Retraction

Retraction: Research on Multimodal Emotion Analysis Algorithm Based on Deep Learning (J. Phys.: Conf. Ser. 1802 032054)

Xuan Xiao1,*, Jie Yang2 and Xinyi Ning3

1 College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai 201418, China
2 North China University of Science and Technology, Tangshan 063000, China
3 College of Information, Mechanical and Electrical Engineering, SHNU Shanghai 201418, China

Published 23 April 2021

IOP Publishing Limited ("IOPP") is retracting this paper following an investigation in accordance with COPE guidelines, which has confirmed translated plagiarism.

This work was originally published in Chinese [1] and has been translated and published without permission or acknowledgement to the original authors. IOP Publishing Limited request any citations to this article be redirected to the original work [1].

IOPP have made multiple attempts to contact the author but has not received a response. The author is encouraged to contact jpcnff@ioppublishing.org.

[1] LIN Min-hong, MENG Zu-qiang. Multimodal Sentiment Analysis Based on Attention Neural Network[J]. Computer Science, 2020, 47(11A): 508-514.

Retraction published: 23 April 2021
Research on Multimodal Emotion Analysis Algorithm Based on Deep Learning

Xuan Xiao1, *, Jie Yang2, Xinyi Ning3

1College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai 201418, China
2North China University of Science and Technology, Tangshan 063000, China
3College of Information, Mechanical and Electrical Engineering, SHNU Shanghai 201418, China

*Corresponding author: xuan_xiao@shnu.edu.cn

Abstract. In recent years, more and more people are keen to express their feelings and opinions in the form of pictures and texts on social media at the same time, which makes the multimodal data with pictures and texts as the main content growing. Compared with monomodal data, multimodal data contains more information and can reveal the real feelings of users. The analysis of the emotion of these massive multimodal data is helpful to better understand people's attitudes and viewpoints and has a wide range of application scenarios. In order to solve the problem of information redundancy in multimodal emotion classification tasks, based on the tensor fusion scheme, a multimodal emotion analysis method based on attention neural network is proposed. This method constructs a text feature extraction model and an image feature extraction model based on attention neural network, which highlights the key areas of image emotional information and the words containing emotional information, which makes the expression of single-modal features more concise and accurate. The tensor product of each mode is regarded as the joint feature expression of multimodal data, and the redundant information of joint features is eliminated by principal component analysis, and then the emotion category of multimodal data is obtained by support vector machine. The proposed model is evaluated on two real Twitter image data sets. The experimental results show that, compared with other emotion classification models, this method has a great improvement in classification accuracy, recall rate, F1 index and accuracy.

Keywords: Social media, multimodal data, emotion analysis, attention mechanism, tensor fusion.

1. Introduction
The research institute WeAreSocial released the latest GlobalDigital2019Reports1 on January 31, 2019). The report shows that the number of global social media users, including Twitter, Facebook, Instagram, has grown to 3.5 billion. The average user spends 1 / 3 of his / her Internet time on social media every day. More and more people are keen to express their opinions or
opinions on social media. Hundreds of millions of data records are generated on social media every day, a large amount of which is in the form of a combination of text and images, forming a huge amount of multimodal data. There is abundant emotional information in the massive multimodal data. The emotional analysis of multimodal data is helpful to understand people's attitudes and views on some events, and has great application value in box office prediction [1], political election [2][3], stock market prediction [4][5] and so on.

Therefore, multimodal emotion analysis has received more and more attention in academia and related industries [6].

Figure 1 Example of Twitter picture and text data

In the picture and text data of social media, text and image contain their own emotional information, which are different and complement each other.

