Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2009, Article ID 802932, 16 pages
doi:10.1155/2009/802932

Research Article

Self-Learning Facial Emotional Feature Selection Based on Rough Set Theory

Yong Yang,[1,2] Guoyin Wang,[1] and Hao Kong[1]

[1] *Institute of Computer Science & Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China*
[2] *School of Information Science and Technology, Southwest Jiaotong University, Chengdou 610031, China*

Correspondence should be addressed to Guoyin Wang, wanggy@ieee.org

Received 16 January 2009; Revised 15 April 2009; Accepted 12 May 2009

Recommended by Panos Liatsis

Emotion recognition is very important for human-computer intelligent interaction. It is generally performed on facial or audio information by artificial neural network, fuzzy set, support vector machine, hidden Markov model, and so forth. Although some progress has already been made in emotion recognition, several unsolved issues still exist. For example, it is still an open problem which features are the most important for emotion recognition. It is a subject that was seldom studied in computer science. However, related research works have been conducted in cognitive psychology. In this paper, feature selection for facial emotion recognition is studied based on rough set theory. A self-learning attribute reduction algorithm is proposed based on rough set and domain oriented data-driven data mining theory. Experimental results show that important and useful features for emotion recognition can be identified by the proposed method with a high recognition rate. It is found that the features concerning mouth are the most important ones in geometrical features for facial emotion recognition.

1. Introduction

In recent years, there has been a growing interest in improving all aspects of the interactions between humans and computers. It is argued that to truly achieve effective human-computer intelligent interaction (HCII), there is a requirement for computers to be able to interact naturally with users, similarly to the way human-human interaction. HCII is becoming more and more important in such applications as smart home, smart office, and virtual reality, and it will be popular in all aspects of daily life in the future. To achieve the purpose of HCII, it is essential for computers to recognize human emotion and to give a suitable feedback. Consequently, emotion recognition attracts significant attention in both industry and academia. There are several research works in this field in recent years and some successful products such as AIBO, the popular robot dog produced by Sony. Usually, emotion recognition is studied by the methods of artificial neural network (ANN), fuzzy set, support

vector machine (SVM), hidden Markov model (HMM), and based on the facial or audio features, and the recognition rate often arrives at 64% to 98% [1–3]. Although some progress has been made in emotion recognition, several unsolved issues still exist. For example, it is still an open problem which features are the most important for emotion recognition. It is a subject that was seldom studied in computer science. However, related research works have been conducted in cognitive psychology [4–6].

There have been several research works related to the important features for emotion in cognitive psychology. Based on the results of psychological experiments, Sui and Ren argue that the information conveyed by different facial parts has diverse effects on the facial expression recognition, and the eyes play the most important role [4]. Wang and Fu argue that the low spatial frequency information is important for emotion [5]. White argues that edge-based facial information is used for expression recognition [6].

In our previous works of emotion recognition in [7–10], attribute reduction algorithms based on classical rough set are used for the purpose of facial emotional feature selection, and SVM is taken as the classifiers. Some useful features concerning eyes and mouth are found. Based on these features, high correct recognition rates are achieved. However, classical rough set theory is based on equivalence relation. There must be a process of discretization in equivalence relation since the measured facial features are continuous values. Information might be lost or changed in the discretization process, thereby affecting the result. To solve this problem, some research works have been taken. Shang et al. proposed a new attribute algorithm, which integrates the discretion and reduction using information entropy-based uncertainty measures and evolutionary computation [11]. Jensen and Shen proposed a fuzzy-rough attribute reduction algorithm and an attribute reduction algorithm based on tolerance relation [12]. Although these research works can avoid the discretization process, the parameters in these methods should be given according to prior experience of domain experts, for example, the fuzzy set membership function in Jensen's fuzzy-rough attribute reduction algorithm, the population amount for Shang's method. If there is no experience of domain experts, these methods will be useless in some extent. In this paper, a novel feature selection method based on tolerance relation is proposed, which can avoid the process of discretization. Meantime, based on the idea of domain-oriented data-driven data mining (3DM), a method for finding suitable threshold of tolerance relation is introduced. Experimental results show that important and useful features for emotion recognition can be identified by the proposed method with a high recognition rate. It is found that the features concerning mouth are the most important ones in geometrical features for facial emotion recognition.

The rest of this paper is organized as follows. In Section 2, a novel feature selection method for emotion recognition based on rough set theory is introduced. Simulation results and discussion are given in Section 3. Finally, conclusions and future works are presented in Section 4.

2. Feature Selection for Emotion Recognition Based on Rough Set Theory

2.1. Basic Concepts of Rough Set Theory

Rough set (RS) is a valid mathematical theory for dealing with imprecise, uncertain, and vague information; it was developed by Professor Pawlak in 1980s [13, 14]. RS has been

successfully used in many domains such as machine learning, pattern recognition, intelligent data analyzing, and control algorithm acquiring [15–17]. The most advantage of RS is its great ability of attribute reduction (knowledge reduction, feature selection). Some basic concepts of rough set theory are introduced here for the convenience of the following discussion.

Definition 2.1. A decision information system is defined as a quadruple $S = (U, C \cup D,\ V,\ f)$, where U is a finite set of objects, C is the condition attribute set, and $D = \{d\}$ is the decision attribute set. For all $c \in C$, with every attribute $a \in C \cup D$, a set of its values Va is associated. Each attribute a determines a function $fa : U \to Va$.

Definition 2.2. For a subset of attributes $B \subseteq A$, an indiscernibility relation is defined by $\mathrm{Ind}(B) = \{(x, y) \in U \times U : \forall_{a \in B}(a_x = a_y)\}$, in which a_x and a_y are values of the attribute a of x and y.

