Clustering-Based Emotion Recognition Micro-Service Cloud Framework for Mobile Computing

PING WANG1, LUOBING DONG2,3, YUESHEN XU2, WEI LIU2, AND NINGNING JING2
1National Key Laboratory of Science and Technology on ATR, College of Electronic Science, National University of Defense Technology, Changsha 410073, China
2School of Computer Science and Technology, Xidian University, Xi’an 710071, China
3College of Liberal Arts and Sciences, National University of Defense Technology, Changsha 410073, China
Corresponding author: Luobing Dong (lbdong@xidian.edu.cn)

This work was supported in part by the Fundamental Research Funds for Central Universities under Grant JB161004.

ABSTRACT In a situation where life becomes more stressful and challenging, people feel compelled to be more concerned about their mental situation. Different emotional statuses are external reactions to different mental states. Therefore, researchers always identify people’s mental situation by monitoring their real-time emotions. At the same time, due to the availability of built-in sensors in a smartphone, applications that can identify real-time emotions of mobile users are constantly emerging. However, compared to most emotion recognition algorithms, computing resources and battery life in mobile phones are always limited. This makes accuracy and latency of these applications are unsatisfactory. In this paper, we propose a micro-service platform for mobile emotion recognition application developers (MSPMERAD) which can supply high performance. First, a classifier fusion emotion recognition algorithm is proposed by using a dynamic adaptive fusion strategy. Second, this new algorithm is encapsulated into a micro-service. With other affiliated micro-services such as data uploading, preprocessing, etc., developers can ignore the implementation of the emotion recognition algorithm and just focus on how to collect sensor data and interact with users. The accuracy and latency of one application based on the MSPMERAD are compared with another application that is implemented using a locale emotion recognition algorithm. Experiments based on the daily behavior data of 50 student volunteers show that the application based on our platform has higher recognition accuracy with a more reasonable time.

INDEX TERMS Classifier fusion method, dynamic adaptive fusion strategy, emotion recognition, micro-service, mobile users.

I. INTRODUCTION

With the development of modern society, The corporeal property of one person is richer and richer. People can use a lot of material resources in their daily life and work. This frees them from worrying about basic livelihoods such as food, clothes, etc. But a lot of factors make life more challenging and pressured than anytime before. These factors come from all aspects of life such as tests in the education system, requests from friends, family relationships, countless tasks at work, etc. The constant pressure will have a far-reaching influence on people’s mental health. Nowadays, people are becoming more aware the fundamental connection between their physical health and their mental health. They increasingly value their mental health. Because emotion regulation is an external reaction to the mental situation of people, it is stated as one of the key features of people’s mental health by researchers [16]. Therefore, it is very important that people’s emotional status is recognized in time to the medical judgment of their mental disorder and their timely mood regulation. Emotion recognition which is known as a promising technology has been widely used in mental health monitoring over the previous decades [29], [30], [35].

Because of the unlimited commercial and academic potential of emotion recognition in the fields of artificial intelligence, it has also become one important research topic. [22]. Emotion detection based on AI algorithms has been a $20bn industry by 2019 [5]. Physiological signals such as
electromyographic signal [31], skin temperature [15], facial expressions [23], blood volume pulse [34], etc. are typically used to identify human emotional status by scientists. However, these methods are very difficult to apply in practice because these signals are impossible to get without special types of equipment and bio-sensors. Doctors always hope to monitor the processes of their patients’ emotions changing, because these processes can help them do some decisions and judgments. For instance, some special carry-on equipment are always needed by patients because their doctors want to watch their long term emotional status. This can help doctors to find the connection between patients’ emotional status and some clinical symptoms. Wearing these equipment is inconvenient. Sometimes it will also make patients depressing. So the timely recording and recognizing the emotions of the observed person who are not informed ahead of time is very important.

In recent years, mobile phones have replaced computers as the main entrance of the Internet [33]. At the same time, mobile phones are more affordable and powerful in functions than ever before. The time people spend on their mobile phones rapidly increases. Moreover, sensor devices including gyrooscope, temperature, lighting and gravity, etc. are equipped not only in high ranking smart phones but also in low ranking and middle ranking smart phones. Many behaviors of users such as the working speed, the sound volume, the tension of the phone’s touch panel when the finger press on it, etc. can be recorded by these sensors. Many types of emotional status can be extracted from these behaviors. The popularity of sensors in mobile phones enables doctors record and recognize the emotional status of their patients by using their mobile phones without inform the patients [13].

Good human emotional state is very important for the creating of good physical and mental health. Therefore, one technology that can identify the external features of human emotional state and recognize the corresponding emotional state timely may bring enormous benefits to our social society. We focus our research on these two issues. A new mobile users’ emotional status recognition method is present based on the analysis of physiological and ecological features that are collected by mobile phone sensors. First, 50 volunteers are selected from Xidian University. Their daily behavior data are collected through our mobile application (EAmobile) which is installed in the volunteers’ mobile phones. Second, a classifier pool including most machine learning based classification algorithms is assembled by thinking about both complementary and diversity. Third, classifiers that have high inconsistent are picked and combined based on a dynamic weighted fusion strategy. The class-conditional probabilities and prior probabilities of base classifiers are used to dynamically decide their fusion weights. Finally, the classification results are map to the emotional status of mobile users.