Figure 1 shows several picture and text examples of Twitter. Among them, the image and text of figure 1 (a) show that the emotion conveyed by this tweet is negative; the picture in figure 1 (b) will have different feelings for different people, some people will feel empty and lonely, some people will feel beautiful, the emotional polarity of the picture is not strong, and the corresponding text expresses a very strong positive tendency, so this tweet is positive. In figure 1 (c), the text is a declarative sentence with no obvious emotional words, but combined with the picture analysis, the overall emotional tendency of the tweet is negative. In picture-text multimodal data, the information contained in text and picture generally complement each other. Compared with the single-modal data of text or image, multi-modal data contains more comprehensive information and can better show and reveal the real feelings of users.

2. Related work

2.1. Text emotion analysis

As early as 2000, affective analysis has become one of the most active research directions in the field of natural language processing. The method of text emotion analysis can be divided into the method based on emotion dictionary and the method based on machine learning.

The method based on the emotion dictionary is to use the emotion dictionary to calculate the emotion score of the text according to the emotion words in the text sentence to obtain the emotion tendency. These emotional dictionaries are constructed artificially or semi-artificially. Literature [9] proposes to extract the phrases of adjectives and adverbs contained in a sentence, take the difference of mutual information between the phrase and "excellent" and "poor" as the semantic emotion of the phrase, and take the average semantic emotional tendency of the phrase as the emotional tendency of the sentence. Reference [10] proposed a vocabulary-based method to extract emotion from text-semantic orientation calculator (SOGCAL), which uses a word dictionary with semantic orientation (polarity and intensity) annotations to analyze the emotion of a text. The implementation of these methods is simple and fast, but it is very difficult to build emotion dictionaries, and most of the existing emotion dictionaries are based on a certain field and are not universal.
The method of emotion analysis based on machine learning is a more common method in the present research. In 2002, the literature [13] applied machine learning method to classify the emotion of film reviews on the task of emotion analysis. With the rapid development of deep learning, superior results have been obtained in the fields of computer vision and natural language processing, and scholars pay more and more attention to the application and research of deep learning in text emotion analysis.

Literature [11] explicitly obtains the local and global information of the text through convolution and other operations, and can quickly process sentences to obtain the text feature expression, so as to classify the emotion of the text. A text emotion classification method based on deep learning is proposed in reference [12]. Cyclic convolution and cyclic correlation operation are used to calculate the correlation weight between the words in the evaluation document and the evaluation object of the evaluation document, and the weighted sum of the word vector in the document is taken as the expression of the document vector, thus the emotion classification is carried out. A document-level method of emotion classification is proposed in reference [13]. Firstly, the convolutional neural network or short-term memory model is used to learn the sentence representation, and then the gated recurrent neural network is used to adaptively encode the sentence to obtain the document representation. The model has achieved good results in the task of emotion classification.

2.2. Image emotion analysis
Because the emotion of the image is more abstract and subjective, the task of image emotion analysis is more complex than that of text emotion analysis. Some literatures have proposed a method based on low-level image features, which uses the image features and color distribution obtained by the visual word bag model to predict image emotion. Some literatures have proposed a method based on image intermediate features to construct 1200 adjective noun pairs (ANP), and extract visual emotion ontology to classify images. With the development of deep learning, the ability of neural network model to obtain advanced features of image is becoming stronger and stronger. In some literatures, a new depth neural network (NIN), which can improve the recognition ability of local region, is proposed for image emotion analysis. Some literatures use the attention mechanism to detect the emotion-related visual areas of the image spontaneously, which confirms the effectiveness of the attention-based mechanism in the emotion analysis task.

2.3. Multimodal emotion analysis
Nowadays, the research of multimodal emotion analysis is still in its infancy, and it is mostly based on the existing technologies of text emotion analysis and image emotion analysis.

