Duan M (2025). SOCIAL MEDIA DATA ANALYSIS AND SENTIMENT RECOGNITION FOR ATHLETES IN SPORT MANAGEMENT. Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte vol. 25 (99) pp. 418-434.
DOI: https://doi.org/10.15366/rimcafd2025.99.027

ORIGINAL

SOCIAL MEDIA DATA ANALYSIS AND SENTIMENT RECOGNITION FOR ATHLETES IN SPORT MANAGEMENT

Mingtao Duan

School of Physical Education, Henan Normal University, Xinxiang 453007, Henan, China
E-mail: dmtqd266@163.com

Recibido 16 de Marzo de 2024 **Received** March 16, 2024
Aceptado 22 de Octubre de 2024 **Accepted** October 22, 2024

ABSTRACT

The emotional state of athletes is crucial to the level of competition and training, especially during the competition, the emotional fluctuations of athletes have a direct impact on the performance of the competition. Recognizing the emotions of athletes through their social media data can detect the fluctuations in their emotions in a timely manner, which can be used to adjust the training plan or intervene early to avoid affecting the competition results. Using machine learning and deep learning techniques, artificial intelligence can analyze multiple data forms such as text, voice, image and video to identify and understand the emotional state of athletes. In this paper, we propose a multi-granularity sentence sentiment analysis method, which constructs the whole sentence sentiment analysis model DABLSTM-L1 by fusing sentiment word vectors and high-level semantic features, and constructs the interactive attention sentence sentiment recognition model Att-CNN-BLSTM to extract the interactive sentiment features of the whole sentence and local sentences, and finally fuses the whole sentence sentiment analysis model DABLSTM-L1, local sentence sentiment analysis model and interactive attention sentence sentiment classification model Att-CNN-BLSTM to predict the sentiment polarity of text. The experimental results show that multi-granularity sentence sentiment analysis can effectively identify the emotional state of athletes.

KEYWORDS: Sentiment Recognition; Sport Management; Social Media

1. INTRODUCTION

With the rapid development of mobile Internet, social media represented

by Weibo and Twitter have gradually become the main platform for netizens to record their lives and express their opinions, which generates a huge amount of data containing emotional tendencies. These data are in various forms, usually coexisting in various modalities such as text, image, video, etc., among which the most common is the combination of text and image (Y et al., 2024). Athlete emotion recognition helps to help sport administrators and coaches better understand and support the emotional state of athletes, thereby improving training and competition. Through emotion recognition, an athlete's anxiety, stress, or other emotional issues can be detected in a timely manner and steps can be taken to provide support and assistance. This helps to safeguard the mental health of athletes, reduce the risk of injury, and improve focus and execution. Additionally, emotion recognition can be used to assess an athlete's response to training programs and competition scheduling so that individualized adjustments can be made to improve performance and satisfaction. Ultimately, through athlete emotion recognition, sport administrators and coaches can create training and management models that are more relevant to the needs of their athletes, promoting their overall development and success. Additionally, emotion recognition can help monitor an athlete's stress response, intervene in a timely manner and provide necessary support. Ultimately, this helps promote team cohesion and individual athlete growth, which can have a positive impact on improving overall athleticism and team performance. Artificial intelligence plays a central role in emotion recognition. Through machine learning and deep learning techniques, AI can analyze multiple forms of data such as text, speech, images, and video to recognize and understand human emotional states (Chen, Han, et al., 2023; Chen, Li, et al., 2023; Li & Cao, 2021). This technology enables AI to help companies better understand the emotional needs of their customers and improve the quality of their products and services. In sports management, AI's emotion recognition technology can help coaches and managers better recognize the emotional states of athletes to optimize training programs and provide more personalized support. Therefore, the role of AI in emotion recognition is very important and extensive. Text feature extraction is generally divided into two steps: text feature representation, text feature extraction. In practice, a pre-trained Word Embedding model is usually used to map the text into a low-dimensional representation, and then input into a neural network for feature extraction. (Kamyab et al., 2021) used the GloVe model to obtain the word embedding, combined with the N-gram features of the text and input it into a deep convolutional neural network for sentiment classification. Rehman et al. [6] input the word embeddings obtained from Word2Vec model into a convolutional neural network to extract local features, and then used Long Short-Term Memory (LSTM) to learn the long-term dependency between word sequences. LSTM can capture the contextual information of the text better, but the model cannot focus on the important information in the sentence, and the high-dimensional text input leads to high model complexity and optimization