The indiscernibility relation defined in this way is an equivalence relation. Obviously, $\mathrm{Ind}(B) = \cap_{b \in B}\mathrm{Ind}(\{b\})$. By $U/\mathrm{Ind}(B)$ we mean the set of all equivalence classes in the relation $\mathrm{Ind}(B)$. The classical rough set theory is based on an observation that objects may be indiscernible due to limited available information, and the indiscernibility relation defined in this way is an equivalence relation indeed. The intuition behind the notion of an indiscernibility relation is that selecting a set of attribute $B \subseteq A$ effectively defines a partition of the universe into sets of objects that cannot be discerned using the attributes in B only. The equivalence classes $E_i \in U/\mathrm{Ind}(B)$, induced by a set of attributes $B \subseteq A$, are referred to as object classes or simply classes. The classes resulted from $\mathrm{Ind}(A)$ and $\mathrm{Ind}(D)$ are called condition classes and decision classes, respectively.

Definition 2.3. A decision information system is a continuous value information system, and it is defined as a quadruple $s = (U,\ C \cup D,\ V,\ f)$, where U is a finite set of objects, C is the condition attribute set, and $D = \{d\}$ is the decision attribute set. For all $c \in C$, c is continuous value attribute.

A facial expression information system is a continuous value information system according to Definition 2.3.

If a condition attribute value is a continuous value, indiscernibility relation cannot be used directly since it requires that the condition attribute values of two different samples are equal, which is difficult to satisfy. Consequently, a process of discretization must be taken, in which information may be lost or changed. The result of attribute reduction would be affected. Since all measured facial attributes are continuous value and imprecise to some extent, the process of discretization may affect the result of emotion recognition. We argue that it is suitable for the continuous value information systems that the attribute values are taken as equal if they are similar in some range. Based on this idea, a method based on tolerance relation that avoids the process of discretization is proposed in this paper.

Definition 2.4. A binary relation $R(x, y)$ defined on an attribute set B is called a tolerance relation if it satisfies

(1) symmetrical: $\forall_{x,y \in U}(R(x, y) = R(y, x))$;

(2) reflexive: $\forall_{x \in U}(R(x, x) = 1)$.

From the standpoint of a continuous value information system, a relation could be set up for a continuous value information system as follows.

Definition 2.5. Let an information system $S = (U, C \cup D, V, f)$ be a continuous value information system; a relation $R(x, y)$ is defined as

$$R(x,y) = \{(x,y) \mid x \in U \wedge y \in U \wedge \forall_{a \in C}(|a_x - a_y| \leq \varepsilon, 0 \leq \varepsilon \leq 1)\}. \tag{2.1}$$

Apparently, $R(x, y)$ is a tolerance relation according to Definition 2.4 since $R(x, y)$ is symmetrical and reflextive. In classical rough set theory, an equivalence relation constitutes a partition of U, but a tolerance relation constitutes a cover of U, and equivalence relation is a particular type of tolerance relation.

Definition 2.6. Let $R(x, y)$ be a tolerance relation based on Definition 2.5, $n_R(x_i) = \{x_j \mid x_j \in U \wedge \forall_{a \in C}(|a_{x_i} - a_{x_j}| \leq \varepsilon)\}$ is called a tolerance class of x_i, and $|n_R(x_i)| = |\{x_j \mid x_j \in n_R(x_i), 1 \leq j \leq U\}|$ is the cardinal number of the tolerance class of x_i.

According to Definition 2.6, for all $x \in U$, the bigger the tolerance class of x is, the more uncertainty it will be and the less knowledge it will contain. On the contrary, the smaller the tolerance class of x is, the less uncertainty it will be and the more knowledge it will contain. Accordingly, the concept of knowledge entropy and conditional entropy could be defined as follows.

Definition 2.7. Let $U = \{x_1, x_2, \ldots, x_{|U|}\}$, $R(x, y)$ be a tolerance relation; the knowledge entropy $E(R)$ of relation R is defined as

$$E(R) = -\frac{1}{|U|}\sum_{i=1}^{|U|} \log_2 \frac{|n_R(x_i)|}{|U|}. \tag{2.2}$$

Definition 2.8. Let R and Q be tolerance relations defined on U, a relation satisfying R and Q simultaneous can be taken as $R \cup Q$, and it is a tolerance relation too. For all $x_i \in U$, $n_{R \cup Q}(x_i) = n_R(x_i) \cap n_Q(x_i)$; therefore, the knowledge entropy of $R \cup Q$ can be defined as

$$E(R \cup Q) = -\frac{1}{|U|}\sum_{i=1}^{|U|} \log_2 \frac{|n_{R \cup Q}(x_i)|}{|U|}. \tag{2.3}$$

Definition 2.9. Let R and Q be tolerance relations defined on U; the conditional entropy of R with respect to Q is defined as $E(Q \mid R) = E(R \cup Q) - E(R)$.

Let $S = (U, C \cup D, V, f)$ be a continuous value information system, let relation K be a tolerance relation defined on its condition attribute set C, and let relation L be an equivalence

relation (a special tolerance relation) defined on its decision attribute set D. According to Definitions 2.7, 2.8, and 2.9, we can get

$$\begin{aligned} E(D \mid C) &= E(L \mid K) \\ &= E(K \cup L) - E(K) \\ &= -\frac{1}{|U|}\sum_{i=1}^{|U|}\log_2\frac{|n_{K\cup L}(x_i)|}{|U|} - \left(-\frac{1}{|U|}\sum_{i=1}^{|U|}\log_2\frac{|n_K(x_i)|}{|U|}\right) \\ &= -\frac{1}{|U|}\sum_{i=1}^{|U|}\log_2\frac{|n_{K\cup L}(x_i)|}{|n_K(x_i)|}, \end{aligned} \tag{2.4}$$

where the conditional entropy $E(D \mid C)$ has a clear meaning; that is, it is a ratio between the knowledge of all attributes (condition attribute set plus decision attribute set) and the knowledge of the condition attribute set.

2.2. Feature Selection Based on Rough Set Theory and Domain-Oriented Data-Driven Data Mining

In this section, a novel attribute reduction algorithm is proposed based on rough set theory and domain-oriented data-driven data mining (3DM) [18, 19].