There are two main challenges in the emotional analysis of social media multimodal data. First of all, the emotional information contained in different modal data is different, so it is necessary to effectively obtain the emotional characteristics of each modal data in the emotional analysis of multimodal data. Secondly, the data of different modes are expressed by the underlying features of different dimensions and attributes. Compared with the traditional single modal emotion analysis, multimodal emotion analysis needs an effective way to combine the modal information correctly in order to maximize the interactive information between each modal information and each modal. Some literatures have proposed a tensor fusion scheme in the video emotion analysis task, and proved that the tensor fusion scheme can better retain the interactive information between the modes in the multimodal emotion analysis task, so as to improve the performance of the multimodal emotion analysis model. The dimensions of the constructed multimodal and joint eigenvectors are very large, which is equal to the product of the dimensions of each modal eigenvector. And the joint feature not only contains effective information, but also contains a lot of redundant information. Therefore, if the model is to achieve better results, it is necessary to effectively extract the emotional information of each modal data and the mechanism to eliminate redundant information. In recent years, attention mechanism has been widely used in natural language processing and image processing. The research shows that the attention mechanism enables the neural network to pay more attention to the relevant
parts of the input and less to the irrelevant parts when performing prediction tasks, so as to improve the performance of the model.

The biggest problem in tensor fusion scheme is information redundancy. Then, if the more refined the feature representation of each modal data, the less redundant information of the multi-modal joint features, and the higher the computational efficiency of the model. Therefore, in this paper, the attention mechanism is introduced into the task of multimodal emotion analysis, in order to retain more parts containing emotional information and ignore the parts that have nothing to do with emotional expression when extracting feature expression from single-modal data, it makes the single-mode feature expression more concise and accurate, and then reduces a lot of redundant information in the multi-modal joint feature expression.

On the basis of the tensor fusion scheme proposed by predecessors, this paper proposes a multimodal emotion analysis model (ANNM). Based on attention neural network.

This model uses the method based on attention mechanism to construct the feature extraction network of text and image in order to highlight the key areas of image emotional information and the words containing emotional information.

The tensor product of each mode is taken as the joint feature expression of the multimodal data, and the principal component analysis method is used to reduce the dimension of the joint eigenvector, and then the support vector machine is used to obtain the emotion category of the multimodal data.

3. Multimodal emotion analysis model

In order to effectively obtain the emotional feature representation of each mode and the interactive information between modes from module learning, this paper uses the method based on attention mechanism to construct a multimodal emotion analysis model. The overall structure of the ANNM model proposed in this paper is shown in figure 2. Firstly, two single-mode feature extraction models based on attention mechanism are proposed to obtain the emotional features of images and texts, and then the tensor fusion strategy is used to obtain multi-modal joint feature representation for emotion classification.

3.1. Image feature extraction network

The emotional information of the image is usually more closely related to a certain part of the visual region. As shown in figure 1 (a), the sad expression of the character can arouse people's emotional resonance more than other parts of the image, and it is the area where the emotional information of the image is more relevant. Therefore, when extracting image features, the local feature of facial expression in figure 1 (a) should be highlighted while the influence of other parts should be weakened. The focused information extraction of the image makes the feature expression more refined and the computational efficiency of the model is higher.

![Figure 2](image-url) Structure of multimodal emotion analysis model based on attention neural network
A weight calculation method considering channel domain attention and spatial domain attention is proposed in reference [4]. For multiple feature graphs generated after each convolution calculation, the model needs to know which feature graph should pay more attention to and which part of the feature graph contains more information. Therefore, the calculation of attention weight is mainly divided into two parts: 1) the weight of each feature graph is calculated, and 2) the local weight of the feature graph is calculated. In the convolution neural network (CNNa) based on attention mechanism, which is the image feature extraction network of this model, the attention weight calculation method in reference [4] is used to calculate the attention weight of the output feature graph of the convolution layer. The gradient is calculated by the neural network, and the attention weight is obtained from the main learning through forward propagation and backward feedback.

![Network structure of image feature extraction](image_url)

**Figure 3** Network structure of image feature extraction

The network structure of image emotional feature extraction in this paper is shown in figure 3. In this multi-layer convolution neural network based on attention mechanism, there are 13 convolution layers, and the size of each convolution kernel is $3 \times 3$. Each convolution step goes through three steps: convolution, attention weight calculation, and feature graph weighting calculation to get the final attention feature graph. Then, input the attention feature graph to the next convolution step to continue the calculation. Finally, the output of the final convolution step is used to obtain the image emotional feature vector through the full connection layer.