difficulties. To solve this problem, the attention mechanism can be used to make the model focus on important words and sentences and use convolutional neural networks for feature extraction and dimensionality reduction (Rehman et al., 2019). Although the convolutional neural network can extract the depth features of the image, it does not consider the different degree of contribution of different regions of the image to the target task. In order to solve this problem, related scholars introduced attention mechanisms, such as channel attention (Hu et al., 2018) and spatial attention (Basiri et al., 2021). In 2020, Google proposed Vision Transformer (ViT) model (Woo et al., 2018), which proved that the model based on the Transformer structure of self-attention can also be used as a visual backbone network. On large datasets, ViT achieves accuracy beyond ResNet. It is shown that there is a difference between ViT and ResNet in their ability to integrate global information (Dosovitskiy et al., 2020). The ResNet model, which takes convolutional pooling as its core operation, strictly adheres to the process of refining global features from local features, and the local information in its generated depth features is less refined, while ViT mixes local and global information in the process, and the attention mechanism enables its depth features to retain local spatial information in a more refined way (Huang et al., 2016). Sentiment recognition based on text and images has made great progress at this stage, but still faces numerous challenges. Most of the existing image sentiment analysis only focuses on the whole image and pays less attention to the influence of saliency targets and face targets on image sentiment analysis, resulting in the inability to fully explore the region that expresses image sentiment. Most of the existing text sentiment analysis methods use LSTM long and short-term memory networks to extract high-level semantic features of the text, while ignoring the low-level semantic features of the text, resulting in the loss of text sentiment features. In this paper, we address the above-mentioned problems and apply deep learning techniques to conduct sentiment recognition research based on social media information of athletes' population. For overall sentence sentiment analysis in text research, high-level semantic sentiment features and low-level word vector sentiment are integrated to make up for the insufficiency of using high-level semantic features or low-level word vector features alone. Analyzing the dependency and complementarity between local sentences and overall sentences, an interactive information feature combining local sentence information and overall sentence information is constructed.

2. Multi Granularity Sentence-Based Sentiment Analysis for Graphic Fusion

2.1 Sentiment Analysis of Multi-Granularity Sentences

In order to simultaneously mine the sentiment features of the whole sentence and the sentiment features of local sentences, a multi-granularity sentence sentiment analysis is proposed. First, the local sentiment of the

sentence is identified according to the local sentence sentiment analysis model. Then, the sentiment recognition model DABLSTM-L1 (double attention Bidirectional Long Short-Term Memory-L1) of the whole sentence is constructed to identify the sentiment of the whole sentence, and the sentence sentiment analysis model Att-CNN-BLSTM with interactive attention is constructed to get the interactive Sentiment between overall sentence and the local sentence. Finally, we integrate the results of the three to get the final sentence sentiment polarity.

2.1.1 Text Preprocessing and Word Vector Representation

(1) Text Preprocessing: The data crawled from athletes' social media not only have sentences with subjective sentiment, but also with some noise data, such as html tags, special symbols and so on. In order to avoid the interference of these noisy data on sentiment analysis, these data are preprocessed:

1. Filter out all html tags, punctuation marks and emoticons, keeping only sentence-related text. And according to the deactivation word list data, remove the words contained in the deactivation word list. 2. Take the sentences obtained in the first step and segment them using jieba segmentation technique. 3. tagging this data, marking positive sentences as 1 and negative sentences as 0.

(2) Word Vector Representation: Since convolutional neural networks process data in batches, a fixed size of data is input each time. However, since each sentence has a different length, a sequence of length L is fixed to be input, and if there are not enough sentences then they are filled with zeros. If that length is exceeded, the later part is truncated. For word embedding techniques, there are static vector representation, non-static vector representation, multi-channel representation and random initialization representation. In this paper, static vectors are used, i.e., word vector representations are obtained by training in a large-scale corpus using word2vec technique. Splice n word vectors to get the sentence matrix, denoted as:

$$S = \{V(x_1) \oplus V(x_2) \oplus \cdots \oplus V(x_n)\} \quad (1)$$

where $\{x_1, x_2, \cdots, x_n\}$ denotes the set of n words contained in a sentence. The splicing operation symbol is $\oplus$ and $\{V(x_1), V(x_2), \cdots, V(x_n)\}$ denotes the word vector corresponding to the n words in this set.

2.1.2 CNN-based Localized Sentence Sentiment Analysis

Convolutional Neural Networks perform well in extracting local semantic features at different locations in a text sequence. It can capture the local key semantic features and filter out some noise and sentiment irrelevant words. In order to capture the sentiment of local sentences (phrases), a local sentence

sentiment analysis model CNN is constructed, which consists of a word vector representation layer, a convolutional layer, a pooling layer, a feature representation layer and a classification layer.