However, as you know, the computing resource and battery capacity of mobile phones are limited [11]. This does not meet the need for fast execution of most classifiers. To address this issue, researchers have attempted to offload computing tasks to remote cloud platforms [12]. But in our system, the classifiers which are used in the final fusion strategy is dynamically changed. And various new classifiers will be continuously added to the classifier pool during use. To ensure its scalability, we adopt a micro-service architecture to implement our system. The fusion framework and all classifiers are encapsulated as individual micro-services. Based on this, mobile phones can just focus on collecting data and displaying results and neglect the recognition process which is resource sensitive. We implemented the system using micro-service architecture and all computing localization architectures respectively. Experiments based on the daily behavior data of 50 student volunteers show that the application based on the micro-service architecture has higher recognition accuracy with a more reasonable time. We provide the following main contributions:

- We study the relation between mobile phone sensor data and emotional status. First, inspired by the circumflex model of affect [27], we describe emotional status from two aspects: pleasure degree and activity degree. Second, the real-time emotional coordinates of mobile users are computed by using the classifier fusion method.
- We study the classifier selection strategy and fusion strategy. Accuracy and difference factors are both considered to ensure that the best classifiers for the current data set can always be selected as base classifiers. The fusion weight of each base classifier is dynamically set by using its historical decision and class-conditional probability.
- A common micro-service based emotion recognition framework is proposed to ease the implementation of similar applications. All functions are encapsulated as individual micro-services which make them loosely coupled. Developers can arbitrarily change the implementation method of a single function without re-developing other functions.

In the following, we describe the organization of this paper. In Section II, we retrospect some theories and tools for resolving the emotion recognition problems. We discuss in Section III the main methodology of the fusion method. In section IV, we construct an emotion recognition micro-service framework. In Section V, we perform experiments to verify the accuracy of our new method and the efficiency of our framework. We finally give the conclusion in Section VI.

II. RELATED WORK

In the fields of emotion recognition, emotion recognition going by text analysis, emotion recognition going by verbal and facial expression, and emotion recognition going by physiological signal analysis are three common categories. We will review these three categories in turn in this section. For the emotion recognition going by text analysis,
two main types are researched: the semantic analysis based method and the word classification based method. The former uses the semantic network to identify the emotional status corresponding to a special text. The richer the semantic knowledge base is, the more efficient this method is. In paper [2], SentiwordNet is present to devise for opinion mining and sentiment classification. A semantic network with common-sense knowledge which is based on the natural language is introduced into the public knowledge classification by Cambria et al. [7]. Sentiment analysis can be achieved through the multiple dimensional expansion of the semantic network. The latter implements emotion recognition by a special dictionary. This dictionary maps different emotionally salient words to different emotional scores. The emotional state implied in one paragraph can be calculated by fetching the emotional score of each word from this dictionary. Obviously, the effectiveness of this method is mainly dependent on the richness of emotionally salient words in the dictionary. Thus, many researchers try to design recognizing and scoring methods for emotionally salient words [8], [9], [20].

The analysis of speech, text, and images are more and more effective because of the DNN’s rapid development and rise. Thus, facial and verbal expressions analysis based emotion recognition comes to be an important issue. People can identify the emotions of other peoples from their facial muscle movements and the corresponding facial expressions because different emotions relate to different expressions. For instance, the emotion status of one person can be judged as happy when his ruffles emanate from his canthus and the corner of his mouth curves up. In contrast, the emotional status of one person can be judged as anger when he frowns and his eyes pop out. Both overall features and local features of the observed person’s facial expression can be used to recognize his emotion. The former considers features of the entire face because people have different overall facial features when they have different emotions. The latter is ground on that people have facial features of various relative positions, sizes and shape in different situations. In this field, a lot of achievements have been made. For instance, a big quantity of pictures of facial expression from various ages, genders, and races are collected to create a facial expression data set by Affectiva [26]. Based on this data set, an artificial intelligence algorithm is designed to recognize human face textures and wrinkles, changes in facial feature shapes, and so on. Many people always do not want other people to know their thoughts and intentions. So they try to be emotionless in their normal life like wearing a poker face. This reduced the effectiveness of emotion recognition technology that is based on the analysis of verbal and facial expressions. Therefore, physiological signals are used to identify the emotions of people [4], [14], [19], [21], [25]. Various emotional characteristics of EEG are studied by Wang et al. by tracking the EEG signals’ changes. The classification of emotions is finally identified by the establishment of the connection between emotional status and EEG characteristics [32]. The EEG signals are combined with the clip database to construct a real-time emotion recognition system by Liu et al. [24]. Algorithms of emotion recognition going by machine learning have valid presentation on extracting signals which have high frequency and spatial characters. Because when machine learning is used to fuse multimodal information like various physiological features and EEG signals a good effect can be achieved [3], [18].