$I = \{I_1, I_2, ..., I_n\}$ represents the number of image data sets $n$. The convolution neural network based on attention mechanism completes the feature mapping from $\text{CNN}_a(I)$ to $V_i$. Input the picture into the CNNa model to obtain the image feature vector $V_i$. In figure 3, $F_i$ represents the feature map obtained from the first picture after the first convolution layer. $F_i^l$ is the attention feature map obtained by attention weighting. Where $F_i, F_i^l \in \mathbb{R}^{C \times H \times W}$ is the number of channels, $H$ is the length of the feature graph, and $W$ is the width of the feature graph. $A_i^l$ is the attention weight of the first feature graph of the first image, and its expression is as follows:

$$A_i^l = \{a_i^{l,c}, a_i^{l,s}\}$$

Where $a_i^{l,c}$ is the channel attention weight of the 1st feature graph of the I th image. And $a_i^{l,s}$ is the spatial attention weight of the 1st feature graph of the I th image.

Channel attention reflects the contribution of each feature graph after convolution to the key information. The formula for calculating the channel attention weight $a_i^{l,c}$ is as follows:
\[ \alpha_{il}^c = \sigma \left( W_1 \left( W_0 \left( \text{globalavg} \left( F_{il} \right) \right) \right) + W_1 \left( W_0 \left( \text{globalmax} \left( F_{il} \right) \right) \right) \right) \]

Where, \( \text{globalavg} \) (equation) denotes the global average pool function and calculates the average of all feature points of each feature graph. The resulting feature space is \( RC \times 1 \times 1 \), where \( C \) is the number of feature graphs; \( \text{globalmax} \) (1) denotes the global maximum pool function, and calculates the maximum eigenvalue of each feature graph. The resulting feature space is \( RC \times 1 \times 1 \), where \( C \) is the number of channels of the feature graph. \( \sigma \) (attention) is a sigmoid function, and the result is mapped to (0) to obtain the standard channel attention weight, and the channel attention weight \( \alpha_{cil} \in RC \times 1 \times 1 \) is the characteristic graph number. In formula (2), \( W_1 \sim (10) \) \( W_0 \) is a parameter of the neural network, which can be learned from the main body by forward propagation and backward feedback.

The spatial attention weight reflects the contribution of the local area of the picture to the key information, and can find out the areas that need to be paid attention to in the picture information. The formula for calculating spatial attention weight \( \alpha_{sil} \) is as follows:

\[ \alpha_{sil}^s = \sigma \left( \mathbf{f}^T \times \mathbf{f} \left( \left[ \text{avg} \left( \alpha_{il}^c \circ F_{il} \right) \right], \text{max} \left( \alpha_{il}^c \circ F_{il} \right) \right) \right) \]

The calculation formula of attention characteristic graph is as follows:

\[ F_{il}' = F_{il} \circ \alpha_{il}^c \circ \alpha_{il}^s \]

Finally, the attention characteristic graph is used as the input of the next convolution layer to continue the calculation. The output of the final convolution structure is converted into an one-dimensional vector through a full connection layer, that is, the final image feature represents \( V_i \).

### 3.2. Image feature extraction network

![Network structure of text feature extraction](image)

**Figure 4** Network structure of text feature extraction

In the task of emotional classification of text, the emotional information of text is often more related to some words. Words such as 'cry' 'damaged' 'killed' can better reflect the emotions conveyed
by the text than words such as 'house' or 'village'. Therefore, the influence of keywords should be increased in the process of feature extraction of text. In the text feature extraction network of this model, bi-directional gated loop unit (BiGGRU) is used to construct the attention-based text feature extraction network. The output of BiGGRU layer is weighted to highlight the influence of key parts, so as to obtain more accurate text feature expression. The structure of the text feature extraction network is shown in Figure 4.