(1) Convolutional layer: Firstly, one-dimensional convolution is used to extract the local feature information, and the filter is used to slide over the text sequence in order to detect the local semantic-emotional feature information at different locations. Let $F \in R^{w \times d \times k_{loc}}$ be the filter w in the convolution operation, d where k_{loc}, and represent the width, number, and dimension of the input of the convolution filter, respectively. The height of the convolutional filter is equal to the input dimension d, which facilitates the filter to scan the text sequence matrix. For the word at position i, the word is embedded in the text subsequence $x_{i-w/2+1:i+w/2}$ as input if w is even, and $x_{i-\lfloor w/2 \rfloor+1:i+\lfloor w/2 \rfloor}$ as input if it is odd. If the number of text subsequences is less than w, then the shortfall is filled with zeros. The formula for the convolution operation is shown in 2.

$$\tilde{c}_i = \begin{cases} f\left(x_{i-w/2+1:i+w/2} \times F + b\right), & w = 2n, n \in (1,2,\cdots,n) \\ f\left(x_{i-\lfloor w/2 \rfloor:i+\lfloor w/2 \rfloor} \times F + b\right), & w = 2n+1, n \in (1,2,\cdots,n) \end{cases} \quad (2)$$

where $x_{i-w/2+1:i+w/2}$ and $x_{i-\lfloor w/2 \rfloor+1:i+\lfloor w/2 \rfloor}$ are both matrices stitched together by row word vectors, representing the word embedding matrices $\left(x_{i-w/2+1},\cdots,x_i,\cdots,x_{i+w/2}\right)$ and $\left(x_{i-\lfloor w/2 \rfloor+1},\cdots,x_i,\cdots,x_{i+\lfloor w/2 \rfloor}\right)$, respectively. * Represents the convolution operation, b is the bias term, f is the activation function, and $\tilde{c}_i$ represents the w meta-local sentiment feature vector at the $i-th$ position in a text sequence. The feature map after convolution is noted as $\tilde{C}$, $\tilde{C} = [\tilde{c}_1, \tilde{c}_2, \cdots, \tilde{c}_T]$, where $\tilde{C} \in R^{W_{out} \times k_{loc}}$. In order to learn local semantic sentiment features for word sequences of different lengths, the input word sequence matrix is scanned using filters of multiple sizes, and the final variable local w-gram features are notated as C, $C = \left[\tilde{C}^{(1)}, \tilde{C}^{(2)}, \cdots, \tilde{C}^{(r)},\right]$, where $C \in R^{T \times r k_{loc}}$.

(2) Pooling layer: Due to the increase in the number of convolutional layers, there is also a high demand for computer performance. Meanwhile, for the sentiment analysis task mainly a two-dimensional vector is obtained to backpropagate the learning parameters, which makes the difference between this vector and the label smaller and smaller, in order to get this two-dimensional vector as soon as possible and to reduce the amount of parameter computation. For this reason, a pooling layer is added between each convolutional layer, specifically, the local sentence representation is obtained

by averaging over the feature mapping through a global average pooling technique:

$$\bar{C} = avg(C) = avg\left(\tilde{C}^{(1)}, \tilde{C}^{(2)}, \cdots, \tilde{C}^{(r)}\right) \quad (3)$$

(3) Localized Sentence Feature Representation: From the previous step the same number of feature maps of different sizes are obtained through the pooling layer, which are stitched together using the concatenate method as new fused semantic features:

$$\hat{C} = \left[avg\left(\tilde{C}^{(1)}\right) \oplus avg\left(\tilde{C}^{(2)}\right) \oplus \cdots \oplus avg\left(\tilde{C}^{(n)}\right)\right] \quad (4)$$

(3) Emotional categorization layer: From the previous step, we get the spliced feature $\hat{C}$, then go through 2 fully connected layers to get $\hat{C}$, and then pick up a softmax layer to get the probability $\hat{y}_{local}$ for local sentence sentiment classification.

$$\hat{y}_{local} = soft\,max\left(W\hat{C} + b\right) + b \quad (5)$$

2.1.3 Sentence Sentiment Analysis Fusing Sentiment Word Vectors and High-Level Semantic Features

In order to judge the sentiment polarity of the overall sentence by combining the high-level semantic features of the sentence and the low-level word vector features, the overall sentence sentiment analysis model DABLSTM-L1 is constructed. First, the sentence vectors are input to the bidirectional LSTM to get the hidden layer features, the important weights of different word vectors are obtained through the attention mechanism, and the hidden layer sentiment features are obtained by combining the word vector weights and the hidden layer features. Then the word vector weights are regularized by L1 to get the sparse attention weights, and then combined with the word vectors to get the emotion word features. Finally, the hidden layer sentiment features and sentiment word features are used together to judge the sentiment polarity of the sentence.

