tag{1}$$

Therefore, for the upper-left element $a_{0,0}$ of the feature map, the convolution calculation method is as follows:

$$\begin{aligned} a_{0,0} &= f\left(\sum_{m=0}^2 \sum_{n=0}^2 w_{m,n} x_{m+0,n+0} + w_b\right) \\ &= \text{relu}(w_{0,0}x_{0,0} + w_{0,1}x_{0,1} + w_{0,2}x_{0,2} + w_{1,0}x_{1,0} + w_{1,1}x_{1,1} + w_{1,2}x_{1,2} + w_{2,0}x_{2,0} + w_{2,1}x_{2,1} + w_{2,2}x_{2,2}) \\ &= \text{relu}(1+0+1+0+1+0+0+0+0+1+0) \\ &= 4 \end{aligned}$$

The above method can be used to calculate the values of all elements in the feature map in turn. In the above calculation process, the stride is 1, and when the stride is 2, the feature map becomes 2 * 2. Then the values of all elements in the Feature Map can be calculated in turn. In the above calculation process, the stride is 1, and when the stride is 2, the Feature Map becomes 2 * 2.

$$W_2 = (W_1 - F + 2P) / S + 1 \tag{2}$$

$$H_2 = (H_1 - F + 2P) / S + 1 \tag{3}$$

W_2 is the width of the feature map after convolution; W_1 is the width of the image before the convolution; F is the width of the filter; P is the amount of Zero Padding (Zero Padding is the number of zeroes around the original image). S is the stride; H_2 is the height of the feature map after convolution; H_1 is the width of the image before convolution.

The above is the calculation method of the convolution layer with a depth of 1. If the image depth before convolution is D , then the depth of the corresponding filter must also be D . Then the convolution calculation formula is:

$$a_{i,j} = f\left(\sum_{d=0}^{D-1} \sum_{m=0}^{F-1} \sum_{n=0}^{F-1} w_{d,m,n} x_{d,i+m,j+n} + w_b\right) \quad (4)$$

In formula (4), D is the depth; F is the size of the filter; $w_{d,m,n}$ represents the weight of the d -th layer of the filter, the m -th row, and the n -th column; $x_{d,i,j}$ represents the pixels of the d -layer, i -th, and j -th column of the image and the other symbols have the same meaning as formula (1). Each convolutional layer can have multiple filters, after each filter and the original image are convolved, a feature map can be obtained. Therefore, the depth of the Feature Map after convolution is the same as the number of filters in the convolution layer.

② 2 pooling layers: reduce the latitude and improve the efficiency of model training. The principle is to take the maximum value of the covered area as the main feature output.

The pooling layer is to compress the feature map. There are two ways, one is *Max Pooling*, and the other is *Mean Pooling*. *Max Pooling* is used in this article. If the output of the pool layer is a^l , where m and n are the regions covered by the corresponding pool kernel in formula (5):

$$a_{i,j}^l = \max(a_{m,n}^{l-1}), i \leq m, n \leq i + 2 \quad (5)$$

③ 2 Full Connection Layers: the Flatten operation is required to Flatten the data into vectors to ensure that the input must be one-dimensional. The fully connected layer plays the role of "classifier" in the whole convolutional neural network, which is essentially a linear transformation from one feature space to another, and the core operation is matrix vector product, as shown in formula (6), where W is weight parameter, b is bias parameter:

$$y = Wx + b \quad (6)$$

④ 4 activation function layers: they introduce the nonlinear characteristics into the neural network. The activation functions used in this paper are *relu* and *softmax*.

The *relu* function is to set all outputs less than 0 as 0, and set outputs greater than 0 equal to the input. The advantage of this function is that it converges quickly.

$$f(x) = \max(0, x) \quad (7)$$

Softmax function maps a k -dimensional real value vector ($a_1, a_2, a_3, a_4, \dots$) into ($b_1, b_2, b_3, b_4, \dots$), where b_i is a constant from 0 to 1, and then multi-classification tasks can be performed according to the size of b_i . The formula for *Softmax* is as follows:

$$P(i) = \frac{e^{V_i}}{\sum_j e^{V_j}} \quad (8)$$

V is an array, V_i represents the i -th element in V , and $P(i)$ is the *Softmax* value of the i -th element in V . Because the calculation amount is small when calculating the *Softmax* loss, it is applied in the last layer of the neural network in this paper.