3DM is a data mining theory proposed by Wang [18, 19]. According to the theory, knowledge could be expressed in different ways; that is, some relationship exists between the different formats of the same knowledge. In order to keep the knowledge unchanged in a data mining process, the properties of the knowledge should remain unchanged during the knowledge transformation process [20]. Otherwise, mistake may occur in the process of knowledge transformation. Based on this understanding, knowledge reduction can be seen as a process of knowledge transformation, in which properties of the knowledge should be remained.

In the application of emotion recognition, no faces are entirely the same nor are emotions. For any two different emotion samples, there must be some different features in the samples. Accordingly, an emotion sample belongs to an emotion state according to its features which are different to the others. From this standpoint, we argue that the discernability of the condition attribute set with respect to the decision attribute set can be taken as an important property of knowledge in the course of knowledge acquisition in emotion recognition. Based on the idea of 3DM, the discernability should be unchanged in the process of knowledge acquisition and attribute reduction.

Definition 2.10. Let $S = (U, C \cup D, V, f)$ be a continuous value information system. If $\forall_{x_i,x_j\in U}(d_{x_i} \neq d_{x_j} \rightarrow \exists_{a\in C}(a_{x_i} \neq a_{x_j}))$, it is certainly discernable for the continuous value information system S.

The discernability is taken as a fundamental ability that a continuous information system has in this paper. According to 3DM, the discernability should be unchanged if feature selection is done for a continuous value information system. From Definition 2.10, we can have $\forall_{x_i,x_j\in U}(d_{x_i} \neq d_{x_j} \rightarrow \exists_{a\in C}(|a_{x_i} - a_{x_j}| > \varepsilon))$. Therefore, according to Definition 2.6, we can

have $\forall_{x_i,x_j\in U}(x_j \notin n_R(x_i) \wedge x_i \notin n_R(x_j) \rightarrow n_R(x_i) \neq n_R(x_j))$. Accordingly, the discernability of a tolerance relation can be defined as follows.

Definition 2.11. Let $R(x, y)$ be a tolerance relation according to Definition 2.5; if $\forall_{x_i,x_j\in U}(d_{x_i} \neq d_{x_j} \rightarrow n_R(x_i) \neq n_R(x_j))$, $R(x, y)$ has the certain discernability.

If $R(x, y)$ has certain discernability, according to Definition 2.11, $\forall_{x_i,x_j\in U}(n_R(x_i) = n_R(x_j) \rightarrow d_{x_i} = d_{x_j})$, therefore, $\forall_{x_i,x_j\in U}(x_i, x_j \in n_R(x_i) \rightarrow d_{x_i} = d_{x_j})$.

Theorem 2.12. *$E(D \mid C) = 0$ is a necessary and sufficient condition of that there is certain discernability for the condition attribute set with respect to the decision attribute set in tolerance relation.*

Proof. Let $S = (U, R, V, f)$ be a continuous value information system, let relation K be a tolerance relation defined on condition attribute set C, and let relation L be an equivalence relation (a special tolerance relation) defined on decision attribute set D.

Necessity

If there is certain discernability for the condition attribute set with respect to the decision attribute set in tolerance relation, according to Definition 2.11, $\forall_{x_i,x_j\in U}(x_i, x_j \in n_K(x_i) \rightarrow d_{x_i} = d_{x_j})$, then

$$\begin{aligned} &n_K(x_i) \subseteq n_L(x_i), \qquad n_{K\cup L}(x_i) = n_K(x_i), \qquad |n_{K\cup L}(x_i)| = |n_K(x_i)|, \\ &E(D \mid C) = E(L \mid K) = -\frac{1}{|U|}\sum_{i=1}^{|U|}\log_2\frac{|n_{K\cup L}(x_i)|}{|n_K(x_i)|} = -\frac{1}{|U|}\sum_{i=1}^{|U|}\log_2 1 = 0. \end{aligned} \tag{2.5}$$

Sufficiency

For all $x_i \in U$, we can have $n_{K\cup L}(x_i) \subseteq n_K(x_i)$, $|n_{K\cup L}(x_i)| \leq |n_K(x_i)|$. Since $E(D \mid C) = E(L \mid K) = -(1/|U|)\sum_{i=1}^{|U|}\log_2(|n_{K\cup L}(x_i)|/|n_K(x_i)|) = 0$, we can have $\forall x_i \in U$, $|n_{K\cup L}(x_i)| = |n_K(x_i)|$, that is, $n_{K\cup L}(x_i) = n_K(x_i)$. Therefore, the decision values should be equal for the different samples included in the same tolerance class. Accordingly, we can have $\forall_{x_i,x_j\in U}(x_i, x_j \in n_R(x_i) \rightarrow d_{x_i} = d_{x_j})$, therefore, $\forall_{x_i,x_j\in U}(d_{x_i} \neq d_{x_j} \rightarrow \exists_{a\in C}(a_{x_i} \neq a_{x_j}))$, and there is certain discernability for condition attribute set with respect to decision attribute set in tolerance relation. This completes the proof. □

From Theorem 2.12, $E(D \mid C) = 0$ can be taken as a measurement for $R(x, y)$ has certain discernability.

For a given continuous value information system S, there could be many different tolerance relations according to different threshold ε under the condition $E(D \mid C) = 0$, but the biggest granular and the best generalization are always required for knowledge acquisition. According to the principle, we can have the following results.

If the threshold ε in tolerance relation is 0, then the tolerance class $n_R(x_i)$ of an instance x_i contains x_i itself only, and we can have $n_{R\cup Q}(x_i) = n_R(x_i) = \{x_i\}$, and $E(D \mid C) = 0$. It is the smallest tolerance class for the tolerance relation, the smallest knowledge granular, and the smallest generalization.

If the threshold ε in tolerance relation is increased from 0, both $n_R(x_i)$ and $n_{R\cup Q}(x_i)$ are increased. If $n_R(x) \subseteq n_Q(x)$, then, $n_{R\cup Q}(x_i) = n_R(x_i)$, $|n_{R\cup Q}(x_i)| = |n_R(x_i)|$, $E(D \mid C) = 0$, and the granular of knowledge is increased.