(1) Emotion word vector feature extraction based on sparse attention mechanism: Bidirectional LSTM Two LSTMs in different directions are connected to each input sequence x_i to obtain the past sequence and future sequence information respectively.

$$h_i = \left[\overrightarrow{h}_i, \overleftarrow{h}_i\right] \quad (6)$$

$\overrightarrow{h}_i$ represents the past hidden layer features learned by the $i-th$ word

using the previous $i-1$ words, and $\overleftarrow{h}_i$ represents the future hidden layer features learned through the latter $n-i+1$ words. In BLSTM network, h_i represents the hidden feature information learned from the $i-th$ word sequence. It is the hidden layer feature with both past and future information obtained through the element-by-element summation of $\overrightarrow{h}_i$ and $\overleftarrow{h}_i$, and the whole input sentence S can be represented as $d=[h_1,h_2,\cdots,h_n]$. The human brain tends to judge the emotional polarity of the whole sentence by some key emotional word information when observing the whole sentence. In recent years, the attention mechanism has become more and more popular, which is combined with neural networks such as LSTM to capture the different importance levels of different words in a sentence, so as to better express the high-level semantic information of the whole sentence. Similarly, in this paper, the self-attention mechanism is firstly used in order to extract the emotional importance coefficient of each word's learned emotional features in the whole sentence hidden layer sequence representation d, which is calculated as follows:

$$\alpha_i = soft\,max(W \times d + b) \tag{7}$$

W represents the weight matrix of the neural network and b is the corresponding bias term of the neural network. The hidden layer representation d is subjected to a linear mapping and then normalized by a softmax function to obtain α_i, which represents the degree of correlation between the sentiment information of the $i-th$ word and the sentiment polarity of the whole sentence. The weight of a sequence of words in a sentence is denoted as $\alpha=[\alpha_1,\cdots,\alpha_i,\cdots,\alpha_n]$, where, $\alpha_i \in (0,1)$ if the larger α_i is, it means that the $i-th$ sentiment word in a sentence will have a greater influence on the sentiment polarity of that sentence. Since there are some words in a sentence that do not have influence on sentiment polarity, in order to avoid the interference of these words with no sentiment expression on the sentiment polarity judgment of the whole sentence, it is considered to de-eliminate them by sparsifying α_i with L1 paradigm, so that this part of the words will have a weight of 0. The calculation formula is as follows:

$$|\alpha|_{L1} = |f(W \times d + b)| \tag{8}$$

The weights, which have been sparsified by the L1 paradigm, are multiplied with the input variables to obtain the sentiment word feature representation of the whole sentence, d', which is calculated as follows:

$$d' = \alpha_1' x_1 + \cdots + \alpha_i' x_i + \cdots + \alpha_n' x_n \tag{9}$$

where x_i represents the $i-th$ word embedding vector of the whole sentence, and α_i' represents the new word vector weights obtained by adding L1 regularization to the hidden layer d after the attention mechanism.

(2) Advanced semantic feature extraction based on attention mechanism:

The sentence sentiment feature representation s' is obtained by multiplying each word importance score learned through the attention mechanism and the hidden layer feature d, which is calculated as follows:

$$s' = \alpha_1 h_1 + \cdots + \alpha_i h_i + \cdots + \alpha_n h_n \quad (10)$$

(3) Fusion of low-level word vector features and high-level semantic sentiment features: The sentence sentiment feature representation s' and the sentiment word feature layer representation d' are spliced together to obtain the sentiment feature q with both sentence and sentiment words, calculated as follows:

$$q = [d' \oplus s] \quad (11)$$

The final q is used as a fusion feature, connected to a fully connected layer and a softmax layer to predict the sentiment polarity of the whole sentence. The formula is as follows:

$$\hat{y}_{global} = soft\,max(Wq + b) \quad (12)$$

$\hat{y}_{global}$ is a two-dimensional vector representing the final predicted sentiment polarity probability (y_{pos}, y_{neg}) of the whole sentence, where $y_{pos} + y_{neg} = 1$, if $y_{pos} > 0.5$, the sentiment polarity of the whole sentence is positive, and vice versa predicts that the sentiment polarity of the whole sentence is negative.

$$L = \frac{1}{N}\sum_i L_i = \frac{1}{N}\sum_i -[y_i \times log(p_i) + (1 - y_i) \times log(1 - p_i)] \quad (13)$$

Assuming $p_i = y_{pos}$, then $1 - p_i = y_{neg}$, y_i for positive labeling is equal to 1 and $(1 - y_i)$ for negative labeling is equal to 0. From this, we can see that when the samples are positive, only $y_i \times log(p_i)$ is computed, and vice versa $(1 - y_i) \times log(1 - p_i)$ is computed. The loss of all the samples is added up and then divided by the total number of samples N to get the average loss of all the samples, and then after the back propagation mechanism, the parameters of the neural network are updated so that that loss is reduced to a low level and is in a stable state, at which point the training is stopped.

2.1.4 Interactive Attention Sentence Sentiment Analysis Based on CNN-BLSTM

Sentence sentiment analysis framework based on CNN-BLSTM with interactive attention is shown in Figure 1. Firstly, local sentence sentiment features are obtained based on CNN, then input to BLSTM and combined with the attention mechanism to get the interactive semantic sentiment features of the overall sentence and the local sentence, and finally the interactive

sentiment polarity of the sentence is obtained after full connectivity and Softmax function.

