



# Emotional computing based on cross-modal fusion and edge network data incentive

Lei Ma<sup>1</sup> · Feng Ju<sup>1</sup> · Jing Wan<sup>2</sup> · Xiaoyan Shen<sup>1</sup>

Received: 31 January 2019 / Accepted: 4 May 2019 / Published online: 17 May 2019  
© Springer-Verlag London Ltd., part of Springer Nature 2019

## Abstract

In large-scale emotional events and complex emotional recognition applications, how to improve the recognition accuracy, computing efficiency, and user experience quality becomes the first problem to be solved. Aiming at the above problems, this paper proposes an emotional computing algorithm based on cross-modal fusion and edge network data incentive. In order to improve the efficiency of emotional data collection and the accuracy of emotional recognition, deep cross-modal fusion can capture the semantic deviation between multi-modal and design fusion methods through non-linear cross-layer mapping. In this paper, a deep fusion cross-modal data fusion method is designed. In order to improve the computational efficiency and user experience quality, a data incentive algorithm for edge network is designed in this paper, based on the overlapping delay gaps and incentive weights of large data collection and error detection. Finally, the edge network is mapped to a finite data set space from the set of emotional data elements inspired by heterogeneous emotional events. In this space, all emotional events and emotional data elements are balanced. In this paper, an emotional computing algorithm based on cross-modal data fusion is designed. The results of simulation experiments and theoretical analysis show that the proposed algorithm is superior to the edge network data incentive algorithm and the cross-modal data fusion algorithm in recognition accuracy, complex emotion recognition efficiency, and computation efficiency and delay.

**Keywords** Emotional computing · Cross-modal fusion · Edge network · Data incentive · Emotional recognition

## 1 Introduction

Intelligent terminals and centralized servers are equipped with software and hardware devices [1, 2] that can observe, understand, and generate various emotional features and have the

ability of emotional computing [3]. By designing an emotional computing algorithm, we can create a system capable of edge perception, real-time recognition, and intelligent understanding [4] of emotions. The system can quickly make accurate, reliable, and friendly responses to dynamic and random emotional events. At present, the research [5] of emotional computing mainly includes emotional mechanism reasoning, emotional data collection, emotional recognition, emotional network modeling, and human-computer communication. However, the research on emotional computing mainly focuses on the acquisition and forwarding of emotional signals, ignoring the networked management and data incentive of emotional terminal devices.

About cross-modal fusion, the authors of ref. [6] proposed a cross-modal correlation learning approach with multi-grained fusion by hierarchical network. Reference [7] improved the performance of recognizing objects by developing a cross-modal attentional context learning framework. The authors of ref. [8] described an alternative method to perform high-level multi-modal fusion that leverages cross-modal

---

✉ Xiaoyan Shen  
mlmyhero@163.com

Lei Ma  
malei@ntu.edu.cn

Feng Ju  
2931613077@qq.com

Jing Wan  
dupihedubao@163.com

<sup>1</sup> School of Information Science and Technology, Nantong University, Nantong 226019, Jiangsu, China

<sup>2</sup> Nantong Rail Transit Co. LTD, Nantong 226019, Jiangsu, China

translation by means of symmetrical encoders cast into a bi-directional deep neural network.

About edge network data incentive, the article [9] studied the coherent optics in topological insulator surface states with broken time-reversal symmetry and develop a theory for the dynamical Hall effect driven by an intense electromagnetic field. The authors of ref. [10] proposed a novel approximation of the participation rate estimator that can significantly improve the tractability and scalability of the resulting mixed integer optimization model. The article [11] incorporated the consideration of data quality into the design of incentive mechanism for crowdsensing and proposed to pay the participants.

About emotional computing, a multi-layer affective decision model was proposed by establishing mapping relation among character, mood, and motion in article [3]. The study of article [5] established an emotional design tutoring system and investigated whether this system influences user interaction satisfaction and elevates learning motivation. The model called the 3D Sphere Wave Database Computing Model proposed in article [12] could search for designated data precisely. The article [13] proposed a technique for transferring knowledge from heterogeneous external sources, including image and textual data, to facilitate three related tasks in understanding video emotion, etc.

However, the above results ignored the cross-modal fusion in large-scale edge network data excitation and emotional computing. Therefore, based on the above research studies, we proposed the emotional computing algorithms based on cross-modal fusion and edge network data excitation.

The rest of this paper is organized as follows. Section 2 develops the cross-modal fusion network model. Section 3 shows the data incentive algorithms for edge networks. Section 4 indicates the emotional computing algorithms for cross-modal data fusion. Section 5 perform the simulation analysis, followed by the conclusion in Section 6.