Nowadays, more and more researchers are working to obtain people’s emotional states by mobile phones. M. SHAMIM et al. use the embedded cameras of smart phones to get the facial video of users [28]. The most dominant bins of the concatenated histograms are extracted from representative frames by the Kruskal–Wallis Feature Selection. The emotion is finally classified by Gaussian mixture model-based classifier which can guarantee high recognition accuracy in a reasonable time. In this work, the speed of emotion recognition is improved by simplifying the computational complexity of the algorithm. However, this will inevitably decrease accuracy. Humaid Alshamsi et al. try to solve this problem by introducing cloud computing [1]. Facial video and Speech emotion signals which are collected by mobile phones are transferred to the back cloud computing platform. They use Mel-Frequency Cepstral Coefficient Facial Landmarks to pre-process the signals. A (SVM) based algorithm is devoted to cluster the emotion. This framework can reduce the computational pressure of mobile phones by offloading the complex emotion recognition process to the cloud platform. As we all know, the amount of data describing facial video and voice emotion signals is usually very large. The transmission of these data between mobile phones and the cloud is always time-consuming and energy-intensive [12]. Therefore, emotion recognition algorithms based on mobile sensors with a small amount of binary non-audio and video data should be considered. These sensors include gravity, gyroscope, pressure, temperature, lighting, etc.. They can perceive user’s actions with a small amount of data such as a value.

With the promotion of a cloud-based application, its user base grows bigger and bigger. The coordination among different development teams becomes more and more difficult as its code gets more and more complicated. Recently, miro-service architecture is gaining popularity due to its granular approach and loosely coupled services [28]. In a miro-service architecture, an application consists of many independent services that answer to different functions. In businesses across different industries—from telecommunications and retail to financial services and manufacturing—Many major companies like Amazon, Netflix, Uber, etc. are choosing miro-service architecture to design their new applications. Researchers are trying to introduce miro-service architecture into many different areas [17]. Svetoslav Zhelev et al. use microservices and event driven architecture to implement big data stream processing [36]. Alexandr Krylovskiy et al. apply the miro-service architecture style to design a Smart City IoT platform [20]. This method can increase the energy efficiency of a city at the district level.
III. MAIN METHODOLOGY
In this section, we discuss the main idea of the proposed classifier fusion method. First, we describe the emotional model. This model is the base of the mobile users’ emotional statuses measurement. Then, we explain how to eliminate useless data and extract meaningful features from the collected daily data of mobile phone sensors, followed by the detail of the proposed dynamic classifier fusion method.

A. EMOTIONAL MODEL
Considering that discrete variables are not suitable for linear analysis and transformation, we propose a new emotional model which is inspired by [27]. Although people may have different paths to their well-being, they have some common external characteristics. When they feel good, they will be pleased and active. Aristotle said: “Happiness is a state of activity”. The good emotion is not just accompanied by pleasure but also positive activity. Therefore, the proposed emotional model describes users’ emotional statuses from two dimensions: activity level and pleasure level. Activity levels are used to measure how active or passive a user is. A coordinate system is constructed with pleasure as the Horizontal coordinate and activity as the Vertical coordinate (shown figure 1). Then, we can position each emotional state in this coordinate system.

![Emotional model](image)

The activity level and pleasure level are divided into five levels. The origin is at \((P3, A3)\). Then \((P5, A5)\) represents very active and very pleasant. \((P1, A1)\) represents very inactive and very unpleasant. Through statistical analysis of the collected information, we can know that the emotional statuses of most mobile users are higher than the third and lower than the fourth levels. Therefore, the emotional status value of one mobile user satisfies the following rule most of the time: \(P3 \leq Pi \leq P4\) and \(A3 \leq Ai \leq A4\). This is because that extreme emotions rarely occur. In normal life, one person is more probable in a happy status than in a sad status. The statistics results of the distribution of mobile user’s emotional statuses are shown in table 1 of paper [10].

| k | silhouette coefficient | k | silhouette coefficient |
|---|------------------------|---|------------------------|
| 2 | 0.51                   | 6 | 0.42                   |
| 3 | 0.62                   | 7 | 0.45                   |
| 4 | 0.54                   | 8 | 0.48                   |
| 5 | 0.47                   | 9 | 0.39                   |

B. CLASSIFIER FUSION METHOD
The main idea of our method is to map the sensor data collected by the mobile phone into the emotional model. Then we can get the real-time emotional status of the mobile phone user. As we all know, there are many classifiers that can be chosen, such as Perceptron, Naive Bayes, Decision Tree, K-Nearest Neighbor, Support Vector Machine, etc. In real pattern recognition applications, different classifiers make “independent” errors. In our system, different sensors will collect different data sets. These data sets would give different results based on different classifiers. Therefore, we use multiple classifier method which achieves better results by integrating classical classifiers. The main problems of traditional multiple classifier methods are how to choose good base classifiers and how to design a fusion strategy. In our method, based on the diversity strategy, complementarity is taken into account simultaneously when selecting a classifier. In the classifier selection phase, we combine clustering algorithms with the silhouette coefficient to determine the number of target base classifiers and use inconsistencies to evaluate the differences between them. In the fusion phase, we set the weights of the basic classifiers based on the prior and conditional probabilities.