If the threshold ε in tolerance relation is increased to a critical point named ε_{opt}, both $n_R(x_i)$ and $n_{R\cup Q}(x_i)$ are increased, and $n_{R\cup Q}(x_i) = n_R(x_i)$, $|n_{R\cup Q}(x_i)| = |n_R(x_i)|$, $E(D \mid C) = 0$, and the granular of knowledge is the biggest under the condition that the certain discernability of condition attribute set with respect to decision attribute set in tolerance relation is unchanged.

If the threshold ε in tolerance relation is increased from ε_{opt} and $\varepsilon < 1$, then $n_{R\cup Q}(x_i) \neq n_R(x_i)$, $|n_{R\cup Q}(x_i)| \neq |n_R(x_i)|$, $E(D \mid C) \neq 0$, and then the certain discernability is changed. If $\forall_{x_i \in U}(n_Q(x_i) \subset n_R(x_i))$, then $n_{R\cup Q}(x_i) = n_Q(x_i)$, $|n_{R\cup Q}(x_i)| = |n_Q(x_i)|$, and $|n_Q(x_i)| < |n_R(x_i)|$. Therefore

$$\begin{aligned} E(D \mid C) &= E(Q \mid R) \\ &= -\frac{1}{|U|}\sum_{i=1}^{|U|} \log_2 \frac{\left|n_{R\cup Q}(x_i)\right|}{|n_R(x_i)|} \\ &= -\frac{1}{|U|}\sum_{i=1}^{|U|} \log_2 \frac{\left|n_Q(x_i)\right|}{|n_R(x_i)|} \\ &> 0. \end{aligned} \tag{2.6}$$

Since $|n_Q(x_i)|$ is held and $|n_R(x_i)|$ is increased with the threshold of ε increase, $E(D \mid C)$ is increased.

If the threshold ε in tolerance relation is increased to $\varepsilon = 1$, then $n_R(x_i) = U$ and $\forall_{x_i \in U}(n_Q(x_i) \subseteq n_R(x_i))$, $n_{R\cup Q}(x_i) = n_Q(x_i)$, $|n_{R\cup Q}(x_i)| = |n_Q(x_i)|$, so,

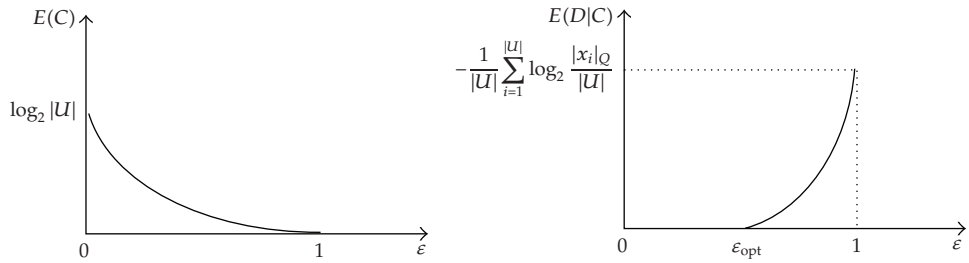
$$\begin{aligned} E(D \mid C) &= E(Q \mid R) \\ &= -\frac{1}{|U|}\sum_{i=1}^{|U|} \log_2 \frac{\left|n_{R\cup Q}(x_i)\right|}{|n_R(x_i)|} \\ &= -\frac{1}{|U|}\sum_{i=1}^{|U|} \log_2 \frac{\left|n_Q(x_i)\right|}{|U|}. \end{aligned} \tag{2.7}$$

Since the equivalence class of Q is held, $E(D \mid C)$ is constant.

The relationship between entropy, condition entropy and ε can be shown in Figure 1.

From Figure 1 and the discussion above, if the threshold value of ε take ε_{opt}, it will make $E(D \mid C) = 0$, and therefore, the certain classification ability of condition attribute set with respect to decision attribute set will be unchanged. At the same time, the tolerance class of x is the biggest. In a sense, the knowledge granular is the biggest in ε_{opt}, and then, the generalization should be the best.

In summary, parameter selection of ε is discussed, and based on 3DM, a suitable threshold value of ε, ε_{opt}, is found. It can keep the classification ability of condition attribute set with respect to decision attribute set, and at the same time, it can keep the generalization the most. It is predominant for the course of finding ε_{opt} since the method is based on



(a) Relationship between $E(C)$ and ε (b) Relationship between $E(D \mid C)$ and ε

Figure 1: Relationship between entropy, condition entropy, and ε.

data only and dose not need experiences of domain experts. Therefore, the method is more robustness.

In this paper, the threshold of ε_{opt} is searched in $[0,1]$ based on binary search algorithm.

2.3. Attribute Reduction for Emotion Recognition

The discernability of condition attribute set with respect to decision attribute set in tolerance relation is a fundamental feature of knowledge of a continuous value information system. The discernability should be unchanged according to 3DM. Since $E(D \mid C) = 0$ is a necessary and sufficient condition for keeping the discernability of condition attribute set with respect to decision attribute set in tolerance relation, therefore, a self-learning attribute reduction algorithm (SARA) is proposed for continuous value information systems as follows.

Algorithm 2.13 (Self-learning attribute reduction algorithm (SARA)).

Input: a decision table $S = (U, C \cup D, V, f)$ of a continuous information system, where U is a finite set of objects, C is the condition attribute set, and $D = \{d\}$ is the decision attribute set.

Output: a relative reduction B of S.

Step 1. Compute ε_{opt}, then set up a tolerance relation on the condition attribute set C.

Step 2. Compute condition entropy $E(D \mid C)$.

Step 3. For all $a_i \in C$, compute $E(D \mid \{a_i\})$. Sort a_i according to $E(D \mid \{a_i\})$ descendant.

Step 4. Let $B = C$, deal with each a_i as in the following.

Substep 4.1

Compute $E(D \mid B - \{a_i\})$.

Substep 4.2

If $E(D \mid C) = E(D \mid B - \{a_i\})$, attribute a_i should be reduced, and $B = B - \{a_i\}$, otherwise, a_i could not be reduced, and B is holding.