## 2 Cross-modal fusion network model

Multi-modal data is very common in large data collection, processing, and analysis applications based on Internet of Things and cloud computing platforms. Examples include the following: different languages and forms of social network news description and its promotion; multi-visual feature extraction and representation in multimedia form [14]; dynamic pictures, diversified texts, and dynamic labels used in human-computer interaction applications to describe. In the deep application of multi-domain, multi-modal fusion can describe and process the same data from different angles and aspects. However, the effective analysis of large data and network fusion shows that the multi-modal network fusion scheme has some shortcomings, such as the inability to detect data vulnerabilities and

the difficulty to accurately capture and optimize the large data characteristics of heterogeneous network sources [15]. Therefore, based on the shortcomings of multi-modal data fusion analysis, this paper adopts a cross-modal approach to in-depth fusion, combined with network feature fusion [16, 17], and carries out data content fusion, large data cross-modal retrieval, and deep mining research.

Deep cross-modal fusion can capture the semantic deviation and design fusion between multi-modals through non-linear cross-layer mapping. In this section, a deep fusion method of cross-modal data fusion is designed. Combining the deep mining network and the cross-modal data fusion method, the fusion semantics of each cross-modal and large data space can be learned in depth, and a cross-modal network model can be constructed.

Suppose that the cross-modal data set  $X = \{X_{i,j}^k\} = \{X_i, X_j\}_L^k$  contains  $L$  data modes and  $i$  data instances belonging to  $j$  data subsets,  $X_L^k \in Z^C$  represents the  $k$ th cross-modal eigenvalue matrix of the  $L$ -level modal for a subset of  $C$  complete modal data samples.  $X_i^{k(L)} \in Z^C$  represents the sample of the first subset of data in the  $k$ th  $L$ -level cross-modal.

Therefore, the data samples contained in the cross-modal fusion have three constraints: cross-layer, mode, and fusion degree. Compared with the multi-modal fusion scheme, each cross-modal data sample has the characteristics of multi-dimensional attribute reconstruction and feature fusion. Because cross-modal data can define and store different dimension features of data samples from the same data set, cross-modal fusion heterogeneous data sets are deeply fused based on semantic bias in network fusion.

In network fusion and deep cross-modal analysis, based on the heterogeneous semantic mapping of feature matrix decomposition and data set clustering, the deep heterogeneous semantic features and cross-layer mapping matrices between cross-modal data clustering are deeply excavated.

As shown in Fig. 1, cross-modal deep network fusion maps each cross-modal data set  $X = \{X_i, X_j\}_L^k$  to the network data sample set in depth and maps  $X$  into multi-dimensional feature normalization processing by active learning, which is expressed as  $DN_i^k = \{DN_C, X_C\}$ .

Based on depth normalization, the cross-modal fusion scheme uses the non-negative data set fusion model of distributed data clustering, based on network fusion to represent each cross-modal data set, and converts it into single-modal matrix and normalization  $N_i^k = \{N_C, X_C^N\}_j$  based on linear matrix decomposition.

Through network fusion, cross-modal deep fusion, heterogeneous class matrix transformation, and cross-modal normalized matrix decomposition, the modal semantic features of complex network models with obvious semantic deviation

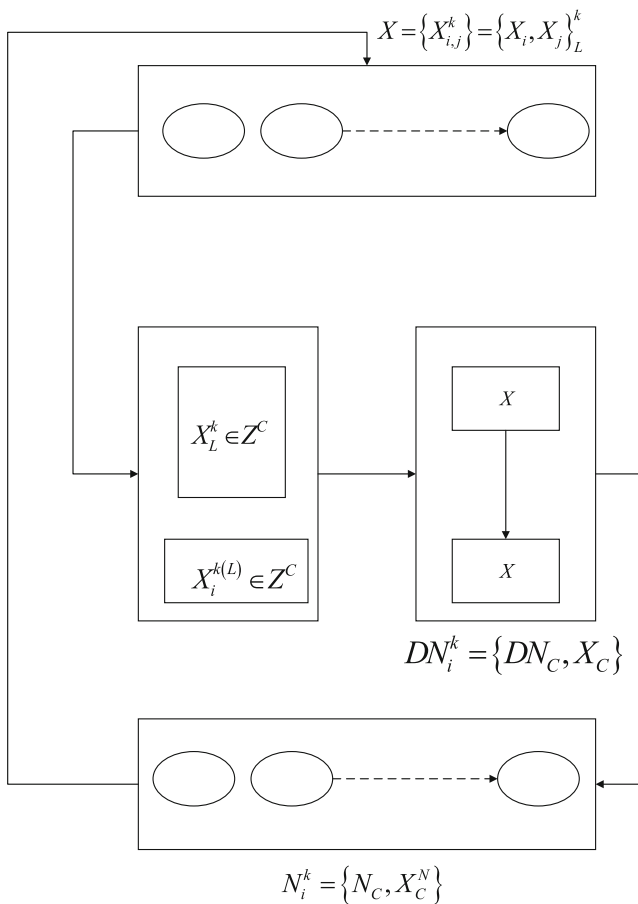


Fig. 1 Heterogeneous semantic features and cross-layer mapping

are simplified into semantically transparent cross-modal network fusion semantics.

The heterogeneous semantic features shown in Fig. 1 refer to the semantic differences between the upper and lower heterogeneous elliptic data samples. Cross-layer mapping refers to the normalization of heterogeneous semantic features through multi-layer semantics and cross-modal fusion and the consistent results.