In the followings, the details of our classifier fusion method are described from three phases: the generation phase, the selection phase, and the fusion phase.

1) GENERATION PHASE
Considering the diversity of base classifiers, we construct a base classifier pool in this stage. The base classifiers are chosen from the classification algorithms that are based on decision trees and neural networks. We train these base classifiers on different or disjoint sub-spaces. To achieve diversity, we choose unstable classifiers like decision trees and neural networks to construct the classifier pool. These classifiers are trained on the same data set using k-fold cross-validation and bagging.
Fig. 3 and Fig. 2 show the training process of our base classifiers. For the same data set, even the same algorithm may generate different classification decisions based on different parameters. In this case, we use Bagging to combine the decision boundaries of these classifiers to reduce the overall error and achieve a better classification effect. For a single individual classifier, it will be trained using 10-fold cross-validation. The original data set is randomly divided into 10 parts using the non-repeated sampling method. 9 of these parts are selected as the training set. Another one is reserved as a testing set. We assume that $T_i$ denotes the training result corresponds to the one using the $i$th part as the training data set. Then the accuracy of the trained classifier is obtained as the average of these 10 $T_i$: 

$$S_{D} = \frac{1}{n} \sum_{i=1}^{10} T_i.$$ 

This cross-validation method can reduce over-fitting to a certain extent. At the same time, it can also obtain as much valid data as possible from the limited information which can reduce the sensitivity of the trained classifier to data partitioning. When the segmentation is performed only once (or less frequently), the distribution of the subset may be approximately the same as that of the original data set. So we use 10-fold to guarantee the diversity of the trained classifier.

2) SELECTION PHASE
For the accuracy of the fusion classifier, we select the most suitable classifiers for the target data set from the classifier pool. The first problem is how many classifiers should be chosen. We use $k$-means method to partition the data set into $k$ clusters. Then use the silhouette coefficients that is correspondent to different $k$ to find the best $k$. Let $S_i$ denotes the silhouette coefficient of data item $d_i$ when we set $k = j$. The silhouette coefficient of data set $D$ can be defined as formula (1). Then the best $k$ can be got as $\arg\max_k S_{D}^k$.

$$S_{D}^k = \frac{1}{n} \sum_{i=1}^{n} S_i^k$$

When people choose classifiers, they always tend to select classifiers which are most suitable for the target data sets. Because suitable classifiers can improve the generalization ability and accuracy of the final fusion classifier [6]. As the unstable classifiers are sensitive to the changing of training data set and learning data set, they tend to be the best base classifiers of combination strategies. They can get very different classification results even there are very subtle changes in the training samples. After we determine the number of base classifiers, we choose the same number of the most mutually inconsistent and the most accurate classifiers for the same data set to fuse as the final decision classifier. First, we set the accuracy threshold. Classifiers whose accuracy are higher than this threshold will be chosen as the spare classifiers. Then the final base classifiers will be chosen from these spare classifiers based on their difference operators.
The difference operator of one classifier $C_i$ can be calculated by formula (2).

$$\Lambda_i = \frac{\sum_{j=1,j\neq i}^{N} N_{C_i C_j} + N_{C_i \bar{C}_j}}{N(N-1)}$$ (2)

Here, $N$ represents the total number of the classifiers in the pool. For any sample which belongs to the $m$th class, if $C_i$ identify that it belongs to the $m$th class and $C_j$ identify that it belongs to another class, $N_{C_i C_j}$ add 1. For any sample which belongs to the $m$th class, if $C_j$ identify that it belongs to the $m$th class and $C_i$ identify that it belongs to another class, $N_{C_i \bar{C}_j}$ add 1. Obviously, $\Lambda_i \in [0, 1]$. A large difference operator indicates the classifier has a large difference with all other classifiers. The $k$ classifiers with the largest difference operators will be selected as the final base classifier.

3) FUSION PHASE

Average fusion, weight fusion, and majority vote fusion are the three most popular fusion strategies. But they all have some known defects. The average fusion and majority vote strategies are sensitive to extreme values. Their performance are uncertain. Researchers try to set different weights to different base classifiers on the basis of their features like accuracy. This is the main idea of the weight fusion strategy.

In our fusion strategy, the weights of the base classifiers are dynamically changing according to the feature of the target data sets. This dynamic adaptive weight strategy is implemented based on the classification confidences and the prior probabilities of the base classifiers.