In this paper, the tensor fusion method [30] is used to fuse the image feature \( V_i = \{v_1, v_2, ..., v_n\} \) and the text feature \( T_i = \{t_1, t_n\} \). The joint feature of the I picture-text data pair is recorded as \( U_i \), and its formula is as follows:

### 3.3. Multimodal fusion

In this paper, the tensor fusion method is used to fuse the image feature \( V_i = \{v_1, v_2, ..., v_n\} \) and the text feature \( T_i = \{t_1, t_n\} \). The joint feature of the I picture-text data pair is recorded as \( U_i \), and its formula is as follows:

\[
U_i = [V_i, 1] \otimes [T_i, 1]
\]

\[
U_i = \begin{bmatrix}
v_1 \cdot t_1 & v_1 \cdot t_2 & ... & v_1 \cdot t_n & v_1 \\
v_2 \cdot t_1 & v_2 \cdot t_2 & ... & v_2 \cdot t_n & v_2 \\
... & ... & ... & ... & ... \\
v_n \cdot t_1 & v_n \cdot t_2 & ... & v_n \cdot t_n & v_n \\
t_1 & t_2 & ... & t_n & 1
\end{bmatrix}
\]

Firstly, the principal component analysis method (PCA) is used to reduce the dimension of joint features in order to reduce the error caused by redundant information and reduce the amount of computation, and then support vector machine (SVM) is used to classify emotions.

Support vector machine (SVM) is one of the most commonly used classification models in machine learning, especially in the classical methods of image classification. Compared with other classifiers, SVM has good adaptability to high-dimensional input data and a large number of samples, and is a prediction tool with good generalization ability. It has achieved good results for both text emotion classification and image emotion classification. Therefore, this method uses SVM to classify the fused feature vectors.

In the fusion feature vector, it contains not only the key interactive information of image-text multimodal data, but also a lot of redundant information, which has little effect on the task of emotion classification and increases the amount of calculation of the classifier; therefore, we need to reduce the dimensionality of the calculation results. In this paper, the principal component analysis method PCA (principalComponentAnalysis) is used to reduce the dimension of the data, and the formula for calculating the emotion category of the unknown sample is as follows:

\[
label = SVM(PCA(U))
\]

### 4. Experiments and results

This section evaluates the performance of the proposed multimodal emotion analysis model through experiments.

#### 4.1. Data set

Because the scarcity of labeled Twitter image data sets is not conducive to the training of the model, this paper uses the existing mature image emotion data sets and text emotion data sets to pre-train the model. The details of the experimental data set are listed in Table 1.
Table 1 Details of the experimental data set

| Type            | Name         | Positive | Negative |
|-----------------|--------------|----------|----------|
| Image dataset   | Twitter1269  | 769      | 500      |
|                 | Twitter10699 | 8443     | 2256     |
| Text dataset    | TwitterText  | 58659    | 73221    |
| Multimodal dataset | Twitter603  | 470      | 133      |
|                 | Twitter2014  | 1008     | 1006     |

There are two image emotion training data sets used in this paper: Twitter1269 data set and Twitter10699 emotion image data set. The former contains 1269 Twitter pictures, of which 769 are positive and the other 500 are negative. The latter is the image emotion analysis data set provided by data tagging company FigureEight2, which contains a total of 10699 Twitter images, of which 8443 are positive and 2256 are negative. The multimodal data sets used in this paper are Twitter603 and TwitGter2014. The former contains a total of 603 picture-text data pairs, including 470 emotional positive picture-text data pairs and 133 emotional negative picture-text data pairs.