It is assumed that there exists a shared semantic space for cross-modal fusion of random dimensions in cross-level large data sample retrieval and visualization processing. At this time, the query results and visualization processing results of large data samples with different dimensions and different layers of heterogeneous semantic mapping relations can be obtained from their respective cross-modal heterogeneous semantic feature spaces to map low-dimensional subspaces in the shared cross-modal fusion semantic space. Thus, the constructor of mapping low-dimensional subspaces can be expressed as

$$\begin{cases} f(X) = F_S \cdot W_S \\ g(N) = W_S \cdot DN^d \end{cases} \quad (1)$$

Here,  $d$  is the dimension of shared cross-modal fusion semantic space, while  $F_S$  and  $W_S$  map the retrieval result matrix and visualization result of large data samples of heterogeneous semantic mapping relationship to the weight matrix of low-dimensional subspace, respectively.

In order to eliminate the dimensionality randomness of cross-modal fusion semantic space in mapping low-dimensional subspace, we use Eq. (2) to obtain normalized semantic correlation between cross-modal fusion retrieval sequence and visualization sequence of large data samples expressed in low-dimensional subspace. Based on the semantic correlation obtained by Eq. (2), we design a shared space that can minimize the difference between cross-modal fusion semantics and low-dimensional subspace in the emotional computation of data-driven edge networks. The description of the optimization objective is detailed in Eq. (3).

$$\begin{cases} R_S^{i,j} = \sum_{i,j=1}^d \|x_i F_S - x_{i,j} W_S\|^2 \\ N_S = \sum_{k=1}^d R_S^k \|R_S - W_S\|^2 \end{cases} \quad (2)$$

The correlation matrix representing the cross-modal semantic space of the first  $j$ -level is presented as  $R_S^{i,j}$ .  $N_S$  represents the normalized matrix of  $F_S$  semantic space based on  $R_S$  correlation.

$$\begin{aligned} \min_{R_S, N_S} & |F_S|(DN \cdot W_S - FS)^T \\ s.t. & DN^T N = 1 \end{aligned} \quad (3)$$

In order to achieve the optimization goal of Eq. (3), the number of data incentives computed by cross-modal edge network data incentives can be used as the normalization basis of matrix and visual semantic consistency for large data fusion samples. In the emotional computing of edge networks, most of the retrieval bases inspired by emotional data are iterated in the form of visualization. The edge network can traverse all edge network data samples before collecting emotional data. Therefore, emotional computing in edge networks tends to be based on the correlation between data incentive criteria and their sample semantic retrieval criteria. Therefore, the mapping of emotional data in low-dimensional semantic space of edge network can be used as the basis of emotional data motivation and consistency of visual processing. Without losing generality, the higher the number of data incentives in the edge network, the smaller the consistency between the computational complexity and visual processing of the corresponding large data emotion in the shared cross-modal fusion semantic space. Therefore, from the semantic space of low-dimensional cross-modal mapping, algorithm II.A,

an emotional computing optimization algorithm based on visual processing design of edge network data, can not only achieve the goal shown in Eq. (3) but also give full play to the advantages of cross-modal fusion.

#### Algorithm II.A

```

1 Input:  $F_S, W_S, L$ 
2 for  $i=1$  to  $|F_S|$ 
3   for  $j=1$  to  $|W_S|$ 
4      $X \leftarrow \{X_i, X_j\}_L^k$ 
5     if  $|X| \leq L$ 
6        $DN_i^k \leftarrow \{DN_C, X_C\}$ 
7        $L \leftarrow L - |DN_i^k|$ 
8     end if
9      $R_S^{i,j} \leftarrow$  invoking the formula (1) and (2)
10    end for
11  end for
12 Obtaining Heterogeneous Semantic Features based on Figure 1
13 while  $DN^T N = 1$  do
14    $F_S \leftarrow (DN \cdot W_S - F_S)$ 
15 end while
16 Output:  $|F_S| (DN \cdot W_S - F_S)^T$ 

```

### 3 Data incentive algorithms for edge networks

In the cross-modal fusion model, based on the cross-modal non-linear relationship of distributed cross-layer aggregation, each edge terminal of the edge network updates the data they collect and their sample space in depth. Then, the time complexity of feature extraction of different edge terminals in low-dimensional space is optimized, and the time complexity of centralized service cluster and the delay of data collection are reduced by cross-modal fusion with centralized service cluster nodes in edge network.

However, in data incentive of edge network, there are several challenges that need to be solved urgently.

- (1) There are overlapping delay gaps in large data collection and error detection. These gaps make the determination

of the incentive weight of large data linear correlation. If the network topology of centralized cluster service nodes changes dynamically, according to the linear correlation of the weights of different large data incentives, the special edge terminals must be reselected for the large data incentive edge network requirements. These terminals must join a centralized server cluster and establish a robust and reliable clustering topology with centralized service nodes of dynamic topology. In the distributed cross-modal aggregation model, the reduction of edge terminals and the unbalanced weight of large data incentives lead to the deviation of each emotional data and a certain number of emotional device entities in the edge network.