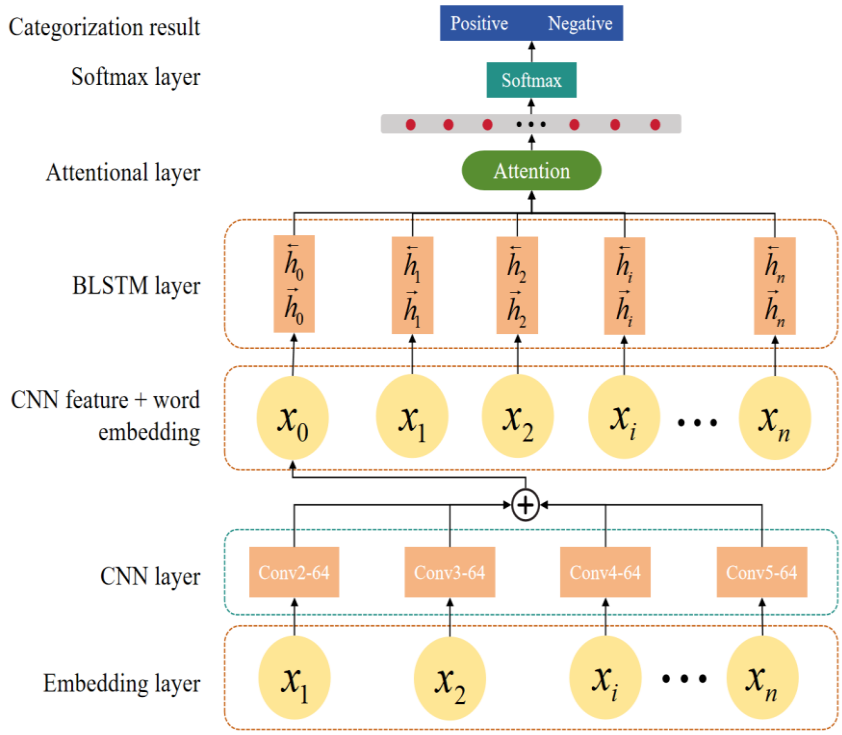

Figure 1: CNN-BLSTM based interactive attention sentence sentiment analysis framework

In order to capture the semantic interaction information between phrase level and word level, a sentence sentiment recognition model Attention-CNN-BLSTM (Attention CNNBLSTM) with cascaded CNN and BLSTM for interactive attention is constructed. The specific process is as follows: firstly, after the sentence input to CNN to get the local sentence features, and then combined with word vectors as a new embedding layer, input to the bidirectional LSTM, to get the interaction semantic information features $(h_0, h_1, \cdots, h_n)$ between the overall sentence and the local sentence, and finally add the attention mechanism to the feature with the weight of the interaction semantic sentiment features.

(1) CNN layer. Local sentence features are extracted using 64 CNNs each with filters of 2×200, 3×200, 4×200, and 5×200, and calculated according to Equation (14):

$$x_0 = CNN(S) \tag{14}$$

where S is the sentence vector matrix, obtained from Equation (1). x_0 denotes the output of the CNN.

(2) BLSTM layer. The feature x_0 output by CNN and the embedding layer sentence matrix S are spliced to get $X = x_0 \oplus S$, and the interaction

semantic feature $H = [h_0, h_1, \cdots, h_n]$ is obtained by BLSTM.

$$H = BLSTM(X) \tag{15}$$

(3) Attention mechanism layer. Input H into the attention mechanism layer to get the weights of individual word vectors of the local sentence and the whole sentence, and combine with the interaction semantic feature H to get the interaction semantic feature $H^{'}$ with weights.

$$H^{'} = H * attention(H) \tag{16}$$

(4) Classification layer. The sentiment polarity of the interacting sentences is obtained by Softmax function, which is calculated as follows:

$$\hat{y}_{inter} = soft\,max\left(WH^{'} + b\right) \tag{17}$$

3. Multi-visual and Multi-Granularity Graphic Fusion Sentiment Analysis

Multi-visual multi-granularity graphic fusion sentiment analysis is composed of multi-visual target image sentiment analysis and multi-granularity sentence sentiment analysis, as shown in Figure 2. Multi-visual target image sentiment analysis is composed of saliency target sentiment analysis model SFPN-CNN, face target sentiment analysis model WSB-CNN and whole image sentiment analysis model FCNN. Multi-granularity sentence sentiment analysis consists of whole sentence sentiment analysis model DABLSTM-L1, local sentence sentiment analysis CNN and sentence sentiment analysis model with interactive attention Att-CNN-BLSTM. Firstly, the sentiment probabilities of image and text are obtained through multi-visual target image sentiment analysis and multi-granularity sentence sentiment analysis, then the feature sentiment probabilities obtained through feature layer fusion, and finally the final graphic and text fusion sentiment analysis prediction is obtained by fusing the sentiment probabilities of the three through decision layer.