The prior probabilities of the base classifiers can be computed by all of their history decision results on the same type of data sets. Suppose that $l$ classifiers are chosen as the base classifiers of the fusion strategy, their prior probabilities can be described by the confusion matrix which is defined as formula (3).

$$CM = \left\{ \begin{array}{c} \sum_{i=1}^{l} \sum_{p=1}^{l} \sum_{q=1}^{l} z_{pq} \\ l \times l \times l \end{array} \right\}$$ (3)

If the $i$th classifier mistakenly partition one sample into the $p$th class and this sample belongs to the $q$th class, the value of $z_{pq}$ will increase by 1. All of the history training results are counted. And the item values of $CM$ are normalized.

The classification confidence of the fusion classifier can be computed by the final weights and the class-conditional probabilities of the base classifiers. For one classifier $C_i$, we assume that $p_i$ denotes its class-conditional probability of the $k$th class. Then the confidence of the final fusion classifier on the $i$th class is defined as formula (4).

$$F_i = \sum_{k=1}^{l} \lambda_{ki} P_i^k$$ (4)

Here, $\lambda_{ki}$ represents the $k$th fusion weight of the $i$th classifier. It can be get from the formula (5).

$$\lambda_{ki} = \frac{z_{kk}}{\pi_k}$$ (5)

In formula (5), $\pi_k = \sum_{q=1,q \neq k}^{l} z_{kq} P_i^q + \sum_{q=1,q \neq k}^{l} z_{qk} P_i^q$. The definition of $\rho$ is shown in formula (6). It denotes the inverse reliability of the $i$th classifier at the $p$th class according to the $q$th class.

$$\rho_{pq}^i = e^{-(P_i^p - P_i^q)\rho}$$ (6)

The final classification decision can be got from the partition result of the fusion classifier which has the highest confidence.

IV. MSPMERAD

Limited by computing power and battery capacity, mobile phones cannot consistently provide sufficiently accurate results just based on local computation for computationally sensitive applications in a reasonable amount of time. So we use the remote cloud to supply assistance. The processing of the emotion recognition algorithm is offloaded to the cloud. Mobile phones are just responsible for the collection of sensor data and showing the recognition results. For the sake of the scalability of the emotion recognition cloud service, each component and classifier is encapsulated into one micro-service which is implemented as one Docker Container.

Fig 4 shows the architecture of our based emotion recognition system. All mobile phones which installed the emotion recognition client application EAmobile will collect users’ behavior data through sensors. These data are transferred to the remote cloud and pre-processed by the pre-processing micro-service. Features that can indicate the emotional status will be extracted from the original data and stored into the HBase database on the cloud. The real-time emotional status of a mobile user can be perceive by the emotion recognition micro-service based on the real-time sensor data of its mobile phone and all historical sensor data of all mobile users and their corresponding emotional status. The recognition result will also be stored into the HBase database on the cloud. The pre-processing, features extraction, classifier implementation, choose of base classifiers and fusion are all encapsulated into micro-services based on the Kubernetes framework. Therefore, service providers can improve the accuracy of their emotion recognition services by only improving individual services, such as replacing classifiers and improving fusion methods, rather than rewrite all services.

In what follows, the useless data eliminating method using the pre-processing micro-service is described firstly. Then, we explain the process of the feature extracting micro-service in detail.

A. DATA PRE-PROCESSING MICRO-SERVICE

In the process of collecting data, the mobile phone may encounter various problems that can cause the data can not be used directly. For instance, the ongoing data acquisition process may be interrupted by an incoming call. This incoming call will make the data not complete. The surrounding magnetic field may cause the compass data of the phone to be inaccurate. The external noise may interfere with the sounds collection sensor and then the collected data are noisy.
The data pre-processing micro-service can eliminate the noise, fill the incomplete data, and delete the inaccurate data. Fig. 5 shows the process flow of the pre-processing micro-service. The initial data which are incomplete, noisy or inaccurate will be replaced or filled with the reference data. These data represent the value of default, average or the neighbor data. If there are no these reference data, these data will be delete.

**B. FEATURE EXTRACTING MICRO-SERVICE**

Different sensors can collect different data. Some of these data indicate different emotional statuses of users. For example, when a user’s sound is louder than usual and the frequency of the shaking of his phones is very high always foreshadow that he is activity. At the same time, if he pressed the phone’s display screen harder than usual, he might be angry with someone. Different sensors have different data features that can be used to infer their user’s emotional status. Therefore, we extract these features from the data which have been pre-processed to enhance the accuracy of the emotion recognition algorithms. The details of the features which are extracted by the feature extracting micro-service are the same as those in paper [10] (Table 2).