- (2) There is a certain contradiction between edge network data processing and incentive weight. These contradictions arise from the unbalanced characteristics of data collection costs and user needs in edge networks. In cross-modal convergent edge networks, there is a certain time interval between the selection of edge terminals and the allocation of centralized server cluster service nodes, and there is a large data incentive imbalance space of cross-modal convergence. The large data query in these unbalanced spaces and the determination of incentive weights become more complex work.
- (3) The choice of emotional data is the most important basis for emotional computing. However, how to determine the source of emotional data, whether it is consistent with large data incentive edge terminals, and whether the weight basis of data incentive is unique, has become a difficult problem in emotional computing.

In order to solve the first problem, i.e., delay gap overlap, after selecting the edge terminal node, we first stimulate the cross-modal fusion of the edge network, which is the most relevant centralized server cluster node entity-related large data information. Then, the edge network updates the data incentive matrix of the terminal nodes of the edge network based on the emotional computing needs. In the process of updating incentive matrix, it is necessary to construct a dynamic topology robustness guarantee strategy for centralized server cluster. In the edge network of cross-modal fusion, the spanning tree is obtained with the centralized server cluster node as the starting point and the minimum forwarding node as the goal. The spanning tree can well describe the overlapping delay gap between large data collection and error detection. In the process of discovering delay gaps, emotional computing needs to be consistent with the linear relationship of edge network data incentives and to be responsible for the generation and distribution of data incentive weights. The spanning tree of delayed gap control in the cross-modal fusion edge network is shown in Fig. 2. The circle in Fig. 2 represents large emotional data, and the arrow represents the delayed

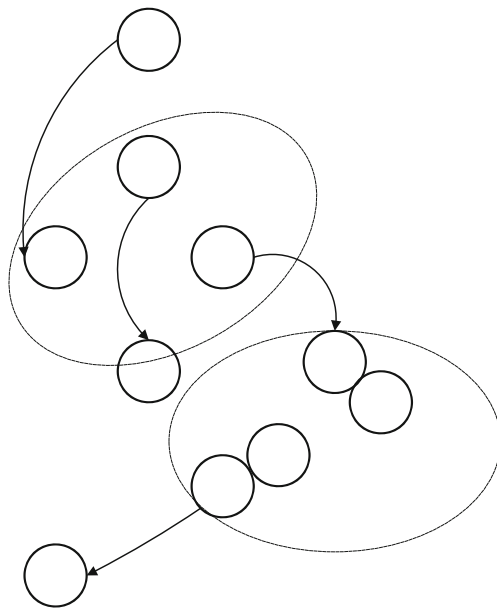


Fig. 2 Spanning tree for delayed gap control

direction of error detection. The dotted ellipse represents the centralized server clustering emotional data sample space.

At the same time, the spanning tree structure shown in Fig. 2 can quickly find the emotional data needed in the emotional computing task in heterogeneous semantic space. Figure 3 shows the cross-layer logic of the spanning tree shown in Fig. 2. The logical relationship LR1

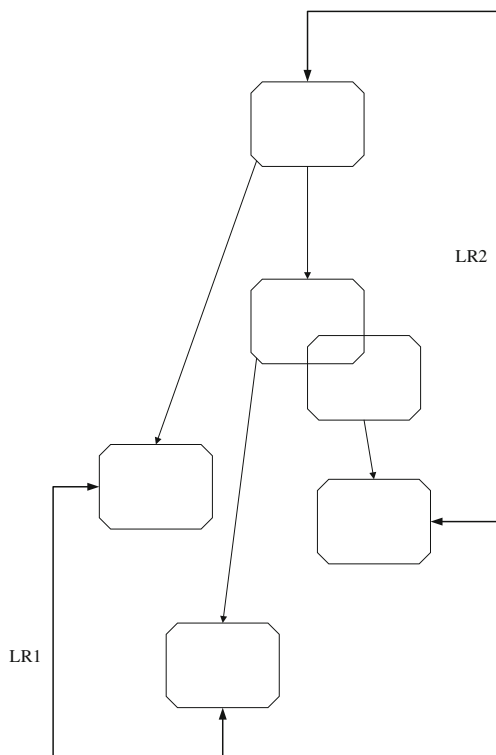


Fig. 3 Cross-layer logical relations of spanning trees

requires data incentives from two tiers of edge network terminals and centralized server clusters. Therefore, in order to weaken the overlapping delay gap and its logical relationship LR1, the spanning tree can be updated by the logical relationship LR2, and the mapping subspace of LR2 can be found in cross-modal fusion. Of course, in order to achieve rapid weakening, the distributed edge network topology and centralized server cluster relationship are further coordinated according to the large data incentive weight matrix.