(1) The face features obtained by the face emotion recognition model WSB-CNN are calculated as follows.

$$A = f_a\left(I_{face}; \theta_a\right) \tag{18}$$

Where I_{face} is the detected face image, $f_a(\cdot)$ represents the face emotion recognition model WSB-CNN, and θ_a represents the set of parameters of WSB-CNN. After obtaining the face features, the face emotion probability Y_a is obtained through two fully connected layers and is calculated as follows:

$$Y_a = f_a(A; \theta_a) \tag{19}$$

where θ_a is the set of fully connected parameters of the face emotion recognition model. The cross-entropy loss function of the face emotion recognition model is defined as follows:

$$L^a(A;\theta_a,\theta_{a'}) = \sum_{i=1}^{n} -log(Y_a) = \sum_{i=1}^{n} -log\left(f_a\left(f_a(I_{face};\theta_a)\right);\theta_{a'}\right) \quad (20)$$

The parameters θ_a and $\theta_{a'}$ are automatically learned by minimizing the cross-entropy loss on the face training set, and this loss function is optimized using stochastic gradient descent.

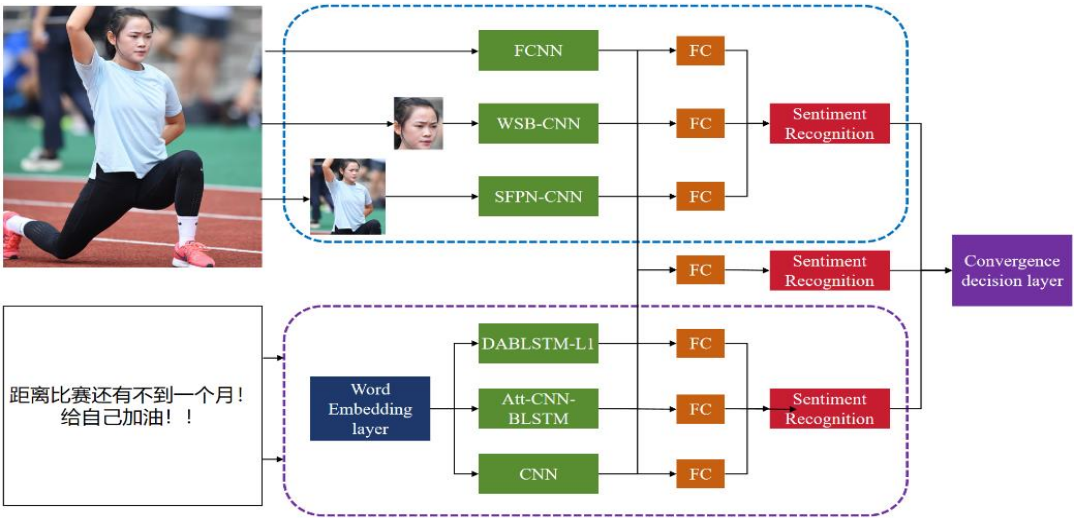

Figure 2: Block diagram of multi-visual multi-granularity graphic fusion sentiment analysis

(2) The salient target features obtained by the salient target sentiment recognition model SFPN-CNN are calculated as follows:

$$B = f_b(I_{object};\theta_b) \quad (21)$$

Where I_{object} is a saliency target in an image, f_b represents the saliency target sentiment recognition model SFPN-CNN, and b represents the set of parameters of SFPN-CNN. Predicting the significance target sentiment probability Y_b through the fully connected layer is calculated as follows:

$$Y_b = f_{b'}(B;\theta_{b'}) \quad (22)$$

where $\theta_{b'}$ is the fully connected parameter of the salience target sentiment recognition model. The cross-entropy loss function of the salience target sentiment recognition model is defined as follows:

$$L^b(B;\theta_b,\theta_{b'}) = \sum_{i=1}^{n} -log(Y_b) = \sum_{i=1}^{n} -log\left(f_{b'}\left(f_b(I_{object};\theta_b)\right);\theta_{b'}\right) \quad (23)$$

(3) The whole image features obtained by the whole image emotion recognition model FCNN are calculated as follows:

$$C = f_{c'}(C; \theta_{c'}) \tag{24}$$

where $\theta_{c'}$ is the set of fully connected parameters of the whole image sentiment recognition model. The cross-entropy loss function of the whole image sentiment recognition model is defined as follows:

$$L^c(C; \theta_c, \theta_{c'}) = \sum_{i=1}^{n} -log(Y_c) = \sum_{i=1}^{n} -log\left(f_{c'}\left(f_c(I_{object}; \theta_c)\right); \theta_{c'}\right) \tag{25}$$

(4) The overall sentence features are obtained by the whole sentence sentiment recognition model DABLSTM-L1 with the following formula:

$$D = f_d(T; \theta_d) \tag{26}$$

where T is a sentence sequence matrix, f_d represents the sentence sentiment recognition model DABLSTM-L1, and d represents the set of parameters of the DABLSTM-L1 model. The entire sentence sentiment classifier is obtained through the fully connected layer and is computed as follows:

$$Y_d = f_d(D; \theta_{d'}) \tag{27}$$

where $\theta_{d'}$ is the fully connected parameter of the overall sentence sentiment recognition model. The cross-entropy loss function of the overall sentence sentiment recognition model is defined as follows:

$$L^d(D; \theta_d, \theta_{d'}) = \sum_{i=1}^{n} -log(Y_d) = \sum_{i=1}^{n} -log(f_{d'}(f_d(T; \theta_d)); \theta_{d'}) \tag{28}$$

(5) Localized sentence sentiment features are obtained by CNN with the following formula:

$$E = f_e(T; \theta_e)$$

where T is a sentence sequence matrix, f_e represents the local sentence sentiment recognition model CNN, and θ_e represents the set of parameters of the CNN.