Table 1: Three facial emotional datasets.

Dataset name	Samples	People	Emotion classes
CKACFE	405	97	Happiness, sadness, surprise, anger, disgust, fear, neutral
JAFFE	213	10	Happiness, sadness, surprise, anger, disgust, fear, neutral
CQUPTE	652	8	Happiness, sadness, surprise, anger, disgust, fear, neutral

(a) Some images of CKACFE database

(b) Some images of JAFFE database

(c) Some images of CQUPTE database

Figure 2: Facial emotion samples.

Let $|U| = n$, $|C| = m$. The time complexity of Step 1 is $O(n)$, the time complexity of Step 2 is $O(mn^2)$, the time complexity of Step 3 is $O(mn^2)$, the time complexity of Step 4 is $O(m^2n^2)$, and therefore, the time complexity of the algorithm is $O(m^2n^2)$.

3. Experiment Results and Discussion

Since there are few open facial emotional dataset, three facial emotional datasets are used in the experiments. The first dataset comes from the Cohn-Kanade AU-Coded Facial Expression (CKACFE) database [21], and the dataset is more representative of Caucasian to some extent. The second one is the Japanese female facial expression (JAFFE) database [22], and it is more representative of Asian women. The third one named CQUPTE [23] is collected from 8 graduate students in the Chongqing University of Posts and Communications in China, in which there are four females and four males. Details of the datasets are listed in Table 1.

Some samples are shown in Figure 2. In each dataset, the samples are happiness, sadness, fear, disgust, surprise, and angry from left to right in Figure 2.

Facial expression of human being is expressed by the shape and position of facial components such as eyebrows, eyes, mouth, and nose. The geometric features, appearance features, wavelet features, and mixture features of facial are popular for emotion recognition in recent years. The geometric facial features represent the shape and locations of facial

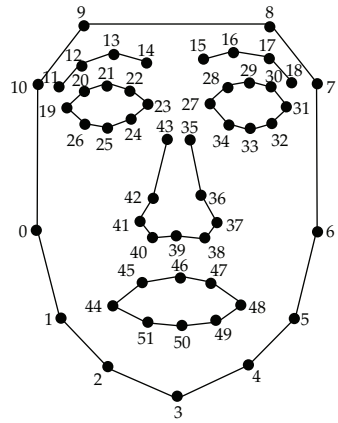


Figure 3: 52 feature points according to FAP parameters.

components, and it is used in the experiments since it is obvious and intuitionistic for the facial expression. The geometric facial features are the distance between two different feature points which are according to a defined criterion. The MPEG-4 standard is a popular standard for feature point selection. It extends facial action coding system (FACS) to derive facial definition parameters (FDP) and facial animation parameters (FAP). There are 68 FAP parameters, in which 66 low parameters are defined according to FDP parameters to describe the motion of a human face. The FDP and low-level FAP can constitute a concise representation of a face, and they are adequate for basic emotion recognition because of the varieties of expressive parameter. In the experiments, 52 low FAP parameters are chosen to represent emotion because some FAP parameters have little effect on facial expression. For example, the FAP parameter named raise_l_ear, which denotes the vertical displacement of left ear. Thus, a feature point set including 52 feature points is defined as shown in Figure 3. Based on the feature points, 33 facial features are extracted for emotion recognition according to [4–7] and listed in Table 2. The 33 facial features can be divided into three groups. There are 17 features in the first group which concern eyes and consists of d_0, d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_{16}, d_{17}, d_{19}, d_{20}, d_{25}, d_{26}, d_{27}, d_{28}, and d_{29}; there are 6 features in the second group which concern cheek and consists of d_9, d_{10}, d_{18}, d_{21}, d_{30}, and d_{31}; there are 10 features in the third group which concern mouth and consists of d_8, d_{11}, d_{12}, d_{13}, d_{14}, d_{15}, d_{22}, d_{23}, d_{24} and d_{32}. In Table 1, A is the midpoint of point 19 and 23, and B is the midpoint of point 27 and 31. dis(i, j) denotes the Euclid distance between point i and j; hei(i, j) denotes the horizontal distance between point i and j; wid(i, j) denotes the vertical distance between i and j. Since the distance between point 23 and 27 is stable for all kinds of expression, we normalize the distance features in the following way.

Firstly, $x_i' = d_i/d$, $i = 0, 1, \ldots, 32$, d is the distance between point 23 and 27.

Secondly, the normalized distance is calculated using the following formula:

$$x_i = \frac{x_i' - \min(x_i')}{\max(x_i') - \min(x_i')}, \quad x_i \in [0, 1]. \tag{3.1}$$

Table 2: 33 features defined on 52 feature points.

feature	description	feature	description	feature	description
d_0	dis(11, 19)	d_{11}	dis(39, 44)	d_{22}	dis(44, 48)/2
d_1	dis(18, 31)	d_{12}	dis(39, 48)	d_{23}	dis(45, 51)
d_2	dis(21, 25)	d_{13}	dis(44, 48)	d_{24}	dis(47, 49)
d_3	dis(20, 26)	d_{14}	dis(46, 50)	d_{25}	dis(14, 23)
d_4	dis(22, 24)	d_{15}	dis(39, 3)	d_{26}	dis(15, 27)
d_5	dis(29, 33)	d_{16}	dis(21, A)	d_{27}	dis(19, 23)/2
d_6	dis(28, 34)	d_{17}	dis(A, 25)	d_{28}	dis(27, 31)/2
d_7	dis(30, 32)	d_{18}	hei(A, 44)	d_{29}	(wid(19, 23) + wid(27, 31))/2
d_8	dis(39, 46)	d_{19}	dis(29, B)	d_{30}	(hei(11, 39) + hei(18, 39))/2
d_9	dis(23, 44)	d_{20}	dis(B, 33)	d_{31}	(hei(14, 39) + hei(15, 39))/2
d_{10}	dis(27, 48)	d_{21}	hei(B, 48)	d_{32}	(hei(44, 39) + hei(48, 39))/2

3.1. Experiments For SARA as a Feature Selection Method for Emotion Recognition

In this section, there are five comparative experiments to test the effectiveness of SARA as a method of feature selection for emotion recognition.