After solving the first challenge, we design the edge network data incentive algorithm. The algorithm aims at solving the second and third challenging problems. Before the design of the algorithm, the data structure and the incentive weight of each emotional data in the edge network are given. For each edge network sentiment data in the cross-modal fusion space, they have their own unique number, data structure, and cross-modal attributes. In addition, each emotional data records emotional events and their data processing intervals. Aiming at the contradiction between edge network data processing and incentive weight, based on the analysis results of contradiction sources, the structure of emotional data is optimized. In the optimization, both the cost of data collection and user needs of edge network are considered. Starting from the data space of emotional events and their preservation and completion events, according to the contradiction imbalance, mapping factors  $Q_1, Q_2, \dots, Q_n$  are established for each data structure of emotional data, where  $n$  is the rank of the corresponding edge network data incentive matrices for cross-modal fusion and cross-layer data incentives. The mapping factor is the weight of the weight matrix to coordinate the edge network data incentives. Mapping factors help to update data incentive weights and discover shared attributes across modal fusion subspaces. These shared attributes can improve the efficiency of emotional data selection and effectively determine the source of emotional data, while maintaining consistency with large data incentive edge terminals.

As the objects of Figs. 2 and 3, the computation of safety data and the corresponding emotional events and the time interval of data incentive in the edge network of the cross-modal fusion model can update the mapping relationship of emotional events, the extraction of emotional features, and the input of large data incentives in real time. Each emotional event producer of the edge network uses vector occurrence and affiliate to store the information of the current emotional event and affiliated emotional event in the spanning tree logic relationship shown in Fig. 3 and uses positive real DPH to record the number of branches from the edge network terminal node of the current emotional event to the leaf node of the spanning tree (Table 1).

**Table 1** Symbols and their definitions

Label	Definition
Occurrence	Information of current emotional events in spanning tree logical relations
Affiliate	Information on affiliated emotional events
DPH	The number of branches from the edge network terminal node of the current emotional event to the generating leaf node

**Algorithm III.A**

```

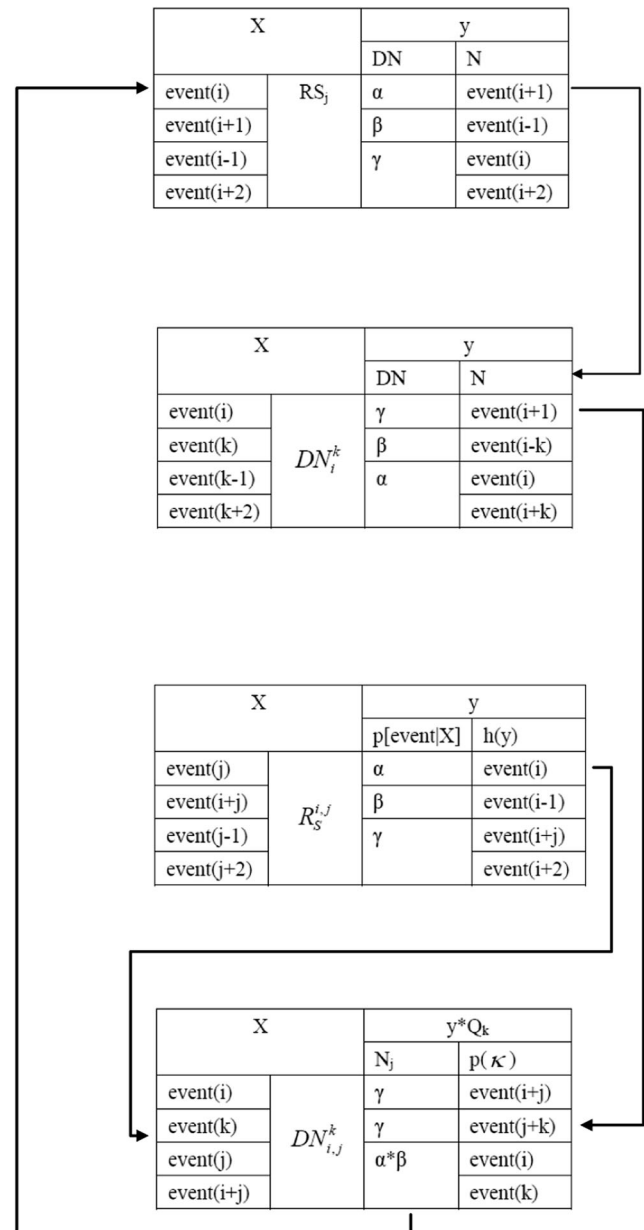
1 Input:  $Q_1, Q_2, \dots, Q_n$ 
2 for  $p=1$  to  $n$ 
3   for  $q=1$  to 2
4      $g \leftarrow \left\lfloor \frac{X}{q} - \frac{X}{q^*} \right\rfloor$ 
5     if  $Q_p < Q_{p-1}$ 
6        $CLR \leftarrow CLR \cup MST(RS)$  with affiliate
7     end if
8      $RS \leftarrow$  Algorithm II.A and Occurrence
9   end for
10 end for
11 Obtaining Data Incentive Weight Matrix based on Figure 2 and 3
12 while  $n \geq 1$  do
13    $CLR \leftarrow CLR \cup MST(RS)$ 
14 end while
15 Output: CLR
    
```