The latter is the multimodal data set collected on the Twitter platform for the experiment. A total of 2250 pairs of picture and text data are collected, and the data are emotionally annotated by three researchers. The data with the same category determined by the three researchers were retained, and the controversial data were eliminated, resulting in a multimodal dataset containing 1008 emotional positive data pairs and 1006 emotional negative data pairs.

4.2. Pre-training of the model
Due to the scarcity of multimodal data sets, the training of the model becomes difficult. In order to solve this problem, the image feature extraction network and text feature extraction network are pre-trained by using image emotion classification data set and text emotion classification data set respectively. In this paper, a three-layer fully connected neural network classifier is added to the image feature extraction network and text feature extraction network, which is constructed into image emotion classification model and text emotion classification model.

In the training of the image emotion classification model, the image classification pre-training model trained on the ImageNet data set is used. 3) the parameters of the convolution module of the image emotion classification model are initialized, and the parameters of the attention calculation module are initialized by normal distribution random decimals. The pre-training model of the image emotion feature extraction network is obtained by fine-tuning the model on the image emotion data set TwitGter10691 and Twitter1269 data set. Similarly, in the training of the text emotion classification model, the parameters of the model are initialized by normal distribution random decimals. The model is trained on the TwitterText data set, and the pre-training model of text emotional feature extraction network is obtained.

4.3. Experimental results
In order to verify the effectiveness of the model proposed in this paper, the proposed model is compared with the single-modal emotion classification model and the multi-modal emotion classification model based on neural network.

- single modal emotion classification model
  VGG16 is a classical image classification model, which is fine-tuned for image emotion classification in this paper.
  CNNA is an image emotion classification model based on the image feature extraction network in this paper.
  BiGRUGText model is a text emotion classification model based on bi-directional GRU network structure.
AbiGrugText model is a BiGRUGText model based on attention mechanism. The calculation of attention weight is consistent with the emotional feature extraction network proposed in this paper.

(1) the multimodal emotion classification model

NNM model is a multimodal emotion analysis model based on neural network with tensor fusion scheme, and it is a simplified model of ANNM without attention mechanism.

| Table 2 | Experimental results on Twitter2014 dataset |
|---------|-----------------------------------------------|
| Model   | Precision | Recall | F1 score | Accuracy |
| VGG16   | 0.6216    | 0.5833 | 0.6018   | 0.6137   |
| BiGRU-Text | 0.7505    | 0.7133 | 0.7314   | 0.7375   |
| NNM     | 0.8107    | 0.7391 | 0.7732   | 0.7830   |
| CNNa    | 0.7325    | 0.6329 | 0.6791   | 0.7096   |
| AbiGrUG-Text | 0.8313    | 0.7917 | 0.8110   | 0.8113   |
| ANNM    | 0.8817    | 0.8284 | 0.8542   | 0.8553   |

| Table 3 | Experimental results on Twitter603 dataset |
|---------|-----------------------------------------------|
| Model   | Precision | Recall | F1 score | Accuracy |
| VGG16   | 0.7762    | 0.7085 | 0.7378   | 0.7443   |
| BiGRU-Text | 0.8063    | 0.7872 | 0.7657   | 0.8049   |
| NNM     | 0.8228    | 0.8290 | 0.8263   | 0.8280   |
| CNNa    | 0.8084    | 0.7891 | 0.7944   | 0.8049   |
| AbiGrUG-Text | 0.8463    | 0.8319 | 0.8391   | 0.8512   |
| ANNM    | 0.8565    | 0.8766 | 0.8665   | 0.8789   |

Tables 2 and 3 show the experimental results of different models on two Twitter graph and text data sets. Generally speaking, the model proposed in this paper performs better than other models on the two experimental data sets. As can be seen from the results in Table 2, compared with the single-modal emotion classification model, the multi-modal emotion classification model has a better performance in each evaluation index. NNM multimodal emotion classification model is compared with text emotion classification model BiGRUGText model and image emotion classification model VGG16 model. The accuracy of ANNM multimodal emotion classification model has been improved by 4.5% and 16.9%, respectively. Compared with text emotion classification model AbiGRUGText model and image emotion classification model CNNa model, the accuracy of ANNM multimodal emotion classification model has been improved by 4.3% and 15.8%. This fully shows that multimodal data can better reveal the true feelings of users. It is also proved that the tensor fusion method combined with each modal feature vector can effectively make use of the complementary information among the modes, so as to improve the effect of emotion classification.

