



# Multimodal emotion recognition algorithm based on edge network emotion element compensation and data fusion

Yu Wang<sup>1,2</sup>

Received: 20 October 2018 / Accepted: 12 December 2018 / Published online: 4 January 2019  
© Springer-Verlag London Ltd., part of Springer Nature 2019

## Abstract

The data feature set of emotion recognition based on complex network has the characteristics of complex redundant information, difficult recognition and lost data, so it will cause great interference to the emotion feature of speech or image recognition. In order to solve the above problems, this paper studies the multi-modal emotion recognition algorithm based on emotion element compensation in the background of streaming media communication in edge network. Firstly, an edge streaming media network is designed to transfer the traditional server-centric transmission tasks to edge nodes. The architecture can transform complex network problems into edge nodes and user side problems. Secondly, the multi-modal parallel training is realized by using the cooperative combination of weights equalization, and the reasoning of nonlinear mapping is mapped to a better emotional data fusion relationship. Then, from the point of view of non-linearity and uncertainty of different types of emotional data samples in the training subset, emotional recognition data compensation evolves into emotional element compensation, which is convenient for qualitative analysis and optimal decision-making. Finally, the simulation results show that the proposed multi-modal emotion recognition algorithm can improve the recognition rate by 3.5%, save the average response time by 5.7% and save the average number of iterations per unit time by 1.35 times.

**Keywords** Emotion recognition · Edge network · Multimodal · Emotion compensation · Emotion recognition · Data fusion

## 1 Introduction

Emotional recognition [1, 2] involves the intersection of signal processing, in-depth learning, edge computing, psychology, and other disciplines [3, 4]. Emotional data not only have signal emotional analysis [5], but also need to consider the transmission of emotional data and deep-level feature extraction and analysis. Then, emotion recognition in complex network environment is faced with such problems as large-scale network data transmission [6, 7], emotion data caching, emotion recognition compensation and parallel recognition, and a series of research results have been achieved such as in [8–11].

About edge streaming networks, Roy et al. [12] demonstrated that a large number of graph algorithms, which could

be expressed using the edge-centric–scatter-gather model. A novel local streaming service concepts based on mobile edge computing and their business models are proposed in article [13] through explorative research, utilizing a relevant theoretical framework. The novel approach denoted as SpanEdge was proposed by the authors of [8], which unifies stream-processing across a geo-distributed infrastructure, including the central and near-the-edge data centers. The article [14] highlighted some of the potentials and prospects of edge computing for interactive media, and presented some preliminary works in the area, such as video and gaming applications.

About emotion compensation, the article [9] constructed the land expropriation compensation RDEU evolutionary game model, which could be used to study the influence of the emotions of farmers and local governments on game strategies. The research in [15] enriched the knowledge about how laypeople make compensation decisions for emotional losses. The authors of article [16] implemented the adaptive sub-layer compensation based facial emotions recognition method for human emotions recognition with various features extraction mechanism. The problems about the ethically guided emotional responses for social robots was studied and researched

✉ Yu Wang  
stonchor@163.com

<sup>1</sup> College of Computer Science, Henan University of Engineering, Zhengzhou 451191, China

<sup>2</sup> State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou 450002, China

in article [10]. The study in ref. [11] investigated the association between schizophrenia polygenic risk and brain activity while testing perception of multisensory, dynamic emotional stimuli.

About data fusion, the problem of data fusion, namely piecing together multiple datasets collected under heterogeneous conditions was studied in [17]. The authors of article [18] provided an overview of established definitions, targeting a generalized understanding of the image fusion. A methodology for estimating the MFD was proposed in article [19] based on both data sources simultaneously. An audio-visual emotion recognition system has been developed by the authors of [20] by using the fusion of both the modalities at the decision level.

Based on the existing research results, the multimodal emotion recognition algorithm was proposed, on the basis of the edge streaming network, emotion element compensation, and data fusion.

The rest of this paper is organized as follows. Section 2 presents the edge streaming network. Section 3 develops the multimodal emotion data fusion algorithm. The multimodal emotion recognition algorithms are indicated in Section 4. Additionally, Section 5 performs simulations and demonstrated the results, followed by the conclusion in Section 6.

## 2 Edge streaming network

The edge streaming media network converts the traditional server centric transfer task to the edge node. The purpose is to combine the existing edge nodes with streaming media to form a new architecture. The architecture is used to distribute streaming media data to distributed edge nodes. Users can download or forward streaming media data through the nearest edge nodes. Edge streaming media network can effectively solve the problems of congestion, low throughput, and large transmission delay in traditional streaming media transmission.

Suppose, according to the topology of streaming media, is built on an edge wireless network. The network consists of evenly distributed  $w$  edge nodes, base stations and streaming media servers. Among them, streaming media server and base station can store a large number of streaming media data. Streaming media server and base station adopt the same queue management and control mechanism. The network can provide service support for multiple concurrent streaming media servers, as shown in Figs. 1 and 2. Different streaming media servers provide different streaming media service data to improve storage space utilization and parallel efficiency. In addition, the streaming media server can select an appropriate edge subnetwork from the edge streaming media network

according to the user-specified mode. This edge subnetwork can reduce the network topology of streaming media transmission and meet the real-time requirements while ensuring the correct content of streaming media services. Due to the strong differences among streaming media servers, base stations and streaming media service contents, the delay of streaming media access on different edge subnetworks is quite different. For example, the streaming media access delay and response requirements of ETS1 in Fig. 1 are quite different at different times. This has caused a great deal of interference to the streaming media transmission scheduling and control on the edge network.

For the above scenarios, we propose a time-based edge node scheduling algorithm for streaming media transmission as follows.