**4 Emotional computing algorithms for cross-modal data fusion**

The edge network is mapped to a finite data set space from the set of emotional data elements inspired by heterogeneous emotional events. All emotional events and emotional data elements in the space are balanced. The probability of emotional events in the cross-modal fusion model is both independent. Emotional computing relies on the current cross-modal fusion of emotional data collection efficiency of edge networks and the input data incentive status of dynamic capture of emotional events. It is assumed that the data incentive state of emotional events has the characteristics of distributed randomness, such as emotional event subject, emotional data forwarding node, emotional data incentive node, and edge network topology. These characteristics can not only help the dynamic updating of the cross-modal fusion model but also improve the accuracy of emotional event data incentive in the edge network. They can also help the mapping variable transformation of multi-level data incentive and cross-modal fusion and the determination of their relationships. Therefore, emotional computing has a cross-modal category of emotions at different times. These different kinds of emotions are directly

related to the random environment of the edge network. The emotional data incentives and computational behaviors of cross-modal data fusion are identified by probability distribution. In order to calculate the probability distribution, we represent the cross-modal emotional data as a distributed edge probability matrix. These probability matrices are helpful for calculating the stochastic state transition matrix, as shown in Fig. 4.

Figure 4 shows the random state transition matrices of probability distribution composed of four different edge probability matrices. Cross-modal fusion edge terminal data excitation is used as the transition entry of Fig. 4. These data incentives are the initial state of emotional events in emotional computing. The transition probabilities of these states are



**Fig. 4** Stochastic state transition matrix

determined by the transition probabilities and iteration delays of the four different matrices in Fig. 4. Each matrix given in Fig. 4 is a dynamic real-time matrix inspired by cross-modal fusion model and edge network data. There is a strong randomness in any emotional computation in the set of heterogeneous and different categories of emotional events. Moreover, for any emotional event capture and emotional computation, they all have the possibility of transferring as shown in Fig. 4.

To sum up, the process of emotional computing is an iterative process of random state capture and transfer. Because this process occurs in the shared space of cross-modal fusion and the subspace of linear mapping, each emotional event is a discrete event independent of each other, and the probability of emotional state transition of each emotion can be obtained by formula (4).

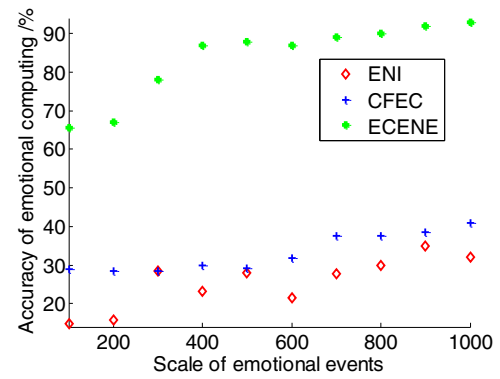
$$\begin{cases} p[event|X] = \frac{\sum_i x_i(\alpha, \beta)^\gamma}{\sum_j h(event_j)} \\ h(y) = \sum_k F_S(y)Q_k \end{cases} \quad (4)$$

The function  $h(y)$  denotes the emotional state of solving  $F_S(y)$  according to the mapping factor  $Q$ .  $p[event | X]$  can be transferred according to the emotional state of  $h(y)$ , and emotional computation is carried out in gamma space under the cross-modal fusion of vectors alpha and beta.

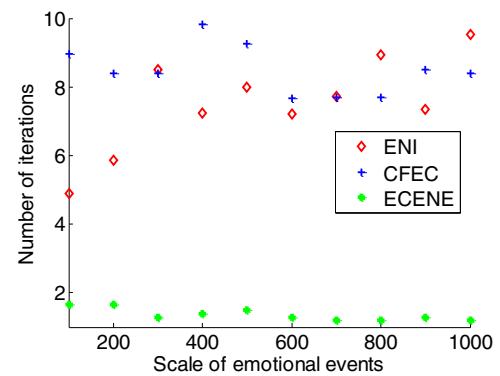
In the process of emotional computing, the edge network data incentive of emotional event mapping directly affects the discrete degree of the state distribution of the emotional computing results. The clustering core position of the weight distribution of emotional events and emotional data incentive

**Table 2** Emotional proportion

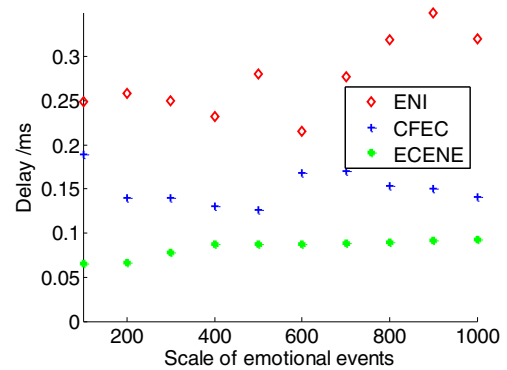
Emotion	Proportion (%)
Happiness	5
Anger	5
Disgust	10
Fear	12
Sadness	10
Surprise	5
Interest	2
Surprise	6
Pain	8
Disgust	10
Pleasure	15
Anger	23
Fear	25
Sadness	26
Shyness	10
Contempt	8
Guilt	9



(a) Accuracy of emotional computing



(b) Number of iterations



(c) Delay

**Fig. 5** Emotional computing average performance: **a** accuracy of emotional computing, **b** number of iterations, and **c** delay

states collected by edge networks plays a decisive role. According to the previous cross-modal fusion model, edge network data collection algorithm, and data incentive mechanism, the probability distribution of emotional data can be calculated according to the centralized server cluster state of emotional events recognized by edge network and the data cluster state of emotional events collected at present, based on the active emotional events, cross-modal states, and distributed edge network. The network state shows the possible

output results and the output probability of emotional computation in cross-modal data fusion.