⑤ 1 Flatten layer: the picture is "flattened", enter full connection layer after the multi-dimensional input is transformed into one-dimensional input.

⑥ 2 Dropout layers: randomly delete neurons according to probability each time to prevent overfitting.

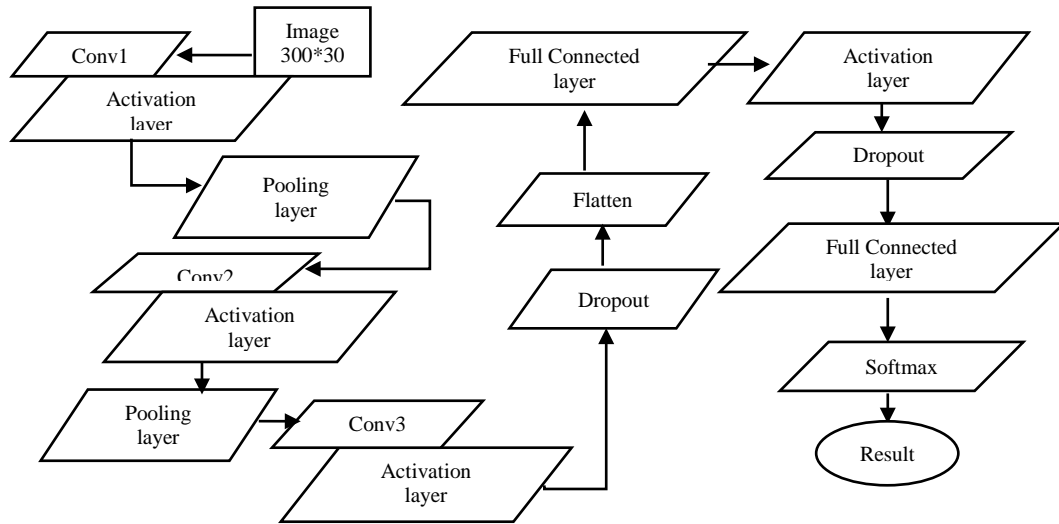


Fig.3. Framework of Picture Emotion Analysis Model.

2.3. Fusion Method Design

The text sentiment classification model is integrated with the picture sentiment classification model to form the sentiment analysis of the fusion of images and texts. At present, the commonly used model fusion methods are decision layer fusion, feature layer fusion and data layer fusion. Decision fusion is used in this paper. According to certain criteria, the probability of each target is weighted and the probability after decision is calculate [10], the specific formula is as follows:

$$P = \lambda P_t + (1 - \lambda) P_i \quad (9)$$

In the formula: P_t is the text sentiment probability output by the text sentiment analysis model, P_i is the picture sentiment probability output by the picture sentiment analysis model, λ is the weight, which represents the importance of the text emotion to the final analysis result, It is usually equal to 0.5, text sentiment and image sentiment are considered to have the same influence on the final result. Then their emotional probabilities are multiplied by their weights, and the products are added to obtain the emotional probability values of the model fusion, if the emotional probabilities are greater than 0.5, they are positive emotions, otherwise they are negative emotions.

3. Results and Discussion

In order to verify the validity of the sentiment analysis method of online homestay reviews based on image-text fusion, this paper collects online reviews for experiment. The following will be explained from three aspects: image-text data set, the description of the experimental process and the analysis of the results.

3.1. Data Source

Data acquisition is to use Selenium to simulate browser click-and-flip operation, with the python crawler implemented by Request. The target of the acquisition is the homestay reviews of Chongqing on Ctrip. The data is 11903 reviews after cleaning. Five graduate students were asked to label the data by artificial voting. If more than three of them considered it was positive emotion, it was masked as positive, otherwise it was negative. Finally, 9428 positive and 4475 negative image-text reviews were obtained.

Since each comment can be matched with more than one picture, in order to achieve overall balance of text-image reviews, 9428 pictures are randomly selected in the positive reviews and 4475 pictures are randomly selected in the negative reviews.

3.2. Evaluation Index

In order to compare the classification effect of the classifier, the accuracy rate (Precision, abbreviated as P) and recall rate (Recall, abbreviated as R) and $F1$ values are used as the evaluation indicators of the experiment [11].