### Algorithm II.A edge schedule algorithm

Input: ETS, number of edge nodes  $N_d$ , number of ETS  $N_e$

Oouput: solution

```

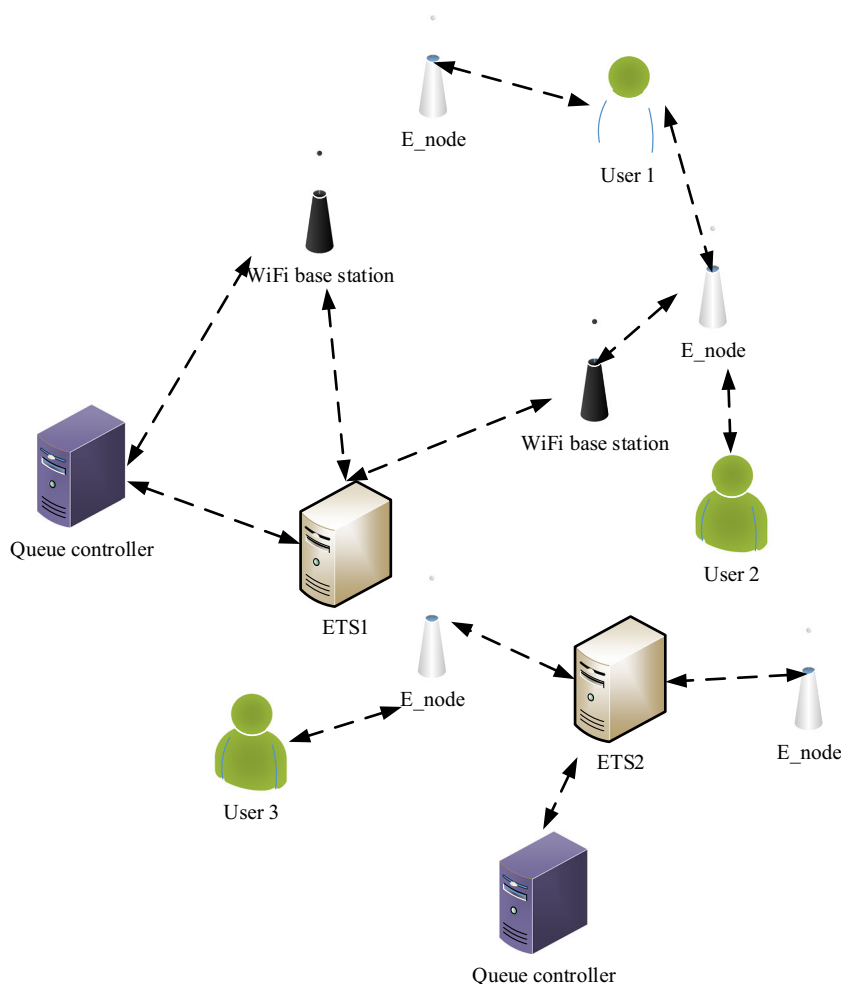
1 for i=1 to  $N_e$ 
2   sort the sequence of ETSi
3   for j=1 to  $N_d$ 
4     compute the route cost of edge node j
5   end for
6   slolution[i]←extract sub edge net from the first
   lowest cost j/i edge nodes
6 end for
7 return solution

```

The vector solution returned by the above algorithm records the edge nodes and their order contained in the streaming media transmission of each edge subnetwork. The queue controller in Fig. 1 solves the queue management problem of streaming media server and wireless hotspot base station. However, more streaming data will be cached at the edge nodes. How to extract data for emotion recognition from streaming media service content and then cache management is a key issue.

It is assumed that the edge streaming media network is partitioned into  $Q$  edge subnetworks. Among them, the three active streaming servers are ETS 1, ETS2, and ETS 3. At the same time, the three streaming servers have the same upper limit for the communication delay of  $Q$  edge subnets. The streaming medium content and emotion recognition data required by User 1 and User 2 in Fig. 1 are stored in ETS1 and 2, respectively. The mapping and constraint relationships are described in Formula (1).

**Fig. 1** Edge streaming network architecture



$$\begin{cases} \sum_{i=1}^2 User_i \cdot \left( \sum_{j=1}^q S_t(i) F(E\_node(j)) \right) \leq ST_{total} \\ F(ETS_i) \geq \sum_{j=1}^q F(E\_node(j)) \end{cases} \quad (1)$$

Among them, function  $F(\cdot)$  can calculate the current node’s emotional recognition data cache.

Assuming that User 1 and User 2 cost  $C1$  and  $C2$  respectively to obtain wireless BS and streaming media servers connected to  $ETS_j$  and  $\sum_{i=1}^2 c_i \leq c(ETS_j)$ . The mapping relationship is shown in Fig. 2. This guarantees that the cost of getting emotional recognition cached data is no higher than the total cost of  $ETS_j$  and the  $j$ -th edge subnet.