### Algorithm IV.A

```

1 Input:  $Q_1, Q_2, \dots, Q_n, \alpha, \beta, \gamma$ 
2 for  $p=1$  to  $n$ 
3   for  $q=1$  to 2
4   invoking Algorithm II.A
5   if  $\frac{\|(DN \cdot W_S - F_S)\|^2}{\|DN \cdot W_S - F_S\|^2} > 1$ 
6     if  $Q_p < Q_{p-1}$ 
7       update  $\alpha$ 
8     end if
9     update  $\beta$ 
10    end for
11 end for
12 update  $\gamma$ 
13 while  $x \neq \text{null}$  do
14   calculating  $p$  with formula (4)
15   calculating  $h$  with formula (4)
16 end while
17 Output:  $p, h$ 

```

## 5 Simulation experiment and analysis

We use the mixed programming of MATLAB and C++ to realize all the contrast algorithms. Due to the disorder of emotional images and multimedia data submitted by edge terminals after network forwarding, the time delay of cross-modal, multi-modal, and fusion perception emotional computing will also be

**Table 3** Accuracy of emotion recognition

Emotion	ENI (%)	CFEC (%)	ECENE (%)
Happiness	35	45.7	94.6
Anger	45	52.7	93.5
Disgust	50	55.6	92.6
Fear	52	62.4	95.2
Sadness	60	71.6	92.6
Surprise	15	15.6	90.5
Interest	52	51.8	95.6
Surprise	61	62.5	95.7
Pain	18	21.8	90.5
Disgust	10	30	89.5
Pleasure	14.5	16	92.5
Anger	20.3	25	87.8
Fear	22.5	28	90
Sadness	76	72	87.9
Shyness	70	71	88.9
Contempt	78	79	90.2
Guilt	69	70	89.5

affected. We classify emotional data across layers according to the size of edge terminals and their delay in submitting emotional data. Each emotional subclass contains 1250 emotional event users in the data set. These emotional events will be randomly divided into 52 groups based on cross-modal fusion and its mapping weights. Each time, we do experiments in a group, and finally we calculate the average and standard deviation of the performance. The proportion of different emotions in the experimental data set is detailed in Table 2.

We compare the performance of different-scale edge network emotional data for heterogeneous emotional computing. The Edge Network data Incentive algorithm (ENI) and Cross-modal data Fusion Emotional Computing algorithm (CFEC) are used as the comparison algorithms. The proposed algorithm is recorded as ECENE. The average performance of emotional computing using different emotional data incentives and different processing methods is shown in Fig. 5 (abscissa is the scale of emotional events), where (a) shows the accuracy of emotional computing, (b) shows the number of iterations, and (c) shows the delay.

From the experimental results, we find that:

- (1) Generally, the performance of cross-modal fusion of different kinds of emotional events is better than that of single-modal and multi-modal feature fusion, which may be because cross-modal feature extraction of emotional events can effectively mine different emotional event features using cross-layer mapping space
- (2) The proposed ECENE algorithm can adapt to almost any scale of emotional events. The average performance of sensory events is better than that of baseline method



- (3) With the increase of emotional events, the characteristics of discriminating emotional data of different modes and edge networks are different; but only by combining cross-modal fusion with emotional incentives of edge networks, and carrying out emotional computation based on random distribution, can we better meet the needs of emotional computing
- (4) For emotional events containing similar emotional states, only the proposed algorithm ECENE has satisfactory accuracy
- (5) The overall accuracy, iteration times, and delays of ENI and CFEC algorithms are subject to external interference, especially the dynamic changes of cluster topology of centralized servers, indicating that only distributed or centralized algorithms are used to deal with heterogeneous emotional events. Recognition and emotional computation are not enough

Table 3 shows the recognition accuracy of three contrast algorithms for all emotions in Table 2. The proposed algorithm ECENE has an average recognition rate of more than 92.5% for all kinds of emotions, 52.5% higher than ENI and 42.5% higher than CFEC. The performance improvement of the proposed algorithm ECENE can be attributed to three reasons:

- (1) The deep cross-modal fusion scheme of the proposed algorithm ECENE can capture the semantic deviation between multi-modals through non-linear cross-layer mapping. The cross-modal data fusion method of the proposed algorithm ECENE can combine the deep mining network and the cross-modal data fusion method to learn the fusion semantics of each cross-modal and large data space and construct the cross-modal network model.
- (2) The low-dimensional space of the proposed algorithm ECENE aggregates the features of different edge terminals and reduces the time complexity of feature extraction. Then, the proposed algorithm ECENE can optimize the time complexity of centralized service cluster and reduce the latency of large data collection through cross-modal fusion with centralized service cluster nodes in edge network.
- (3) The proposed algorithm ECENE can not only update the cross-modal fusion model dynamically but also improve the data incentive accuracy of emotional events in the edge network. It can also accurately determine the mapping variable transformation between multi-level data incentive and cross-modal fusion and their relationships.