True Positive (TP): The number of reviews that are positively predicted as positive;

False Positive (FP): The number of reviews that are negatively predicted as positive;

True Negative (TN): The number of reviews that are negatively predicted as negative;

False Negative (FN): The number of reviews that are positively predicted as negative.

$$\text{Accuracy rate: } P = \frac{TP + TN}{TP + FP + TN + FN} \quad (10)$$

$$\text{Recall rate: } R = \frac{TP}{TP + FN} \quad (11)$$

$$\text{F1 value: } F1 = \frac{2P * R}{P + R} \quad (12)$$

3.3. Text Sentiment Classification

3.3.1. Text Preprocessing

This paper uses machine learning method to classify text emotion, which mainly includes the following steps:

(1) Word segmentation: splitting sentences into collections of words. First, jieba is used for word segmentation. Jieba supports three word segmentation modes(Precise mode, full mode, search engine mode)[12], traditional Chinese word segmentation, custom dictionary and other features, which greatly meet the needs of Chinese word segmentation; Then, part-of-speech tagging is carried out. Finally, words with the part of speech of `n'(noun) are selected.

(2) Removal of stop words: delete the symbols in the comments such as "#, ¥, @, &" that have no effect on the results of sentiment analysis.

(3) Feature extraction: text sentiment analysis is to analyze the features extracted from the text, so how to extract the features is very critical [13].

Firstly, the word frequency of selected nouns is counted to filter out irrelevant nouns, and the feature words are selected according to the ranking rate from high to low to obtain the feature words collection of comments{ transportation, location, experience, service, features, environment, facilities, food, price, hygienic ... }.

Then, because the input of the model is tuple, the word combination of each data is converted into a numerical vector, common conversion algorithms are: Bag of Words, TF-IDF and Word2Vec. The Word2Vec method is chosen in this paper[14]. There are two models in Word2Vec, one is CBOW, and the input words are inferred from the context; the other is Skip-gram, and the context is predicted by the input words, the opposite of which is the opposite. In the experiment, Word2Vec is used to train the skip-gram model to convert words into vectors.

Finally, the feature dictionary is built on the basis of features by summarizing the high-frequency words, as shown in table 2 below. The corresponding feature attribute dictionary can be found by using the central words, and then the recognition and analysis of the emotional tendency are carried out [15].

Table 2. Feature attribute dictionary

Theme features	Attribute dictionary
Environment	environment, scenery, air, river, view, community, scenic spot, night view, street, surrounding, scenic, spot, sound
Price	price, cost, unit price, expensive, cheap, affordable, worth
Features	features, layout, architecture, structure, style, decoration, design, style, sound insulation
Facilities	facilities, equipment, condition, hardware, room, hot water, toilet, elevator, balcony, air conditioner, quilt, bed, hall, telephone, electricity, decoration
Food	dining, breakfast, coffee, taste, rice, vegetables, fruits, specialties, meals, delicacies, barbecues, snacks, ingredients, restaurants, snacks
Transportation	transportation, drive, distance, parking, airport, station, location, geography, center, elevation, wharf
Service	service, attitude, front desk, waiter, boss, staff
Experience	experience,feeling

(4) Text sentiment analysis, through the experiment of Bayesian, SVM and decision tree, the best classifiers is selected for text sentiment analysis.

3.3.2. Classifier Effect Comparison

Record the accuracy and recall rate of each classifier for each feature, as shown in Table 3 below, the average accuracy of the three classifiers is 82.5%, 79.8% and 80.6%, respectively. The Bayesian classifier has higher accuracy and is more suitable for the text data set in this paper.