In order to further improve the buffer efficiency and reduce the transmission delay of streaming media affective data,  $ETS_j$ ,  $3 < j < q$ , which is in the waiting state, is activated. When any of the Formulas (1) is not satisfied, the remaining streaming media server is activated. The

streaming media emotional data in paired cache on the wireless base station and edge node is sent to the activated frontal streaming media server in the same edge subnet by the queue controller. At this point, the transmission delay and buffer condition of streaming media emotional data on edge subnet are like formula (2).

$$\begin{cases} d = \log(F(ETS_i)) \frac{\alpha \sum_{j=1}^q F(E\_node_j)}{\sum_{i=1}^2 User_i} t_{slot} \\ \eta_i = 1 - \sum_{i=1}^q \sqrt{\frac{\Pi_j}{j} (1-p_i^j)} \end{cases} \quad (2)$$

Among them,  $D$  represents the transmission delay of streaming media emotional data on edge subnets. At this point, we consider the amount of data transmitted by the streaming media server and the amount of streaming media cache per edge node. In order to approach the cache situation of the actual streaming media emotional data, we multiplied by a weight coefficient  $\alpha$ . The parameter  $t$ -slot

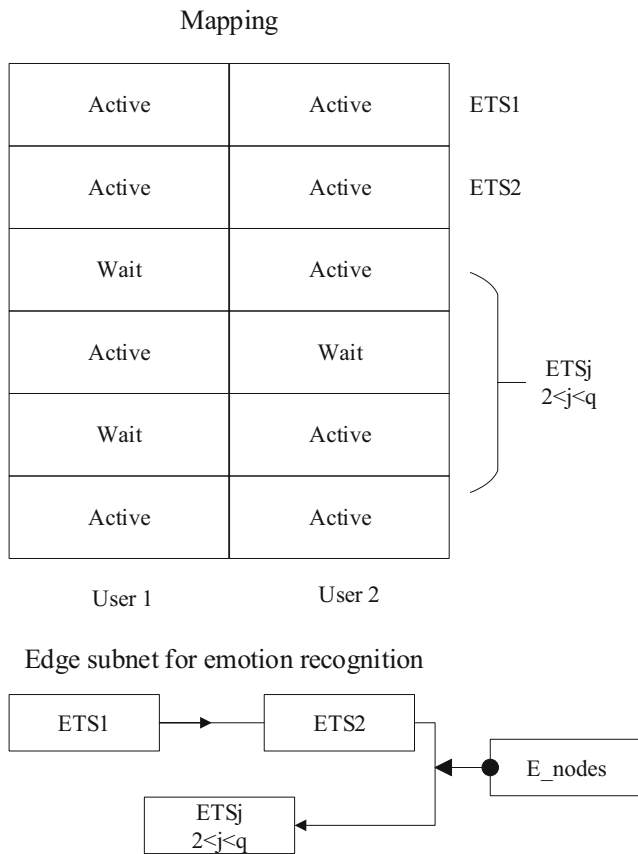


Fig. 2 Mapping between edge subnet and stream emotion recognition data

represents the time cost of the transmission of the emotional data of the unit streaming media. Parameter  $I$  represents the cache utilization of the  $I$  edge subnet. The parameter represents the cache invalidation rate of the  $j$  edge node of the  $I$  edge subnet.

However, in the process of streaming emotional recognition data transmission in edge streaming media network, the emotional recognition data is vulnerable to external topology changes and streaming media cache delay interference. If the emotion recognition data of edge nodes do not satisfy the constraints of Formula (3), the emotion recognition data will be mismatched.

$$\delta = \frac{\sum_{i=1}^n E(X_i)}{\sum_{j,i=\lfloor \sqrt{j} \rfloor} D(X_i, X_j)} \leq \bar{\eta} \tag{3}$$

Here, the parameter  $\delta$  indicates mismatch ratio. If the ratio exceeds the average cache usage, it indicates that the emotional recognition data cannot continue to be used. In Formula (3), we use the sample interpolation method, by solving the mean of the same sample, the variance of different samples to get the mismatch ratio.

At this point, we use time information based data recognition algorithm for streaming media emotion recognition.

Algorithm II.B emotion recognition data compensation algorithm for streaming media based on edge temporal information

Input: emotion recognition data  $De$ ,  $\delta$ , edge sub net number  $Nse$ , stream data sample  $X$

Oouput: emotion recognition data afer compensation  
Dec

```

1 for i=1 to Nse
2   reduce the subnet of ESNi
3   Invoke algorithm III.A
4   If length of solution is larger than Nse
5     compute  $\delta$  by using formula (3)
6     if  $\delta$  is larger than  $\bar{\eta}$ 
7       for j=1 to length of X
8         compute the E(xj)
9          $k = \lfloor \sqrt{j} \rfloor$ 
10        compute  $D(X_j, X_k)$ 
11      end for
12    end if
13  end if
14 end for
15  Dec ← resort the sequence with  $\bar{\eta} \sum_{j=1}^q F(De_j)$ ;
16 return Dec

```

### 3 Multimodal emotion data fusion algorithm

In order to effectively integrate emotional recognition data and emotional multi-modal data transmitted by edge streaming media network, this section adopts multi-core learning method to construct emotional data fusion model. The kernel function of multi-kernel learning can effectively deal with the non-linearity of emotion recognition data and the inseparability of data stream. From the previous section, we find that the streaming media emotion recognition data transmitted cooperatively between edge subnets have the following characteristics:

- (a) The discrete samples of streaming media data contain heterogeneous information of different edge subnets, which makes the emotional data redundant.
- (b) The scale of streaming media continuous sample is large and the emotion recognition data is nonlinear, which makes the dimension of emotion recognition data higher. In high-dimensional feature space, the probability distribution of streaming media emotion recognition data is dynamic and not smooth.
- (c) Compensation of time dimension for mismatched emotional recognition data is easy to lose the constraint of spatial dimension on emotional recognition data.