## 6 Conclusions

This paper presents an emotional computing algorithm which combines cross-modal fusion and edge network data

incentive. The proposed deep cross-modal fusion scheme can give full play to the advantages of non-linear cross-layer mapping, deeply capture the semantic deviation between multi-modals, and design the corresponding deep fusion method. The above research designs a cross-modal data fusion method based on deep fusion. Starting from the overlapping delay gap, incentive weight, and balance contradiction between large data collection and error detection, this paper designs a data incentive algorithm for edge network. Based on the above research, a cross-modal data fusion algorithm is designed by combining edge network data incentive and cross-modal fusion. Emotional computing mapping and emotional data element set inspired by heterogeneous emotional events are shared, and a finite data set space is constructed. We have carried out a series of simulation experiments and theoretical analysis. Our main contribution concludes the following: (1) The cross-modal fusion network model was developed. (2) The data incentive algorithms for edge networks was designed. (3) We indicated the emotional computing algorithms for cross-modal data fusion.

The results show that the average performance of the proposed algorithm is 25% higher than that of ENI and 18.5% higher than that of CFEC. The proposed algorithm has an average recognition rate of more than 92.5% for all kinds of emotions, 52.5% higher than that of ENI and 42.5% higher than that of CFEC.

## References

1. Ren G, Zhang X, Duan S (2018) Articulatory-acoustic analyses of mandarin words in emotional context speech for smart campus[J]. *IEEE Access* 6:48418–48427
2. Laurence Likforman-Sulem, Anna Esposito, Marcos Faundez-Zanuy et.al EMOTHAW: a novel database for emotional state recognition from handwriting and drawing[J]. *IEEE Trans Human-Machine Syst* 2017, 47(2): 273–284
3. Zhou Q (2018) Multi-layer affective computing model based on emotional psychology[J]. *Electron Commer Res* 18(1):109–124
4. Dumitras T, Prakash BA, Subrahmanian VS et al (2017) Understanding the relationship between human behavior and susceptibility to cyber attacks: a data-driven approach[J]. *ACM Trans Intell Syst Technol* 8(4):1–25
5. Wang C-H|L, Koong H-C (2018) Emotional design tutoring system based on multimodal affective computing techniques.[J]. *Int J Dist Educ Technol* 16(1):103–117
6. Peng Y, Qi J, Xin H et al (2017) CCL: cross-modal correlation learning with multigrained fusion by hierarchical network[J]. *IEEE Trans Multimedia* 20(2):405–420
7. Li G, Gan Y, Wu H, Xiao N, Lin L (2018) Cross-modal attentional context learning for RGB-D object detection[J]. *IEEE Trans Image Process* 28(4):1591–1601
8. Vukotić V, Raymond C, Gravier G (2018) A crossmodal approach to multimodal fusion in video hyperlinking[J]. *IEEE Multimedia* 25(2):11–23

9. Tse WK (2016) Coherent magneto-optical effects in topological insulators: excitation near the absorption edge[J]. *Phys Rev B* 94(12):125430
10. Ng TS, Wang S (2017) Recycling systems design using reservation incentive data[J]. *J Oper Res Soc* 68(10):1–23
11. Dan P, Fan W, Chen G (2018) Data quality guided incentive mechanism design for crowdsensing[J]. *IEEE Trans Mob Comput* 17(2): 307–319
12. Lo LY, Li WO, Lee LP et al (2018) Running in fear: an investigation into the dimensional account of emotion in discriminating emotional expressions[J]. *Cogn Process* 19(3):1–11
13. Xu B, Fu Y, Jiang YG, Li B, Sigal L (2018) Heterogeneous knowledge transfer in video emotion recognition, attribution and summarization[J]. *IEEE Trans Affect Comput* 9(2):255–270
14. Dean J, Krusienski, Guoxu Zhou, Jing Jin et.al Discriminative feature extraction via multivariate linear regression for SSVEP-based BCI[J]. *IEEE Trans Neural Syst Rehabil Eng* 2016, 24(5): 532–541
15. Rayn Sakaguchi, Kenneth D. Morton, Leslie M. Collins, et.al. A comparison of feature representations for explosive threat detection in ground penetrating radar data[J]. *IEEE Trans Geosci Remote Sens* 2017, 55(12): 6736–6745
16. Liu Y, Liu Y, Ding L (2018) Scene classification based on two-stage deep feature fusion[J]. *IEEE Geosci Remote Sens Lett* 15(2):183–186
17. Yong Ye, Chunlong He, Bin Liao et.al Capacitive proximity sensor array with a simple high sensitivity capacitance measuring circuit for human–computer interaction[J]. *IEEE Sensors J*, 2018, 18(14): 5906–5914

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