4. Experiments

4.1 Experimental Data Set

Since multimodal sentiment analysis is in the preliminary research stage,

there are few publicly available image and text data. Firstly, we crawl Sina Weibo data (including images and text), and then we do pre-process on the crawled data to filter news and advertisements. After pre-processing the data, it is assigned to 5 workers who label the data, and each piece of data is labeled by these 5 workers, if there are more than 3 workers labeled as positive, then the microblogging data is identified as positive, and vice versa as negative. Finally, 6587 positive tweets and 3621 negative tweets were obtained, as shown in Table 1.

Table 1: Image Sentiment Dataset

POLARITY OF EMOTION	NO. OF SOCIAL MEDIA
POSITIVE	6587
NEGATIVE	3621
TOTAL	10208

4.2 Experimental Results and Analysis

4.2.1 Evaluation Indicators

In order to verify the effectiveness of the multi-granularity sentence sentiment analysis method based on the interactive attention mechanism, the accuracy rate, recall rate, precision rate and F-value were adopted as the evaluation criteria for sentiment classification, respectively, and the formulas were calculated as follows:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{29}$$

$$Recall = \frac{TP}{TP+FN} \tag{30}$$

$$Precision = \frac{TP}{TP+FP} \tag{31}$$

$$F = \frac{Precision \times Recall \times 2}{Precision+Recall} \tag{32}$$

Where TP is the number of positive samples predicted correctly, TN is the number of negative samples predicted correctly, FN is the number of positive samples predicted as negative samples, and FP is the number of negative samples but predicted as positive samples.

4.2.2 Hyperparameter Settings

In order to choose the appropriate hyperparameters, the experimental word vector dimensions were tested in 100, 200 and 300 dimensions, the

number of convolutional filters in CNN was taken as 64 and 128 for testing one by one, and the hidden layer nodes of BLSTM were taken as 100, 128 and 256 for testing. After several experiments, it is found that the best effect of text sentiment classification is achieved when the model parameters are as shown in Table 2.

Table 2: Model Parameters

PARAMETER	VALUE
WORD VECTOR DIMENSION	200
NUMBER OF CONVOLUTION FILTERS	64
NUMBER OF HIDDEN LAYER NODES OF BLSTM	128
LEARNING RATE	0.001
OPTIMIZER	ADAM
BATCHSIZE	128

4.3 Comparative Experiments

In order to verify the effectiveness of the multi-granularity sentence sentiment analysis algorithm based on the interactive attention mechanism, comparative experiments are conducted with the currently used neural network models, and the results are shown in Table 3.

(1) CNN: a localized sentence sentiment analysis model from literature (Raghu et al., 2021). (2) LSTM: LSTM model from literature (Kim, 2014). (3) Att-BLSTM (word): add attention weights to word vectors. (4) Att-BLSTM-L1(word): Att-BLSTM with L1 regularization. (5) BLSTM (hidden): bidirectional LSTM model. (6) Att-BLSTM (hidden): add attention weights to hidden layer features. (7) DABLSTM-L1: Overall sentence sentiment analysis model, which is a feature layer fusion of experiment (3) and experiment (5). (8) Att-CNN-BLSTM: Sentence sentiment recognition model based on interactive attention of CNN and BLSTM.

(9) SAOMS (Sentiment analysis of multi-grain sentences): The algorithm of multi-grain sentence sentiment analysis proposed in this paper, which is used to get the interaction sentiment probability, whole sentence sentiment probability and local sentence sentiment probability by (1), (7) and (8), and decision fusion. For sentiment word vectors, with the addition of L1 regularization, the accuracy is improved, indicating that the addition of L1 sparse sentiment word vectors has a helpful effect on the sentiment analysis of sentences. The high-level sentiment semantic features improve the classification accuracy by about 3 percentage points over the low-level word vector features, which fully indicates that compared with the high-level sentiment semantic features, the features learned from the low-level word vectors alone do not express the sentence sentiment well.