Based on the above characteristics, the efficiency and accuracy of emotional data processing of single modal learning function are seriously restricted. The result of emotional data is difficult to reflect the real situation of the data. Multi-modal emotion data processing, which realizes the complementary combination of multiple kernel functions, is helpful to improve the performance of emotion recognition. Converts the emotional data of the upper edge network streaming media into a set  $T = \{t_1, t_2, \dots, t_{|TF|}\}$  of training instances. Here,  $t_i = (x_i, y_i, z_i)$ . The first metadata represents the edge network streaming data sample, which satisfies the  $x_i \in X \subseteq R^n$  relationship. The second metadata represents the streaming media emotional data sample, which satisfies the  $y_i \in Y \subseteq R^n$  relationship. The third metadata represents the multimodal emotion data sample, which satisfies the  $z_i \in Z \subseteq R^n$  relationship. In addition, we defined a subset  $TF := \{(y_1, z_1), (y_2, z_2), \dots, (y_{|TF|}, z_{|TF|})\}, TF \subseteq T$ . Thus, the following nonlinear mapping exists.

$$\begin{cases} \Phi : Y \rightarrow Z \\ x \rightarrow \Phi(x) \end{cases} \quad (4)$$

$$\begin{cases} M_e = \cup_i (\Phi(x_i), z_i) \subset T \times Y \\ W = \frac{\|X-Z\|}{\|M_e\|^2} \end{cases} \quad (5)$$

The Formula (4) defines the nonlinear mapping function  $\Phi(\cdot)$  from Y to Z. The function takes the element of X as the input parameter. Formula (5) By comparing the result of Formula (4) with the sample set X and Z, the weight coefficient set W of multimodal function is obtained.

Multimodal learning method combines different kernel functions based on weight according to the training results of samples. Then, a better emotion data fusion relation is obtained through reasoning from nonlinear mapping. Because the type of multi-modal training has a great

influence on the cooperation of different kernel functions, in order to construct multi-modal emotional data model better, we consider the multi-modal weighting coefficients set functions, namely  $W := \{w_1, w_2, \dots, w_{|W|}\}$ . Thus, the edge network streaming media data feature extraction of emotional data samples is transformed into a multi-modal kernel function and weight parameter adaptive scheduling problem. Formula (6) gives a formal description of the scheduling probability for this problem.

$$P_{se}(z_i) = w_j \frac{\exp\{TF^{z_i} | ETS\}}{|\Phi| \sum M_e(i)}, j \neq i \quad (6)$$

$P_{se}(z_i)$  represents the scheduling probability of the j kernel function and the I emotion compensation data sample that is compensated.

Then, considering the difference and weight balance between the multi-modal data, the multi-modal emotional data is fused with the flow shown in Fig. 3.

First, we mismatch and compensate for the multi-mode of edge network, streaming media, and emotional data. Multimodal samples are recovered through the transmission of the edge subnet and arriving at the user's emotional data. This can maintain emotional data with good recognition characteristics.

$$Z_R(M_e, X) = \frac{\sum_{i=1}^{|W|} w_i \Phi(x_i)}{\|X\|^2 \operatorname{argmax}\{TF\}} \quad (7)$$

This function  $Z_R(M_e, X)$  is used to restore the emotion recognition features of multimodal samples.

Secondly, in order to ensure the integrity of different modal emotional data and the sufficiency of time information, the update method of modal function as shown in Table 1 is constructed in the process of multi-modal emotional data fusion.

Finally, the restored multi-modal emotion data and the cooperative update process based on weight control are fused on the edge nodes, such as Formula (8)

$$\begin{cases} F_U(w, z) = \rho \left( \log(|TF|) \sum_{i=1}^{|W|} \frac{w_i \Phi(x_i)}{\|X\|^2} \right) + \eta A_k(t) + \lambda T_M(x, y, z) \\ (\delta + \eta + \lambda) \leq 1 - P_{se} \end{cases} \quad (8)$$

Among them,  $\rho$ ,  $\eta$ , and  $\lambda$  are three weight equilibrium coefficients. Coefficient  $\rho$  is used to balance the recovery of multimodal emotion recognition features. The coefficient  $\eta$  is used to balance the updating way of each state function. Coefficient  $\lambda$  is used to balance X, Y, and Z's emotional data sample transfer process.

Algorithm III.A: Multimodal data fusion algorithm

Input:  $0 < \rho < 1, 0 < \eta < 1, 0 < \lambda < 1, X$ , set  $T$ ,  $Me$ .

Oouput: multimodal sample after training  $M_{at}$ , multimodal kernel state  $Sk$ =fusion result by using the scheme shown in figure 3.

```

1 for i=1 to |X|
2   If |Y| = 0 then
3     Invoke algorithm II.A
4     compute  $Z_R$  by using formula (7)
5     compute  $P_{se}$  by using formula (6)
6     if  $(\delta + \eta + \lambda) \leq 1 - P_{se}$  then
7       for j=1 to length of |ZR|
8          $q1 \leftarrow \log(|TF|) \sum_{i=1}^{|W|} \frac{w_i \Phi(x_i)}{\|X\|^2}$ 
9          $q2 \leftarrow A_k(t)$ 
10         $q3 \leftarrow T_M(x, y, z)$ 
11      end if
12       $F_U(w, z) \leftarrow \rho q1 + \eta q2 + \lambda q3$ 
13    end if
14  update the  $Sk$ 
15 end for
16   $Mat \leftarrow$  average of  $F_U(w, z)$ 
17 return  $Sk$ 

```

4 Multimodal emotion recognition

Based on the analysis of different types of topology and transmission behavior, such as edge networks, streaming media and emotion recognition, and the multi-modality analysis of different types of emotion recognition capabilities described in Sections 2 and 3, this section divides the entire streaming media emotional data into multiple distributed training subsets. Algorithm III.A is used to synchronize the number of samples and multi-modal characteristics of streaming media emotional data. In order to further eliminate the non-linearity and uncertainty of different types of emotional data samples in