The accuracy of CNNa model based on attention mechanism is 8.69% higher than that of VGG16 model, and that of AbiGRUGText model based on attention mechanism is 7.75% higher than that of BiGRUGText model, which proves that attention mechanism can extract emotional information of image and text more effectively, thus improving the classification effect. In particular, the ANNM model has more than 7% improvement in each index compared with the NNM model. The model based on attention neural network can greatly improve the classification results, which shows that the ANNM model in this paper is feasible.

As can be seen from the results in Table 3, the ANNM model proposed in this paper is still the best among many models. In addition, the accuracy, recall and F1 score results of all models in Table 3 are better than those of other models. The accuracy is lower than other evaluation indexes. This situation shows that the model has a strong ability to identify positive examples, that is, positive data, while weak ability to identify negative data, that is, negative examples. This may be due to the large proportion of positive and negative samples in Twitter603 data sets. However, even if there is the problem of data imbalance, the model proposed in this paper still has a good performance, which shows that the ANNM model is effective.
5. Conclusions
In this paper, a multimodal emotion analysis method based on attention mechanism is proposed. The results show that the model proposed in this paper produces better classification results. Due to the limitation of existing data resources and level, there is room for further improvement in this paper. Our proposed model only considers the perfect picture-text multimodal data of social media, but in real life, social media data is a multimodal data set in which text, image and picture-text data pairs coexist. In the future, we will supplement our data resources and propose a more effective model in the case of modal loss in multimodal data.

References
[1] Asur S. Predicting the future with social media (informal publication) [J]. ArXiv e-prints, 1003.5699, 2010.
[2] O'Connor B. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series [J]. Icwsm, 2010.
[3] Unankard S, Li X, Sharaf M, et al. Predicting Elections from Social Networks Based on Subevent Detection and Sentiment Analysis [C]// International Conference on Web Information Systems Engineering. Springer, Cham, 2014.
[4] Woo S, Park J, Lee JY, et al. CBAM: Convolutional Block Attention Module [J]. 2018.
[5] Li X, Xie H, Chen L, et al. News impact on stock price return via sentiment analysis [J]. Knowledge Based Systems, 2014, 69(oct.): 14-23.
[6] Nguyen TH, Shirai K. Topic Modeling based Sentiment Analysis on Social Media for Stock Market Prediction [C]// Meeting of the Association for Computational Linguistics & the International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing. 2015.
[7] Zhang L, Zhao Y, Zhu ZF. Advances in Semantically Shared Subspace Learning for Cross-Media Data [J]. Chinese Journal of Computers, 2017.
[8] Zhang L, Wang S, Liu B. Deep Learning for Sentiment Analysis: A Survey [J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2018: e1253.
[9] Thumbs Up or Thumbs Down, bạ o cá o khoa họ, mồ hinh ngón ngữ. Bạ cọ khoa hc: Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews [J]. Tailieu Vn.
[10] Taboada M, Brooke J, Tofiloski M, et al. Lexicon-Based Methods for Sentiment Analysis [J]. Computational Linguistics, 2011, 37(2):267-307.
[11] Baccianella S, Esuli A, Sebastiani F. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining[C]// Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010, Valletta, Malta. DBLP, 2010.
[12] Hu M, Liu B. Mining and summarizing customer reviews [C]// Tenth Acm Sigkdd International Conference on Knowledge Discovery & Data Mining. ACM, 2004.
[13] Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment Classification using Machine Learning Techniques [J]. Proceedings of Emnlp, 2002.