In the first experiment, SARA is taken as the method of feature selection for emotion recognition. In the second one, an attribution reduction algorithm named CEBARKNC [24] is taken as a method of feature selection for emotion recognition. CEBRKNC is selected in this comparative experiment since it is an attribute reduction algorithm based on conditional entropy in equivalence relation. In this experiment, a greedy algorithm proposed by Nugyen [25] is taken as a discretization method, and it is done on the platform RIDAS [26]. In the third experiment, an attribute reduction algorithm named MIBARK [27] is taken as a method of feature selection. It is a reduction algorithm based on mutual-information as the measure of importance of attribute. And a greedy algorithm proposed by Nugyen [25] is taken as a discretization method, and it is done on the platform RIDAS also. In the fourth experiment, a traditional feature selection method, Genetic Algorithm (GA) [28], is used as the feature selection method for emotion recognition. This experiment is done on WEKA [29], a famous machine learning tool, and CfsSubsetEval is taken as the evaluator for feature selection in WEKA. In the fifth experiment, all the 33 features are used for emotion recognition, and the feature selection course is omitted, SVM is a new machine learning method, and it is famous for its great ability for small samples applications. Therefore, SVM are taken as classifiers for all the comparative experiments. SVM are given same parameters in all the experiments. 4-fold cross-validation is taken for all the experiments.

The results of the comparative experiments are shown in Table 3. CRR is the percentage of the correct recognition rate, and RAN is the number of attributes after attribute reduction.

From the experiment results of SARA + SVM and SVM from Table 3, we can find that SARA can use nearly one third features and get nearly the same correct recognition rate; therefore, SARA can be taken as a useful feature selection method for emotion recognition. When we compare the experimental results of SARA + SVM and CEBARKNC + SVM from Table 3, we can find SARA selects as much features as CEBARKNC, but SARA gets a better correct recognition rate than CEBARKNC. Furthermore, from the comparative experiment

Table 3: The results of comparative experiments on the three dataset.

Database	SARA + SVM		CEBARKNC + SVM		MIBARK + SVM		GA + SVM		SVM	
	CRR	RAN	CRR	RAN	CRR	RAN	CRR	RAN	CRR	RAN
CKACFE	76.01	11.25	73.07	12.5	75.05	17.75	73.09	14.25	79.80	33
JAFFE	69.37	11.5	63.17	11	63.98	14.5	55.89	14.25	74.46	33
CQUPTE	92.45	14	78.83	13.5	87.90	13.5	88.95	14.75	93.86	33
average	79.28	12.25	71.69	12.33	75.64	15.25	72.64	14.42	82.71	33

Table 4: The common features in the three datasets.

Database	SARA	CEBARKNC	MIBARK	GA
CKACFE	x_{13}, x_{14}, x_{15}	x_1, x_8, x_{13}, x_{16}, x_{21}	x_1, x_8, x_{16}, x_{17}, x_{22}, x_{28}, x_{30}, x_{31}, x_{32}	x_{14}, x_{25}
JAFFE	x_5, x_{13}, x_{14}, x_{15}, x_{26}	x_0, x_8, x_{13}	x_1, x_{15}, x_{24}, x_{26}, x_{32}	x_7, x_{25}, x_{32}
CQUPTE	x_7, x_{13}, x_{14}, x_{15}, x_{24}, x_{30}, x_{31}	x_0, x_1, x_4, x_8, x_{13}, x_{26}, x_{27}, x_{32}	x_0, x_1, x_4, x_8, x_{13}, x_{23}, x_{24}, x_{32}	x_{11}, x_{14}, x_{22}, x_{23}, x_{24}, x_{32}
Common features	x_{13}, x_{14}, x_{15}	x_8, x_{13}	x_1, x_{32}	—

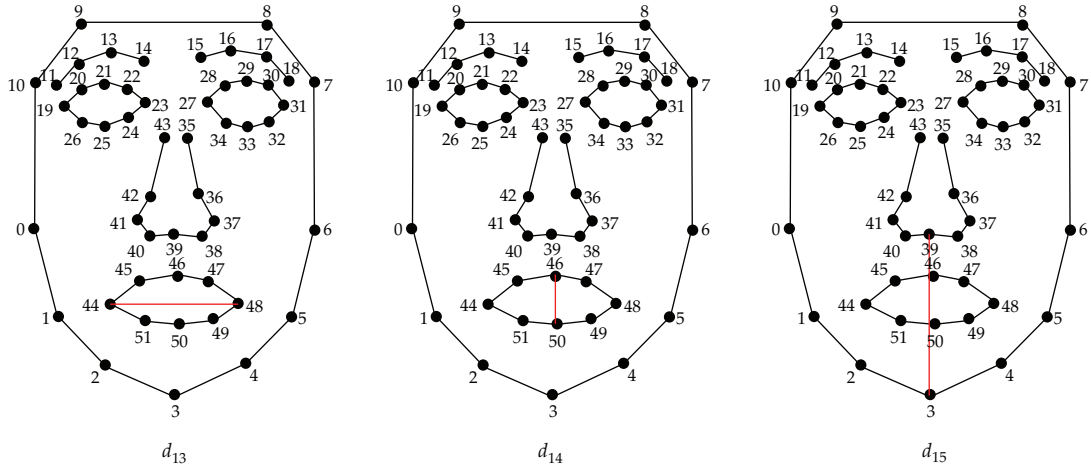
results between SARA + SVM and MIBARK + SVM, or experimental results between SARA + SVM and GA + SVM from Table 3, we can find that SARA can use fewer features than MIBARK or GA but get higher recognition rate. Therefore, SARA can be taken as an effective feature selection method for emotion recognition than CEBARKNC, MIBARK, and GA, since the features selected by SARA have better discernability in emotion recognition.

Common features reserved by the four feature selection methods are listed in Table 4.