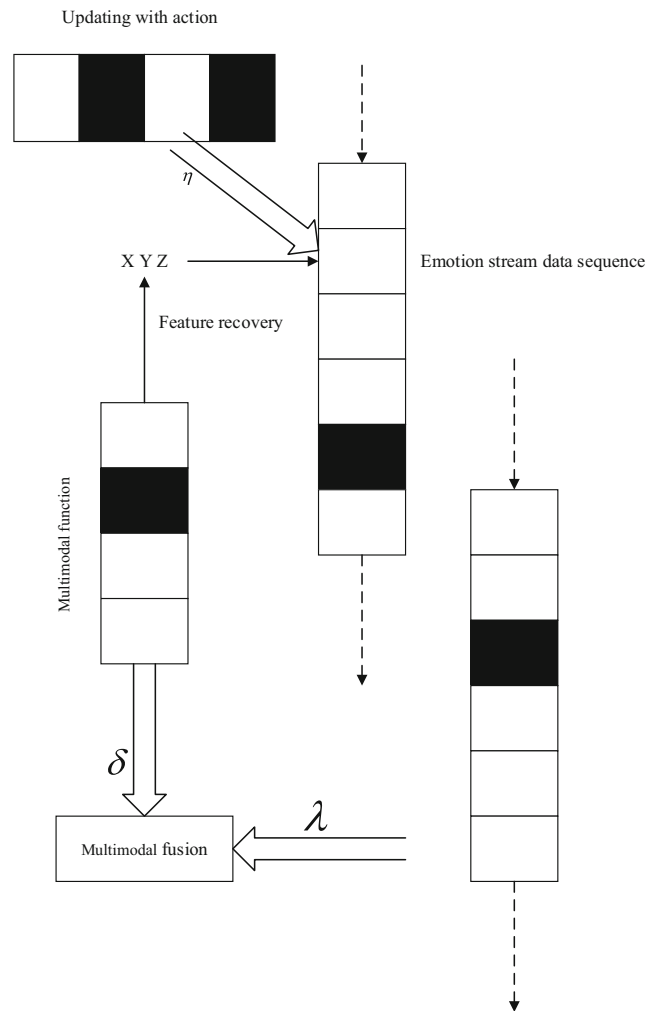


Fig. 3 Multimodal data fusion scheduling

the training subset, the emotional recognition data compensation in Section 2 is evolved into emotional element compensation. These two different modes of compensation schemes are qualitatively analyzed and optimized according to local topological features of edge subnet and streaming media service classification. Emotion element compensation can transform non-linear and uncertain emotional mismatch into qualitative model, which is helpful to solve the problem of uncertain transformation of emotional recognition from quantitative expression to qualitative description.

In addition, Sections 2 and 3 define three sets of samples of different modes: X, Y, and Z. We can qualitatively

Table 1 Updating schemes of multimodal kernel functions

Action	Function
Kernel action	$x^t = w^t(TF^t y^t + M_e^t)$
Modal action	$w^t P_{se} = x^t(TF^t z^t + M_e^t)$
Cell action	$z^t = x^t(w^t TF^t + y^t)$
Sample action	$\bar{x}^t = w^t(TF^t \bar{y}^t + \Phi(X))$

identify emotional features by solving the expected values  $Ex$ , Entropy, Covariance  $Cv$ , and Correlation Coefficient  $Cef$  of each modal. Because the multi-modal kernel functions have different dimensions for feature extraction of emotion recognition results, we improve and enhance Section 3’s Algorithm III.A. Algorithm III.A’s fusion algorithm evolves into a low-dimensional fusion of multiple partitions, which makes it easy to calculate the weights of the edge subnet and update the kernel function. This evolution helps to achieve multi-modal compensation and evolution of emotional characteristics in high dimensional space. When constructing a relational reasoning network for multi-modal emotion recognition, the expectation  $Ex$  and correlation coefficient  $Cef$  are trained by the inverse reasoning and inverse transformation of Fig. 4, which can effectively improve the accuracy and consistency of different modal emotion features. The inverse transformation of Fig. 4 also helps to transform the single modal emotion recognition data into emotional elements. The transformation can guarantee the expected solution and the determination of entropy. The entropy  $En$  will play an important

role in the iterative quantitative process of multimodal emotion recognition. This process can further reduce the dimension of the emotion recognition data set while reducing the ambiguity of the dataset.

The multimodal emotion recognition algorithm proposed in this paper can be divided into three stages.

Stage 1. Digital signature calculation. Formulas (9)–(12) were used to calculate the expected values  $Ex$ , entropy  $En$ , covariance  $Cv$ , and correlation coefficient  $Cef$  of all emotional data in sequence.

$$Ex = \frac{1}{|X|} \sum_{i=1}^{\min\{Y|Z\}} x_i \tag{9}$$

$$En = \sqrt{\frac{\pi}{2}} \times \sum_{i=1}^{\min\{Y|Z\}} |x_i - Ex| \tag{10}$$

$$Cv = \frac{1}{|X|} \left( \sum_{i=1}^{\min\{Y|Z\}} x_i - \sum_{i=1}^{\min\{Y|Z\}} x_i \sum_{i=1}^{\min\{X|Z\}} y_i \right) \tag{11}$$

$$Cet = \frac{|X|Cv}{\sum_{i=1}^{\min\{Y|Z\}} x_i \sum_{i=1}^{\min\{X|Z\}} y_i} \tag{12}$$

Stage 2. Multimodal data dimensionality reduction. In order to describe the inverse transformation process of multi-modal streaming emotional data in different edge subnets, the function  $Mf(y, z)$  is used to reduce the dimension adaptively. Adaptive scheduling is controlled by the inverse reasoning and digital characteristics of the weight  $W$  vector (which can be obtained from Formula 5) for each mode, as shown in Formula 13.