**V. EVALUATION**

In this section, we will describe the evaluation results of the proposed system. We construct an android application that is named EAmobile. It can collect daily sensor data of the user of its host phones. The sensors which are monitor is consistent with table 3 of the paper [10]. 50 volunteers are chosen from

---

**TABLE 2. Accuracy of different groups of base classifiers.**

| no. | base classifier 1 | base classifier 2 | base classifier 3 | base classifier 4 | accuracy (%) |
|-----|------------------|------------------|------------------|------------------|--------------|
| 1   | LAD              | CART             |                  |                  | 64.48        |
| 2   | LAD              | DTNB             |                  |                  | 70.08        |
| 3   | LAD              | PART RULE        | DTNB             |                  | 69.17        |
| 4   | LAD              | CART             | DTNB             |                  | 71.67        |
| 5   | LAD              | REP              | DTNB             |                  | 60.75        |
| 6   | LAD              | MLP              | DTNB             |                  | 59.45        |
| 7   | LAD              | CART             | PART RULE        | DTNB             | 72.37        |
| 8   | LAD              | CART             | DTNB             | REP              | 70.58        |
| 9   | LAD              | CART             | DTNB             | PART RULE        | 70.43        |
| 10  | LAD              | CART             | DTNB             | MLP              | 69.31        |
| 11  | LAD              | CART             | DTNB             | SVM              | 67.56        |
the undergraduate students of Xidian University. They install EAmobile on their phones and keep it running for 7 months. They do some necessary labels on the daily data which can help us identify the mapping relationship between data and in time emotional statuses. Our remote cloud server collects these data every day. The pre-processing micro-service and feature extraction micro-service will handle these data and store them into the HBase database in the system. We use these data to train our fusion classifier and get the best weights. We choose 8 popular classifiers and put them into our classifier pool. Table 3 in paper [10] shows the original accuracy without fusion when these 8 classifiers are trained on our sample. To compare the effectiveness of our method, we use some normal classifiers: Linear Discriminant Analysis (LDA), Classification And Regression Tree (CART), REP, MLP, etc.

### A. NUMBER OF BASE CLASSIFIERS

We change $k$ from 2 to 9 and compute the silhouette coefficients of the data set after applying k-means clustering. Table 1 shows the results. We can know that, the silhouette coefficients achieves its maximum value when $k = 3$.

We choose 2, 3 and 4 classifiers as the base classifiers of our adaptive weight fusion classifier respectively. The recognition results are shown in Table 2. Group no. 4 and 7 have the best performance. This results are consistent with the results when we change the number of base classifiers. When we choose 3 base classifiers, higher performance could be reached. When we choose two base classifiers, the accuracy would not always be higher than when only one basic classifier is selected. For group no. 1, 2, and 3, the accuracy are improved. This is because they choose two classifiers which are different types that the mutual complementary is aroused. But for group no. 5 and 6, the accuracy drop because the two base classifiers come from the same class. The same rule can be got from the four base classifiers groups. Therefore, in the following experiments, we try to choose three classifiers from different classes as the base classifier for the fusion classification method.

### B. EFFECT OF OUR FUSION STRATEGY

We do experiments using our strategy and another two fusion strategies and compare their results. The accuracy based method and the majority vote method are chosen. Data in table 3 shows the accuracy of these three strategies on our samples. All results are the average results of experiments that are repeated 5000 times. Results show that the proposed method can always reach the highest performance. We construct two classifier pools. The first one consists of REP, CART, and LAD. They have the largest average difference and accuracy. Thus, the performance of the accuracy strategy is better than that of the vote strategy. The system’s accuracy is significantly improved when we replace the strategy as the proposed strategy. In the second group, DTNB is replaced by REP. REP possesses the lowest accuracy percentage.
Then the system’s emotion recognition accuracy is improved under all strategies.

### C. EFFICIENCY OF THE MICRO-SERVICE PLATFORM

We compare the efficiency of emotion recognition in two computing modes, local computing and micro-service based computing. The computing time is got from the average computing time of 500 individual experiments. The average computing time of local computing mode is 3s. At the same time, the result of micro-service based computing mode is 50ms. Table 4 shows the parameters of the mobile phones and cloud servers we used in the experiments.

### VI. CONCLUSION

A microservice-based emotion recognition system for mobile phone users is present in this paper. We train the dynamic adaptive fusion weights of base classifiers on the behavior data of 50 student volunteers which are collected by their mobile phone sensors. The highest average accuracy can achieve 72.37%. The micro-service platform can easy the construction of emotion recognition mobile applications. Application providers and algorithm researchers can improve their work without affecting each other. In the following work, we will open this architecture to APP developers for real application.

### REFERENCES

[1] H. Alshamsi, V. Kepuska, and H. Meng, “Automated facial expression recognition app development on smart phones using cloud computing,” in Proc. IEEE 8th Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf. (U/MCON), Oct. 2017, pp. 577–583.

[2] S. Baccianella, A. Esuli, and F. Sebastiani, “Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining,” in Proc. Int. Conf. Lang. Resour. Eval. (LREC), vol. 10, 2010, pp. 2200–2204.