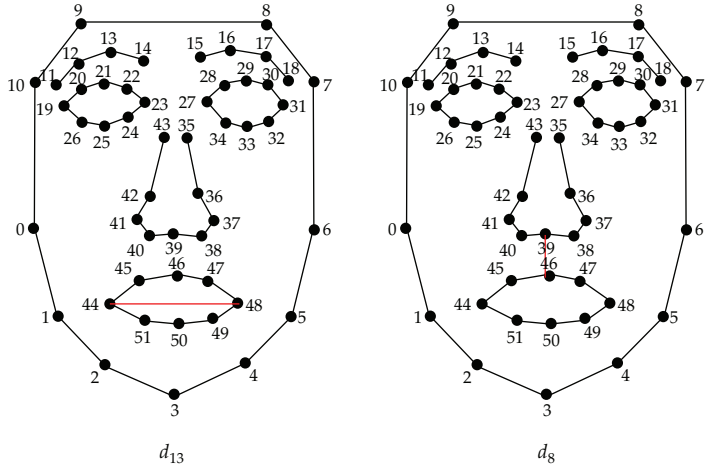
From Table 4, we can find that the four feature selection algorithms can select different features for emotion recognition. Among all the experiment results, SARA selects three common features, x_{13}, x_{14}, and x_{15} for all the three emotion datasets, meanwhile, CEBARKNC selects two common features, x_8 and x_{13}, and MIBARK selects two common features, x_1 and x_{32}; however, GA cannot find any common feature for all the three datasets. Since better correct recognition rate can be achieved if SARA is used as a method of feature selection for emotion recognition, therefore, x_{13}, x_{14}, x_{15} can be seen more important for emotion recognition. Although the features of x_{13}, x_{14}, x_{15} are normalized features, the importance of original features of d_{13}, d_{14}, d_{15} is also evident. Since the features of d_{13}, d_{14}, d_{15} are all concerning mouth, therefore, we can draw a conclusion that the geometrical features concerning mouth are the most important features for emotion recognition. The original selected features of SARA, CEBARKNC, and MIBARK are shown in Figure 4.

3.2. *Experiments for the Features Concerning Mouth for Emotion Recognition*

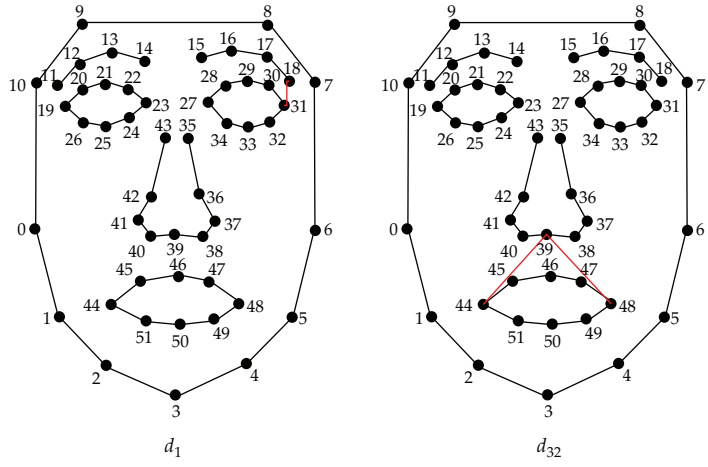
From the last section, we draw a conclusion that the geometrical features concerning mouth are important for emotion recognition. In this section, there are four experiments for the purpose of testing the importance of the geometrical feature concerning mouth for emotion recognition. In the first experiment, all the 33 facial features are used for emotion recognition. In the second experiment, only the features selected by SARA are used for



(a) Common features selected by SARA



(b) Common features selected by CEBARKNC



(c) Common features selected by MIBARK

Figure 4: Common features.

Table 5: The results of comparative experiments on the three dataset.

	SARA reserved		ALL features		No mouth		No eyes	
	CRR	RAN	CRR	RAN	CRR	RAN	CRR	RAN
CKACFE	76.01	11.25	79.80	33	45.18	19	75.15	12
JAFFE	69.37	11.5	74.46	33	53.32	19	58.29	12
CQUPTE	92.45	14	93.86	33	69.02	19	89.72	12
average	79.28	12.25	82.71	33	55.84	19	74.39	12

emotion recognition. In the third experiment, all the features concerning mouth are deleted, and there are 19 features that are used for emotion recognition, in which there are 17 features concerning eyes d_0, d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_{16}, d_{17}, d_{19}, d_{20}, d_{25}, d_{26}, d_{27}, d_{28}, d_{29}, and two features d_{30}, d_{31} concerning cheek but not mouth. In the fourth experiment, all the features concerning eyes are deleted, and there are 12 features that are used for emotion recognition, in which there are 10 features concerning mouth d_8, d_{11}, d_{12}, d_{13}, d_{14}, d_{15}, d_{22}, d_{23}, d_{24}, d_{32}, and two features d_{30}, d_{31} concerning cheek but not eyes. SVM is taken as classifier in the four experiments and is given the same parameters. Experiment results are listed in Table 5.

From Table 5, we can find that the correct recognition rate is decreased greatly if there is no feature concerning mouth. Therefore, it is concluded that the features concerning mouth are the most important geometrical features for emotion recognition. On the other hand, we can find that the correct recognition rate is not affected so much if there are no features concerning eyes. Therefore, the geometrical features concerning eyes do not play an important role in emotion recognition. But from the psychological experiments of [4], Sui and Ren found that the eyes play an important role in emotion; therefore, we may draw a conclusion that the geometrical features concerning mouth are the most important in the geometrical features for emotion recognition, and the geometrical features concerning eyes are not so important. Furthermore, the important features concerning eyes for emotion recognition should be discovered and used in emotion recognition in the further work. Meanwhile, we can find that the correct recognition rate is decreased in CKACFE more than in JAFFE and CQUPTE. Therefore, we can draw a conclusion that the geometrical features concerning mouth are more important for emotion expression for the Caucasian than the eastern people.