$$\begin{cases} w_i = \frac{\tau}{1 - P_{se}(i)} \cdot \sqrt{\exp\left(\frac{Cet}{En + Ex + Cv}\right)} \\ \tau = \sqrt{Ex(\|Z\|^2 - \|W\|^2)} \end{cases} \tag{13}$$

Coefficients can map the weights of emotional data feature extraction after inverse reasoning to a controllable range, so as

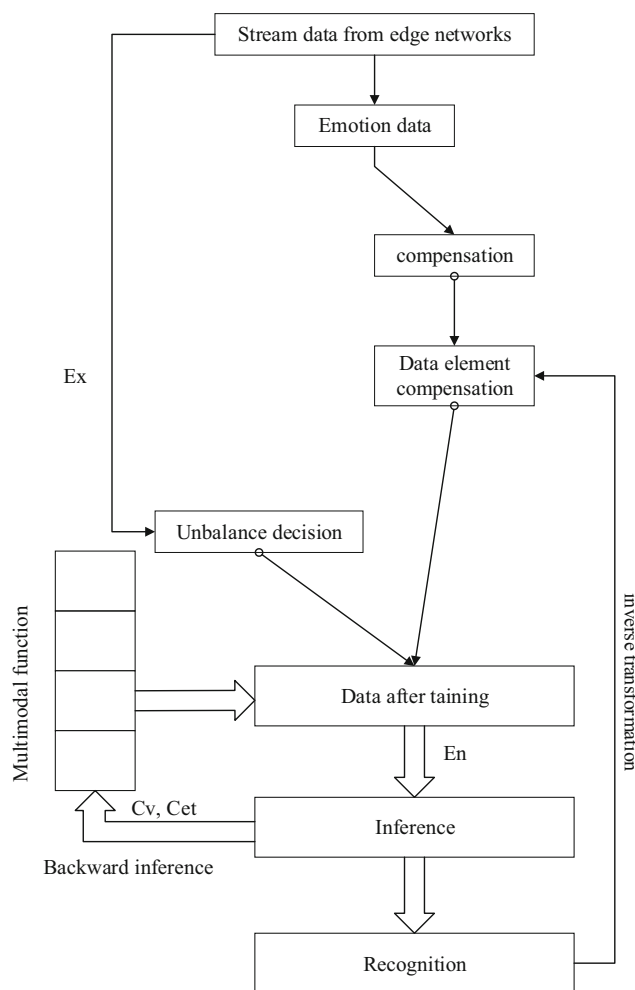


Fig. 4 Backward inference and inverse transformation training

Table 2 Recognition rates of MEC-EECF on voice database (%)

Emotion	Neutral	Scared	Happy	Annoying	Angry
Neutral	70.02				
Scared		80.78			1.78
Happy			92.13		
Annoying	2.56			79.12	10.56
Angry				10.34	82.09

**Table 3** Recognition rates of MEC-EECF on image database (%)

Emotion	Neutral	Scared	Happy	Annoying	Angry
Neutral	<b>80.02</b>				
Scared		<b>82.57</b>			1.28
Happy			<b>93.41</b>		
Annoying	3.46			<b>77.32</b>	9.75
Angry				8.34	<b>88.25</b>

to avoid high-dimensional errors in expectations, entropy, covariance and correlation coefficients, and training samples of emotional data.

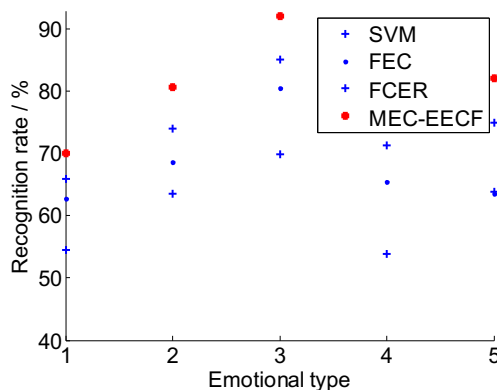
Stage 3. Multimodal emotion recognition. On the basis of stages 1 and 2, multi-modal emotion recognition is carried out by combining multi-dimensional digital features with multi-modal emotion data fusion using edge streaming media network as data source.

The implementation of the above three stages of multimodal emotion recognition algorithm is divided into two different states:

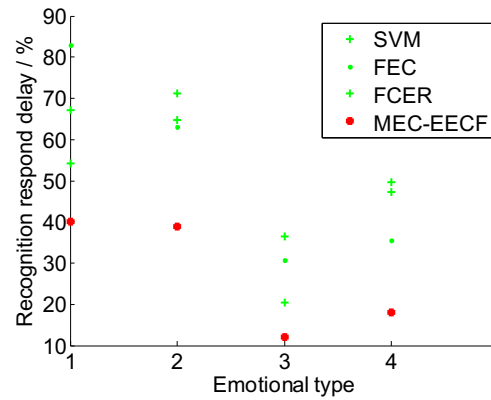
- (a) emotion\_compensation. In this state, the algorithm takes the algorithm of section III as the core.
- (b) emotion\_element\_compensation. If the probability Pse is biased and the Algorithm III.A can not be executed properly, the emotion recognition algorithm is converted to the emotion element compensation state.