Table 3. Classification effect of different classifiers

Features Classifier	Bayesian		SVM		Decision Tree	
	P/%	R/%	P/%	R/%	P/%	R/%
Environment	84.2	79.7	83.5	79.4	79.4	78.5
Price	87.0	84.3	85.2	80.0	83.7	85.8
Features	77.5	79.6	73.9	74.5	75.3	74.6
Facilities	81.6	85.2	76.5	78.3	75.8	79.2
Food	75.0	75.8	81.6	77.8	81.4	85.5
Transportation	80.2	79.3	78.7	80.5	87.4	80.5
Service	88.7	85.5	84.6	81.9	84.8	85.6
Experience	86.4	82.1	74.6	75.1	77.3	71.1
Average	82.5	81.4	79.8	78.4	80.6	80.1

The sentiment analysis of the feature-extracted text comments is performed on the feature topic attribute table, and the emotional trends corresponding to each feature are respectively obtained. The abscissa is the feature's sentiment score, the ordinate is the number of corresponding comments, the farther to the right, the better the current feature evaluation. Here are the theme features{“Transportation”,“ Price”,“ Experience”,“ Service”,“ Features”,“ Environment”, “ Facilities”, “Food”},Select the theme features of " Transportation " for visualization, and the results are shown in Fig.4 below.

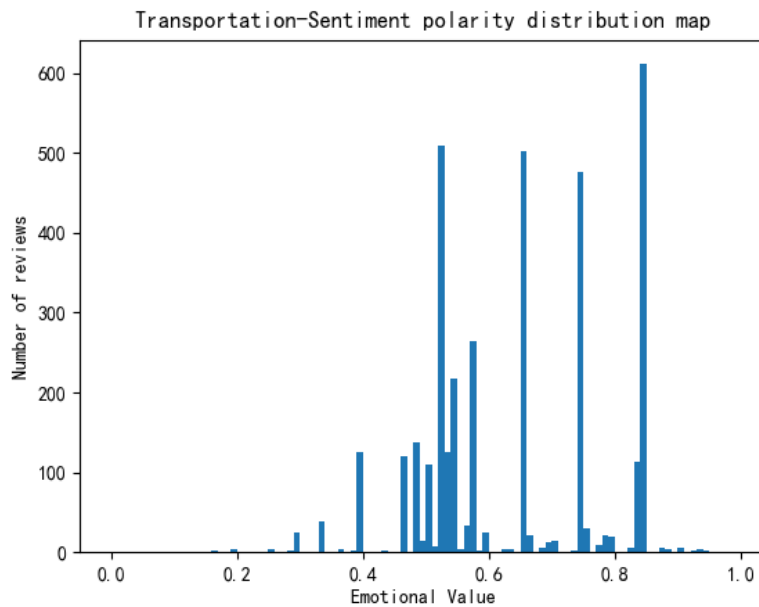


Fig.4. Emotional trend distribution map corresponding to the feature

3.4. Image Sentiment Classification

3.4.1. Image Preprocessing

The study in document [8] used two data sets, one from Flickr (including pictures and text descriptions) and the other from Twitter. Two sets of experiments were set up. One group was trained on Flickr data set and tested on Twitter data set. The other group was trained with 80% Flickr data set, and the rest were used for testing. In this paper, 80% of the data collected is used as training set, and the remaining 20% as test set. One fourth of the data is reserved as validation set in training set.

Table 4. Quantitative table of H value

Range of H	Value of H	Colour
$25 \leq H < 41$	0	red
$41 \leq H < 75$	1	orange
$75 \leq H < 156$	2	yellow
$156 \leq H < 201$	3	green
$201 \leq H < 272$	4	blue
$272 \leq H < 285$	5	indigo
$285 \leq H < 330$	6	purple
$330 \leq H < 360$ Or $0 \leq H < 25$	7	pink

The key to feature extraction in the image sentiment analysis model is the extraction of color features. First, the RGB parameters of each pixel in the image are extracted, and then the RGB parameters are converted into hue, saturation and intensity, i.e. HSV [16]. When H is in different ranges, it has different values, as shown in table 4 below.

After quantification, S has two values: when $S \leq 0.1$, the image is close to gray, and H can be ignored, that is, the influence of hue; When $0.1 < S \leq 0.65$, $S = 0$, representing a dark color; When $0.65 < S \leq 1$, $S = 0$, representing a clear color. After quantification, V has a value: when $V < 0.15$, the color of the image is close to black, and the influence of H can be ignored; when $0.15 < S \leq 1$, $V = 0$. Finally, the HSV values are summed according to formula (13). Q_s and Q_v are the quantification coefficients of S and V respectively, $Q_s = 2$, $Q_v = 1$. After summation, the color feature can be represented by a value, and the range is [0, 15].