4. Conclusion

In this paper, based on rough set theory and the idea of domain oriented data driven data mining, a novel attribute reduction algorithm named SARA is proposed for feature selection for emotion recognition. The proposed method is found to be effective and efficient, and the geometrical features concerning mouth are found to be the most important geometrical features for emotion recognition.

Acknowledgment

This paper is partially supported by the National Natural Science Foundation of China under Grants no. 60773113, Natural Science Foundation of Chongqing under Grants no. 2007BB2445, no. 2008BA2017, and no. 2008BA2041.

References

[1] R. W. Picard, *Affective Computing*, MIT Press, Cambridge, UK, 1997.

[2] R. W. Picard, "Affective computing: challenges," *International Journal of Human Computer Studies*, vol. 59, no. 1-2, pp. 55–64, 2003.

[3] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: analysis of affective physiological state," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1175–1191, 2001.

[4] X. Sui and Y. T. Ren, "Online processing of facial expression recognition," *Acta Psychologica Sinica*, vol. 39, no. 1, pp. 64–70, 2007 (Chinese).

[5] Y. M. Wang and X. L. Fu, "Recognizing facial expression and facial identity: parallel processing or interactive processing," *Advances in Psychological Science*, vol. 13, no. 4, pp. 497–500, 2005 (Chinese).

[6] M. White, "Effect of photographic negation on matching the expressions and identities of faces," *Perception*, vol. 30, no. 8, pp. 969–981, 2001.

[7] Y. Yang, G. Wang, P. Chen, J. Zhou, and K. He, "Feature selection in audiovisual emotion recognition based on rough set theory," in *Transaction on Rough Set VII*, pp. 283–294, 2007.

[8] Y. Yang, G. Y. Wang, P. J. Chen, et al., "An emotion recognition system based on rough set theory," in *Proceeding of the Active Media Technology*, pp. 293–297, 2006.

[9] P. Chen, G. Wang, Y. Yang, and J. Zhou, "Facial expression recognition based on rough set theory and SVM," in *Proceedings of the 1st International Conference on Rough Sets and Knowledge Technology (RSKT '06)*, pp. 772–777, Chongqing, China, July 2006.

[10] J. Zhou, G. Y. Wang, Y. Yang, et al., "Speech emotion recognition based on rough set and SVM," in *Proceedings of the 5th IEEE International Conference on Cognitive Informatics (ICCI '06)*, pp. 53–61, 2006.

[11] L. Shang, Q. Wan, W. Yao, J. Wang, and S. Chen, "An approach for reduction of continuous-valued attributes," *Journal of Computer Research and Development*, vol. 42, no. 7, pp. 1217–1224, 2005 (Chinese).

[12] R. Jensen and Q. Shen, "Tolerance-based and fuzzy-rough feature selection," in *Proceedings of the 16th IEEE International Conference on Fuzzy Systems*, pp. 877–882, 2007.

[13] Z. Pawlak, "Rough sets," *International Journal of Computer & Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982.

[14] Z. Pawlak, "Rough classification," *International Journal of Man-Machine Studies*, vol. 20, no. 5, pp. 469–483, 1984.

[15] Z. Pawlak, "Rough set theory and its applications to data analysis," *Cybernetics and Systems*, vol. 29, no. 7, pp. 661–688, 1998.

[16] R. W. Swiniarski and A. Skowron, "Rough set methods in feature selection and recognition," *Pattern Recognition Letters*, vol. 24, no. 6, pp. 833–849, 2003.

[17] N. Zhong and A. Skowron, "A rough set-based knowledge discovery process," *International Journal of Applied Mathematics and Computer Science*, vol. 11, no. 3, pp. 603–619, 2001.

[18] G. Y. Wang and Y. Wang, "3DM: domain-oriented data-driven data mining," *Fundamenta Informaticae*, vol. 90, no. 4, pp. 395–426, 2009.

[19] G. Y. Wang, "Introduction to 3DM: domain-oriented data-driven data mining," in *Proceedings of the 3rd International Conference on Rough Sets and Knowledge Technology (RSKT '08)*, pp. 25–26, Chengdu, China, May 2008.

[20] S. Ohsuga, "Knowledge discovery as translation," in *Foundations of Data Mining and Knowledge Discovery*, T. Y. Lin, et al., Ed., pp. 1–19, Springer, Berlin, Germany, 2005.

[21] The Cohn-Kanade AU-Coded Facial Expression Database, http://vasc.ri.cmu.edu/idb/html/face/facial_expression/index.html.

[22] The Japanese Female Facial Expression (JAFFE) Database, http://www.kasrl.org/jaffe.html.

[23] Chongqing University of Posts and Telecommunications Emotional Database (CQUPTE), http://cs.cqupt.edu.cn/users/904/docs/9317-1.rar.

[24] G. Y. Wang, H. Yu, and D. C. Yang, "Decision table reduction based on conditional information entropy," *Chinese Journal of Computers*, vol. 25, no. 7, pp. 759–766, 2002 (Chinese).

[25] H. S. Nugyen and A. Skowron, "Quantization of real value attributes: rough set and Boolean reasoning approach," in *Proceedings of the 2nd Joint Annual Conference on Information Science*, pp. 34–37, Wrightsville Beach, NC, USA.

[26] G.-Y. Wang, Z. Zheng, and Y. Zhang, "RIDAS—a rough set based intelligent data analysis system," in *Proceedings of the 1st International Conference on Machine Learning and Cybernetics (ICMLC '02)*, vol. 2, pp. 646–649, Beijing, China, November 2002.

[27] D. Q. Miao and G. R. Hu, "A heuristic algorithm for reduction of knowledge," *Journal of Computer Research and Development*, vol. 36, no. 6, pp. 681–684, 1999 (Chinese).
[28] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley Longman, Boston, Mass, USA, 1989.
[29] G. Holmes, A. Donkin, and I. H. Witten, "WEKA: a machine learning workbench," in *Proceedings of the 2nd Australian and New Zealand Conference on Intelligent Information Systems*, pp. 357–361, Brisbane, Australia, August-November 1994.