### 5 Performance analysis

To verify the validity of multi-modal data fusion and compensation mechanism in emotion recognition, a network topology with 50 streaming media servers and 200 edge nodes is



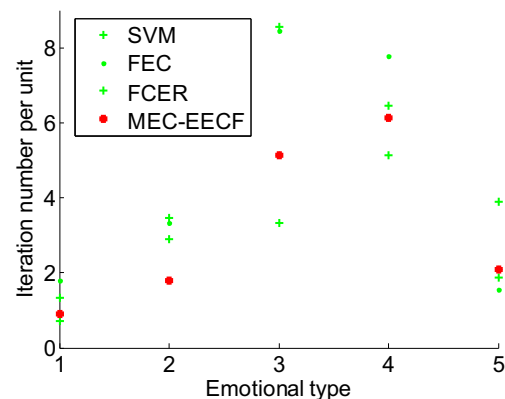
**Fig. 5** Recognition rate



**Fig. 6** Recognition response delay ratio

deployed in the campus. The experiment carried out 30 days data acquisition and transmission. The emotion database cached on the cloud server contains 656 voices, which can be divided into five different emotions: neutral, fear, joy, annoyance, and anger. In addition, the database has 1378 emotional images, which can be divided into six emotions—neutral, fear, joy, annoyance, anger, and boredom. SVM, fusion-based emotion recognition (FER), fusion and compensation-based emotion recognition (FCER) and MEC-EECF strategy are used in the experiment. Weighted accuracy is used in evaluation criteria. Multimodal classifier adopts softmax training times more than 200 times. The evaluation index includes recognition rate, recognition response delay, iteration number, and so on.

- (1) Recognition rate. The ratio of correct recognition number to total number of edge nodes transferred to the user’s emotional data.
- (2) Identify response delay ratio. Emotional recognition takes up the ratio of the time it takes to send from the streaming media server, transmit the emotional data to the user within the fine edge of the network, and identify it correctly.
- (3) Number of iterations per unit. Run the emotion recognition algorithm on the streaming media server, edge node



**Fig. 7** Iteration number per unit



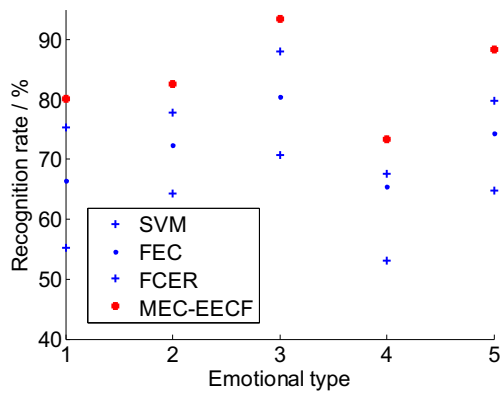


Fig. 8 Recognition rate

and client, and complete the correct recognition twice in succession as an iteration. The number of such operations completed in unit time.

In the first group of experiments, 656 speech data sets were used for emotion recognition. The recognition rate of the different emotions of the proposed MEC-EECF strategy is shown in Table 2.

Table 2 shows that the recognition rate of the five emotions is higher, especially the recognition rate of emotional “happiness” is significantly higher than other types of emotions, and the recognition rate here is 92.13%. From the point of view of misjudgment rate, “neutral,” “anger,” and “hate” have certain misjudgment rate, the remaining misjudgment rate is 0. The experimental results of Table 2 verify the effectiveness and feasibility of the proposed MEC-EECF method for emotional recognition in speech database.

Second groups of experiments, 1378 emotional pictures as data sets, emotion recognition. The recognition rate of the different emotions of the proposed MEC-EECF strategy is shown in Table 3.

Table 3 shows that the proposed MEC-EECF can identify the image emotion database very well. In addition, the proposed MEC-EECF performs better on the basis of image database.

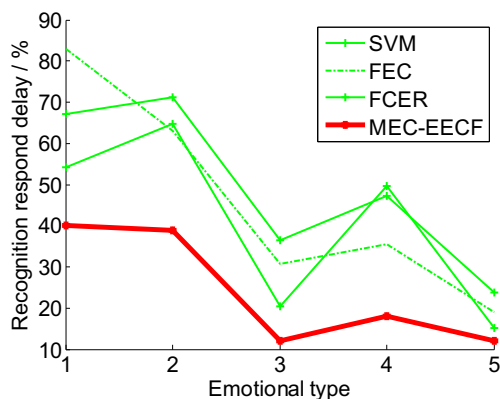


Fig. 9 Recognition respond delay ratio

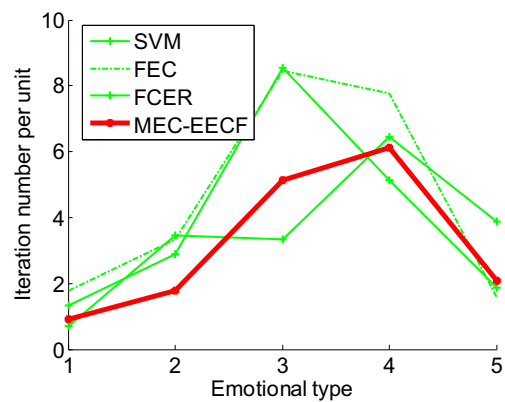


Fig. 10 Iteration number per unit

In Figs. 5, 6, 7, 8, 9, and 10, one in abscissa is neutral, two is afraid, three is joy, four is annoyance, and five is anger. Compared with Figs. 5 and 7, the proposed MEC-EECF performs best in recognition rate, which is 3.5% higher than other strategies on average. The results show that MEC-EECF can give full play to the advantages of edge node transmission, send emotion recognition data to the client through streaming media, and improve the correct recognition number.