$$L = H * Q_s * Q_v + S * Q_v + V \quad (13)$$

3.4.2. Parameter Settings

The image emotion classification model based on CNN consists of three convolution layers, each of which has 32 convolution cores with a size of 3×3 . The activation function is *relu*: In the first layer, the *border_mode* is set to valid (If *border_mode* is selected as same, the input and output sizes will remain the same after convolution. If valid is selected, the size will be smaller after the convolution.); 2 pooling layers: Downsampling data with *max-pooling*, and *poolsize* is (2, 2); After three layers of convolution and two layers of pooling, each image is processed to 4×4 size; 2 dropout layers: Set the probability to set the value of some elements to zero randomly, and the value of the remaining elements will be enlarged to prevent over-fitting, here set the probability is 0.5; 1 flatten layer: the two-dimensional feature map output by the previous layer is flattened into one-dimensional after being processed by the flatten layer; 2 fully connected layers: 512 neurons are designated in the first fully connected layer; 1 softmax layer: the *softmax* activation function is used; The *batch* is set to 4 and the number of iterations is 20.

3.5. Fusion Model Performance Analysis

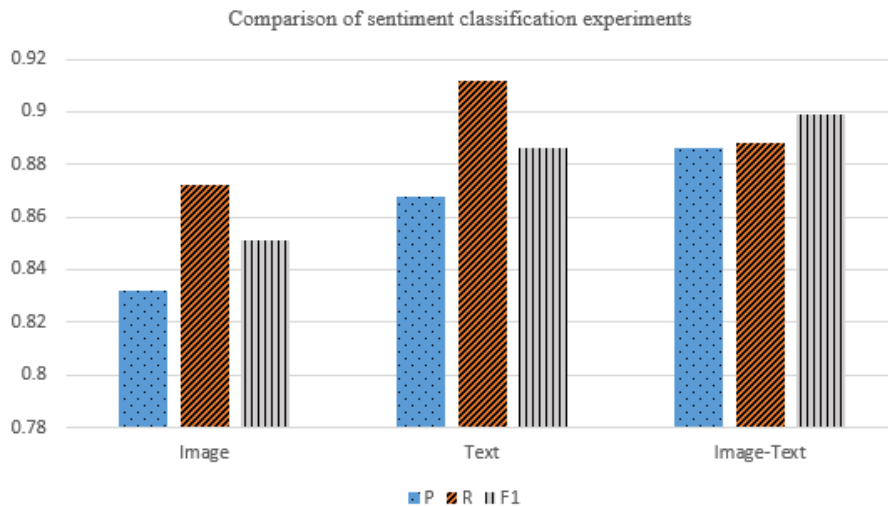


Fig.5. Comparison of sentiment classification experiments

The sentiment analysis model of online homestay reviews based on image-text fusion adopts the method of decision-level fusion, and the emotional probability of text and picture is multiplied by their respective weights, and then added to get the final emotion probability after image-text fusion. Judging whether the final probability of emotion is greater than 0.5, if it is greater than 0.5, it is considered to be positive emotion, otherwise it is negative emotion. The experimental results of emotion classification are shown in Fig.5 below.

From the comparison results in Fig.5, the sentiment classification model of image-text fusion has higher accuracy, which indicates that the classification effect of online homestay reviews analysis model based on image-text fusion is better than that of single text sentiment classification model or image sentiment

classification model. It is proved that the emotion analysis method based on image and text fusion is effective for the analysis of homestay reviews, and the existence of pictures has a great impact on the emotional polarity. However, from the perspective of recall rate, the recall rate of the text emotion classification model is relatively large. This is because the classification method with supervised mark samples by machine learning can fully extract the feature topics of text comments, so it can well identify the comments with positive emotions. The feature extraction of image emotion classification model mainly lies in the extraction of color features, which can't fully extract image features, resulting in low recall rate of image emotion classification model and fusion model.

As shown in Fig.6 below, using the learned model to test the comments of a certain homestay and select the theme of "facilities" for analysis, it can be seen that the user's score is "very satisfied", while the average emotional probability value of the output is less than 0.5, which is seriously inconsistent with the actual situation, indicating that the user rating is not enough to be a real and effective reference. In addition, businesses can also know which aspects need to be strengthened through the statistics of topics with low emotional probability value.