[3] M. Balconi and G. Mazza, “Brain oscillations and BIS/BAS (behavioral inhibition/activation system) effects on processing masked emotional cues,” Int. J. Psychophysiol., vol. 74, no. 2, pp. 158–165, Nov. 2009.

[4] D. Baur. (2019). AI-Based Emotion Detection has Become a 20B Industry. Accessed: Jul 1, 2019. [Online]. Available: https://www.theguardian.com/technology/2019/mai/06/facial-recognition-software-emotional-science

[5] L. Breiman, “Bias, variance, and arcing classifiers,” Dept. Statist., Univ. California, Berkeley, CA, USA, Tech. Rep. 460, 1996.

[6] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, “New avenues in opinion mining and sentiment analysis,” IEEE Intell. Syst., vol. 28, no. 2, pp. 15–21, Mar. 2013.

[7] Z.-H. Deng, K.-H. Luo, and H.-L. Yu, “A study of supervised term weighting scheme for sentiment analysis,” Expert Syst. Appl., vol. 41, no. 7, pp. 3506–3513, Jun. 2014.

[8] L. Dong, Q. Guo, and W. Wu, “Speech corpora subset selection based on time-continuous utterances features,” J. Combinat. Optim., vol. 37, no. 4, pp. 1237–1248, May 2019.

[9] L. Dong, Y. Xu, P. Wang, and S. He, “Classifier fusion method based emotion recognition for mobile phone users,” in Proc. Int. Conf. Broadband Commun., Newt. Syst. Chan, Switzerland: Springer, 2019, pp. 216–226.

[10] L. Dong, M. N. Satpute, J. Shan, B. Liu, Y. Yu, and T. Yan, “Computation offloading for mobile-edge computing with multi-user,” in Proc. IEEE 39th Int. Conf. Distrib. Comput. Syst. (ICDCS), Jul. 2019, pp. 841–850.

[11] L. Dong, W. Wu, Q. Guo, M. N. Satpute, T. Znati, and D. Z. Du, “Reliability-aware offloading and allocation in multilevel edge computing system,” IEEE Trans. Rel., to be published.

[12] A. A. Farhan, “Modeling human behavior using machine learning algorithms,” Ph.D. dissertation, Univ. Connecticut, Storrs, CT, USA, 2016.

[13] C. A. Frantzidis, C. Bratsas, C. L. Papadelis, E. Konstantinidis, C. Pappas, and P. D. Bamidis, “Toward emotion aware computing: An integrated approach using multichannel neurophysiological recordings and affective visual stimuli,” IEEE Trans. Inf. Technol. Biomed., vol. 14, no. 3, pp. 589–597, May 2010.

[14] H. Gao, Y. Duan, L. Shao, and X. Sun, “Transformation-based processing of typed resources for multimedia sources in the IoT environment,” Wireless Netw., pp. 1–17, Nov. 2019.

[15] J. J. Gross and R. F. Muñoz, “Emotion regulation and mental health,” Clin. Psychol., Sci. Pract., vol. 2, no. 2, pp. 151–164, 1995.

[16] M. S. Hossain and G. Muhammad, “An emotion recognition system for mobile applications,” IEEE Access, vol. 5, pp. 2281–2287, 2017.

[17] D. Iacoviello, A. Petracca, M. Speziello, and G. Placidi, “A real-time classification algorithm for EEG-based BCI driven by self-induced emotions,” Comput. Methods Programs Biomed., vol. 122, no. 3, pp. 293–303, Dec. 2015.

[18] M. Jenke, A. Peer, and M. Buss, “Feature extraction and selection for emotion recognition from EEG,” IEEE Trans. Affect. Comput., vol. 5, no. 3, pp. 327–339, Jul. 2014.

[19] F. H. Khan, U. Qamar, and S. Bashir, “Lexicon based semantic detection of sentiments using expected likelihood estimate smoothed odds ratio,” Artif. Intell. Rev., vol. 48, no. 1, pp. 113–138, Jun. 2017.

[20] M. Khezri, M. Firoozbadabi, and A. R. Sharafat, “Reliable emotion recognition system based on dynamic adaptive fusion of forehead biopotentials and physiological signals,” Comput. Methods Programs Biomed., vol. 122, no. 2, pp. 149–164, Nov. 2015.

[21] B. Ko, “A brief review of facial emotion recognition based on visual information,” Sensors, vol. 18, no. 2, p. 401, 2018.

[22] M. Li, H. Xu, X. Huang, Z. Song, X. Liu, and X. Li, “Facial expression recognition with identity and emotion joint learning,” IEEE Trans. Affect. Comput., to be published.

[23] Y.-J. Liu, M. Yu, G. Zhao, J. Song, Y. Ge, and Y. Shi, “Real-time movie-induced discrete emotion recognition from EEG signals,” IEEE Trans. Affect. Comput., vol. 9, no. 4, pp. 550–562, Oct. 2017.