Compared with Figs. 6 and 9, the proposed MEC-EECF has the lowest recognition response time occupancy rate, which saves an average of 5.7% compared with other strategies. It is shown that MEC-EECF can compensate for the mismatch of edge network, streaming media, and multi-modality of emotional data. It can give full play to the multi-modal collaboration advantage and effectively reduce the time cost of emotion recognition. Compared with Figs. 7 and 10, the proposed MEC-EECF runs in parallel in the number of unit iterations. In the process of two consecutive correct multi-modal emotion recognition, digital feature extraction and updating, and dimensionality reduction are used to reduce the requirement of iterative operation per unit time. The average number of iterations per unit of the proposed MEC-EECF is close to that of the best performing FCER, which saves an average of 1.35 iterations compared with the other two strategies.

## 6 Results and conclusion

Emotional recognition data transmitted through complex networks are easy to mix redundant information with fuzzy features, which increases the difficulty of emotion recognition and recognition response speed. In order to eliminate or reduce the complexity of emotional features in speech or image recognition, a high recognition rate and real-time emotion recognition algorithm based on emotion element compensation and multi-modal data fusion is proposed in edge streaming media network architecture. On the one hand, the streaming media data service is distributed to the edge node to

achieve the consistent emotional data cache between the edge node and the client. On the other hand, based on the sub-network partition of edge streaming media network, the multi-modal parallel training and reasoning are implemented by updating the balance weight of multi-modal. This ensures that the reasoning of nonlinear mapping has better emotional data fusion relationship. Especially, emotional recognition data compensation is evolved into emotional element compensation, which weakens the non-linear and uncertain characteristics of different types of emotional data samples and transforms them into qualitative analysis and optimal decision-making. Finally, several different experiments are carried out from the speech database and the image database. The statistical results fully verify the feasibility and effectiveness of the proposed algorithm. For example, the recognition rate is increased by 3.5%, the recognition response time is saved by 5.7% and the average number of iterations per unit time is saved by 1.35 times.

**Funding information** This work is supported in part by the Key scientific research projects of Henan Province Education Department (No.18A520004), Henan Province Science and Technology projects (No. 182102310925), and National Natural Science Foundation of China (No. 61802115).

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

- Hossain MS, Muhammad G, Alhamid MF et al (2016) Audio-visual emotion recognition using big data towards 5G[J]. *Mobile Netw Appl* 21(5):1–11
- Fan Y, Lu X, Li D et al (2016) Video-based emotion recognition using CNN-RNN and C3D hybrid networks[C]// *ACM international conference on multimodal interaction*. ACM:445–450
- Dhall A, Goecke R, Joshi J, Emoti W et al (2016) Video and group-level emotion recognition challenges[C]// *ACM international conference on multimodal interaction*. ACM 2016:427–432
- Zheng WL, Lu BL (2017) Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks [J]. *IEEE Trans Auton Ment Dev* 7(3):162–175
- Zhao M, Adib F, Katabi D (2017) Emotion recognition using wireless signals[C]// *international conference on mobile computing and networking*. ACM:95–108
- Papageorgiou A, Cheng B, Kovacs E (2016) Real-time data reduction at the network edge of internet-of-things systems[C]// *International Conference on Network and Service Management*. IEEE:284–291
- Kunz J, Becker C, Jamshidy M, et al. (2016) OpenEdge: a dynamic and secure open service edge network[C]// *Network operations and management symposium*. IEEE:257–264. Edge streaming
- Sajjad HP, Danniswara K, Al-Shishtawy A et al (2016) SpanEdge: towards unifying stream processing over central and near-the-edge data centers[C]// *edge computing*. IEEE:168–178
- Hong K R, Sun D, Zhao X, et al. (2017) The influence of Participant's emotion on land expropriation compensation strategy—based on RDEU evolutionary game theory[J]. *Modern Finance and Economics-Journal of Tianjin University of Finance and Economics* (9):40–51
- Ojha S, Williams MA (2016) Ethically-guided emotional responses for social robots: should I be angry?[C]// *International Conference on Social Robotics*. Springer International Publishing, Switzerland, pp 233–242
- Dzafic I, Burianová H, Periyasamy S et al (2018) Association between schizophrenia polygenic risk and neural correlates of emotion perception. *Psychiatry Res Neuroimaging* 276:33–40 Data fusion
- Roy A, Mihailovic I, Zwaenepoel W (2013) X-stream:edge-centric graph processing using streaming partitions[C]// *twenty-fourth ACM symposium on operating systems principles*. ACM:472–488
- Mäkinen O (2016) Streaming at the edge: local service concepts utilizing Mobile edge computing[C]// *international conference on next generation mobile applications, services and technologies*. IEEE:1–6
- Bilal K, Erbad A (2017) Edge computing for interactive media and video streaming[C]// *International Conference on Fog & Mobile Edge Computing*. IEEE:68–73. Emotion compensation
- Zhang S, Hsee CK, Yu X (2018) Small economic losses lower total compensation for victims of emotional losses [J]. *Organ Behav Hum Decis Processes* 144:1–10
- Kulkarni P U, Bharate V D, Chaudhari D S (2016) Human emotions recognition using adaptive sublayer compensation and various feature extraction mechanisms[C]// *international conference on wireless communications, signal processing and networking*. IEEE
- Bareinboim E, Pearl J (2016) Causal inference and the data-fusion problem [J]. *Proc Natl Acad Sci U S A* 113(27):7345–7352
- Schmitt M, Zhu XX (2016) Data fusion and remote sensing: an ever-growing relationship [J]. *IEEE Geosci Remote Sens Mag* 4(4):6–23
- Ambühl L, Menendez M (2016) Data fusion algorithm for macroscopic fundamental diagram estimation [J]. *Transp Res Part C Emerg Technol* 71:184–197
- Sahoo S, Routray A (2017) Emotion recognition from audio-visual data using rule based decision level fusion[C]// *Technology Symposium*. IEEE:7–12