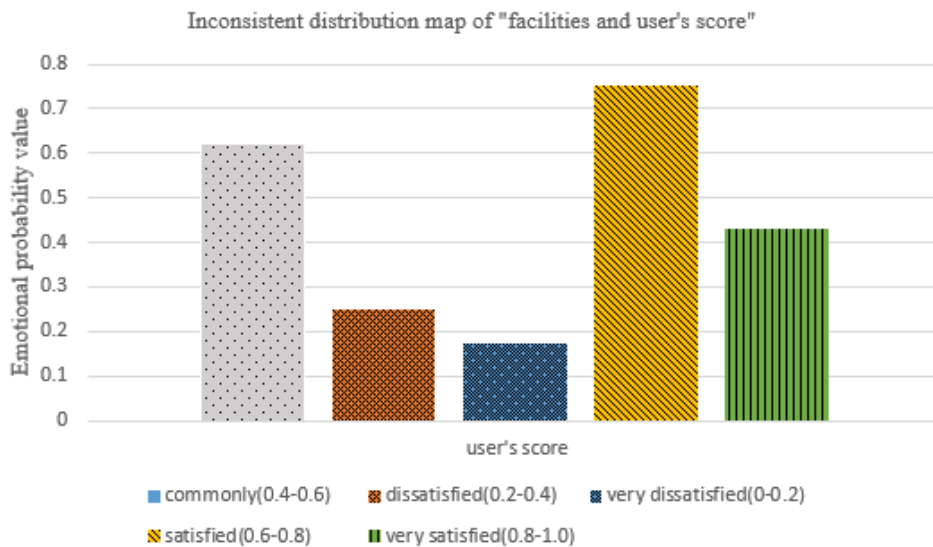


Fig.6 Inconsistent distribution of " user's score and emotional polarity"

4. Conclusions

Based on the phenomenon that users' rating is inconsistent with their comments, this paper proposes to take the homestay reviews in Chongqing as the research object. Due to the diversity of comments, image features and text features are fused to train the sentiment analysis model of image-text fusion. Text sentiment classification based on machine learning, In the experiment, Bayesian, SVM and decision tree methods were used to compare the average accuracy of them; On the basis of CNN, image sentiment classification model is trained; Finally, the decision fusion method is adopted to synthesize the emotion probability of text and image to get the emotion probability of image-text, judge the polarity of emotion, and compare with the user's score can prove that the comment and score are inconsistent. Experimental results show that the image-text fusion method has better classification effect. The problem of inconsistent comments and scores can be more effectively avoided. At the same time, according to the emotional probability to give a reasonable score and ranking, can ensure that other users get more accurate recommendation and reference.

Studies have shown that the criteria for customers to choose accommodation in different areas are different [17]. This paper only collected the reviews on homestay in Chongqing, resulting in a single data set, and the analysis results are not applicable to other regions. The next study will collect comments on homestay in multiple regions for experiments, so as to have better classification effect and generalization ability.

Acknowledgments

The author would like to thank the support received from Xinjiang Uygur Autonomous Region Innovation Environment (talent, base) Construction Foundation (Xinjiang NSFC Program Foundation 2020D01A132): Research and implementation of horizontal well inversion optimization interpretation method. Jingzhou Science Technology Foundation (2019ec61-06) and Hubei Science, Technology Demonstration Foundation (2019ZYYD016) and Qinglan Project Subsidized by Colleges and Universities in Jiangsu Province.