[24] N. Martini, D. Menicucci, L. Sebastiani, R. Bedini, A. Pingitore, N. Vanello, M. Milanesi, L. Landini, and A. Gemignani, “The dynamics of EEG gamma responses to unpleasant visual stimuli: From local activity to functional connectivity,” NeuroImage, vol. 60, no. 2, pp. 922–932, Apr. 2012.

[25] D. McDuff, R. Kaliboub, T. Senechal, M. Amr, J. Cohn, and R. Picard, “Affectiva-MIT facial expression dataset (AM-FED): Naturalistic and spontaneous facial expressions collected,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, Jun. 2013, pp. 881–888.

[26] J. A. Russell, “A circumplex model of affect,” J. Personality Social Psychol., vol. 39, no. 6, p. 1161, Dec. 1980.

[27] V. Singh and S. K. Peddoujo, “Container-based microservice architecture for cloud applications,” in Proc. Int. Conf. Comput., Commun. Autom. (ICCCA), May 2017, pp. 847–852.

[28] G. Trigeorgis, F. Ringeval, R. Brueckner, E. Marchi, M. A. Nicolaou, G. Stratou, and R. Cowie, “A VEC 2016: Depression, mood, and emotion recognition workshop and challenge,” in Proc. IEEE Int. Conf. Affect., Speech Signal Process. (ICASSP), Mar. 2016, pp. 5200–5204.

[29] M. Valstar, M. Pantic, J. Gratch, B. Schuller, and S. Zafeiriou, “Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network,” in Proc. IEEE Trans. Inf. Technol. Biomed., vol. 20, no. 3, pp. 589–597, May 2016.  

### TABLE 4. Experiment environment.

| Cloud Server | Mobile Phone |
|---------------|--------------|
| CPU: 32 Cores 2 GHz, RAM: 64GB, OS: windows 7 | HuaWei P9, Android 7.1, 3G RAM, 64GB ROM, CPU: 8 Cores 1.84 GHz |

P. Wang et al.: Clustering-Based Emotion Recognition Micro-Service Cloud Framework for Mobile Computing
[31] A. E. Vijayan, D. Sen, and A. P. Sudheer, “EEG-based emotion recognition using statistical measures and auto-regressive modeling,” in Proc. IEEE Int. Conf. Comput. Intell. Commun. Technol., Feb. 2015, pp. 587–591.

[32] X.-W. Wang, D. Nie, and B.-L. Lu, “Emotional state classification from EEG data using machine learning approach,” Neurocomputing, vol. 129, pp. 94–106, Apr. 2014.

[33] Y. Yin, L. Chen, Y. Xu, J. Wan, H. Zhang, and Z. Mai, “QoS prediction for service recommendation with deep feature learning in edge computing environment,” Mobile Netw. Appl., pp. 1–11, Apr. 2019.

[34] B. Zhao, Z. Wang, Z. Yu, and B. Guo, “EmotionSense: Emotion recognition based on wearable wristband,” in Proc. IEEE Smart-World, Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People Smart City Innov. (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), Oct. 2018, pp. 346–355.

[35] M. Zhao, F. Adib, and D. Katabi, “Emotion recognition using wireless signals,” in Proc. 22nd Annu. Int. Conf. Mobile Comput. Netw., 2016, pp. 95–108.

[36] S. Zhelev and A. Rozeva, “Using microservices and event driven architecture for big data stream processing,” in Proc. 45th Int. Conf. Appl. Math. Eng. Econ. (AMEE), vol. 2172, 2019, Art. no. 090010.

PING WANG received the Ph.D. degree from the National University of Defense Technology, Changsha, China. He is currently a Professor with the National Key Laboratory of Science and Technology on ATR, College of Electronic Science, National University of Defense Technology. His research interests include big data processing, mobile cloud computing, and image processing.

LUOBING DONG received the Ph.D. degree from Xidian University. He was a Visiting Scholar with The University of Texas at Dallas. He is currently an Associate Professor with the School of Computer Science and Technology, Xidian University. His research interests include big data processing, mobile cloud computing, and document summarization.

YUESHEN XU received the Ph.D. degree from Zhejiang University and a co-trained Ph.D. Student at University of Illinois, Chicago. He is currently a Lecturer with the School of Computer Science and Technology, Xidian University. He has published more than thirty articles in international journals or conferences, such as ESWA, EAAI, WISE, WAIM, and Sensors. His research interests include recommender systems, text mining, and service computing. He is also a member of ACM and CCF. He is also a Reviewer of several journals and conferences, including the IEEE TRANSACTIONS ON SERVICES COMPUTING, KBS, IEEE ACCESS, INCA, Neurocomputing, and ICDCS.

WEI LIU received the master’s degree in engineering from Xidian University. He is currently an Associate Professor with the School of Computer Science and Technology, Xidian University. His research interests include recommender systems, text mining, and big data processing.

NINGNING JING is currently pursuing the master’s degree with the School of Computer Science and Technology, Xidian University. His research interests include big data processing, mobile cloud computing, and document summarization.

* * *