Table 3: Comparison Experiments of Different Models

NO.	MODEL	ACCURACY	RECALL	F-VALUE
1	CNN	0.8315	0.8357	0.8325
2	LSTM	0.8447	0.8494	0.8452
3	ATT-BLSTM(WORD)	0.8115	0.8207	0.8111
4	ATT-BLSTM-L1(WORD)	0.8242	0.8106	0.8290
5	BLSTM(HIDDEN)	0.8534	0.8394	0.8574
6	ATT-BLSTM(HIDDEN)	0.8579	0.8574	0.8594
7	DABLSTM-L1	0.8555	0.8754	0.8527
8	ATT-CNN-BLSTM	0.8687	0.8570	0.8715
9	SAOMS	0.8929	0.8957	0.8932

The accuracy of the multi-visual target fusion image sentiment analysis method is 0.8815 under the maximum pooling fusion strategy. By adjusting the hyper-parameters, it is observed that the accuracy of the graphic fusion method with different parameters shows that the highest accuracy is achieved when the weights of the multi-target visual sentiment probability and the multi-granularity sentence sentiment probability are both 0.5.

The experimental comparisons show that adding multi-granularity sentence sentiment analysis or multi-visual target image sentiment analysis on the basis of feature layer fusion has significantly improved the accuracy rate, indicating that either multi-granularity sentence sentiment analysis or multi-visual target image sentiment analysis can help to improve the effect of graphic fusion sentiment classification. Also consider that both multi-granularity sentence sentiment analysis and multi-visual target image sentiment analysis have higher accuracy than adding a single modality. It indicates that multi-granularity sentence sentiment analysis and multi-visual target image sentiment analysis can compensate each other to improve the sentiment classification accuracy of graphic fusion methods. The time for training of multi-visual target fusion image sentiment analysis is mainly consumed in the whole image sentiment recognition model FCNN and saliency target sentiment recognition model SFPN-CNN. Multi-visual target fusion image sentiment analysis improves about 3 percentage points compared to single modality accuracy.

5. Conclusion

This paper takes athletes' social media data as the research objective and uses deep learning methods to construct a graphic fusion sentiment analysis method based on multi-visual and multi-granularity, in order to be able to better understand athletes' emotional state and improve the ability of sports management. In this paper, through sentiment recognition for multi-granularity sentence sentiment analysis, a whole sentence-based sentiment analysis

model DABLSTM-L1 is designed, which integrates low-level word vector features and high-level semantic features. The outstanding effect of the model is verified through experiments. The main conclusions of this paper are as follows.

(1) Based on the overall sentence sentiment analysis model DABLSTM-L1, low-level word vector features and high-level semantic features are fused, and the sentence sentiment recognition model Att-CNN-BLSTM with interactive attention is constructed to extract the features of interactive information between the overall sentence and local sentences. (2) Multi-Visual Target Sentiment Analysis predicts the sentiment of images through saliency targets, face targets and whole images. Multi-Visual Target Sentiment Analysis predicts the sentiment of microblogging image modalities and fuses the predicted sentiment of both modalities at the decision level.

Reference

Basiri, M. E., Nemati, S., Abdar, M., Cambria, E., & Acharya, U. R. (2021). ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis. *Future Generation Computer Systems*, *115*, 279-294.

Chen, J., Han, P., Zhang, Y., You, T., & Zheng, P. (2023). Scheduling energy consumption-constrained workflows in heterogeneous multi-processor embedded systems. *Journal of Systems Architecture*, *142*, 102938.

Chen, J., Li, T., Zhang, Y., You, T., Lu, Y., Tiwari, P., & Kumar, N. (2023). Global-and-local attention-based reinforcement learning for cooperative behaviour control of multiple uavs. *IEEE Transactions on Vehicular Technology*.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., & Gelly, S. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. Proceedings of the IEEE conference on computer vision and pattern recognition,

Huang, M., Cao, Y., & Dong, C. (2016). Modeling rich contexts for sentiment classification with lstm. *arXiv preprint arXiv:1605.01478*.

Kamyab, M., Liu, G., & Adjeisah, M. (2021). Attention-based CNN and Bi-LSTM model based on TF-IDF and glove word embedding for sentiment analysis. *Applied Sciences*, *11*(23), 11255.

Kim, Y. (2014). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

Li, Y., & Cao, J. (2021). WSN node optimal deployment algorithm based on adaptive binary particle swarm optimization. *ASP Transactions on Internet of Things*, *1*(1), 1-8.

Raghu, M., Unterthiner, T., Kornblith, S., Zhang, C., & Dosovitskiy, A. (2021). Do vision transformers see like convolutional neural networks? *Advances in Neural Information Processing Systems*, *34*, 12116-12128.

Rehman, A. U., Malik, A. K., Raza, B., & Ali, W. (2019). A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis. *Multimedia Tools and Applications*, *78*, 26597-26613.

Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. Proceedings of the European conference on computer vision (ECCV),

Y, O., M, Y., Y, X., & H, Y. (2024). The impact of workplace marginalization on organizational citizenship behavior of property management personnel in public health emergencies. *IECE Transactions on Social Statistics and Computing*, *1*(1), 9-14.