References

- [1] Zhang Sijia, and Qian Yipei. "Research on the Development Obstacles and Countermeasure of Hostel in the Background of Shared Economy. " *Market Modernization* ,2018(18):194-195.
- [2] Campos, Víctor, Brendan Jou, and Xavier Giró-I-Nieto. "From Pixels to Sentiment: Fine-Tuning Cnns for Visual Sentiment Prediction *." *Image & Vision Computing* (2017): S0262885617300355.
- [3] Bo, Pang, and Lillian Lee. "Opinion Mining and Sentiment Analysis." *Opinion Mining & Sentiment Analysis*.
- [4] Shi Wei, Wang Hongwei, and He Shaoyi. "Sentiment Analysis of Chinese Online Reviews Based on Semantics. *Opinion Mining & Sentiment Analysis*." *Journal of the China Society for Scientific and Technical Information*, 2013, 32(8): 860-867.
- [5] Arnold, Thomas M. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks." *Journal of Finance* 66, no. 1 (2011): 35-65.
- [6] Deng Pei, Tan Changgeng. "Multimedia sentiment analysis on micro-blog based on transition variable." *Application Research of Computers*,2018,35(07):124-127.
- [7] You, Quanzeng, Jiebo Luo, Hailin Jin, and Jianchao Yang. "Robust Image Sentiment Analysis Using Progressively Trained and Domain Transferred Deep Networks." (2015).
- [8] Tan J, Xu M, Lin S, Jia X, editors. "Sentiment Analysis for Images on Microblogging by Integrating Textual Information with Multiple Kernel Learning." *Pacific Rim International Conference on Trends in Artificial Intelligence*; 2016.
- [9] Cai Guoyong, and Xia Binbin. "Multimedia Sentiment Analysis Based on Convolutional Neural Network." *Journal of Computer Applications*. 2016;36(2):428-431.
- [10] Miao Yuqing, Wang Junhong, Liu Tonglai, et al. "Joint Visual-textual Approach for Microblog Sentiment Analysis".*Computer Engineering and Design*,2019,40(04):1099-1105.
- [11] Jiao Feng. "Analysis of the emotional tendency of hotel review based on naive bayes." *Modern Computer*,2018(20):45-49.
- [12] Zhang Decheng, Wang Yang, Zhao Chuanxin, Zhen Lei, and Li Chang. "Minimal text classification model based on bayesian decision." *Journal of Chongqing University of Science and Technology*,2018,20(04):82-85.
- [13] Tian Zhu. "Research on sentiment analysis based on deep feature extraction". Shandong University,2017.
- [14] Rehurek R, Sojka P. "Software framework for topic modelling with large corpora//Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks.", 2010.
- [15] Kanayama, Hiroshi, and TETSUYA NASUKAWA. "Unsupervised Lexicon Induction for Clause-Level Detection of Evaluations." *Natural Language Engineering* 18, no. 1 (2012): 83-107.
- [16] Vadivel A , Sural S, Majumdar A K . "Color-texture feature extraction using soft decision from the HSV color space// International Symposium on Intelligent Multimedia." IEEE, 2004.
- [17] Wu Weifang, Gao Baojun, Yang Haixia, et al. "Impact of review text on hotel satisfaction: method based on sentiment analysis." *Data Analysis and Knowledge Discovery*,2017,1(03):62-71.

Authors' Profiles



Wenguang Song was born in Wuhan, Hubei, P.R. China, in 1979. Now, he works in school of Computer Science, Yangtze University. His research interest includes software engineering, petroleum software technology and big data analysis. Visiting scholar of Regina University, Canada. E-mail: wenguang_song@yangtzeu.edu.cn



Corresponding Author: **Hanyu Li** was born in Zaoyang, Hubei, P.R. China, in 1995. Now she is studying in the school of Computer Science, Yangtze University. Her research interest includes data mining and natural language processing. E-mail: 201871340@yangtzeu.edu.cn



Qian Yu (1981.07-), Ph.D. Now, he works in University of Regina, Chartered Engineer, Canada. IEEE embedded systems letter, Elsevier microprocessors and Microsystems, IEEE Transactions on industrial electronics, Member of the review committee or organizing committee of IEEE global telecommunications conference and other academic journals, and editor of International Journal of Engineering & Technology for home network management. E-mail: qian.yu@hotmail.com

Zhigang Liu, male, 1980.05, Associate Professor. Now, he works in the Control Technology Institute, Wuxi Institute of Technology. His research interest includes multi-agent collaborative/coordinated control technology, Advanced Manufacturing Digital Management and Digital Enterprise Integration Technology, robot integration technology.

How to cite this paper: Wenguang Song, Hanyu Li, Qian Yu, Wan Li, Bingxin Zhang, Qiujuan Zhang, Zhigang Liu, " The Multimedia Sentiment Model Based on Online Homestay Reviews ", International Journal of Engineering and Manufacturing (IJEM), Vol.10, No.4, pp.13-23, 2020. DOI: 10.5815/ijem.2020.04.02