Application of J48 Decision Tree Classifier in Emotion Recognition Based on Chaos Characteristics

Nie Chun yan[1, a], Wang Ju[1,b], He Fang[2,c], Sato Reika[3,d]

[1]Electronic Information Institute Changchun University, China

[2]School of Electronics and Information Engineering, Changchun University of Science and Technology, Changchun 130022，China

[3]School of Design, Jiangnan University, 1800 Lihu Dadao BinHu Qu, Wuxi, China

[a]e-mail: ncy66@163.com, [b]e-mail: youyou8221@yahoo.com.cn, [c]e-mail: hefang0724@sina.cn, [d]e-mail:satoreika@dg.osakac.ac.jp

Keywords: multiple physiological signals, emotion recognition, chaos, decision tree

Abstract. Physiological signals are external manifestations of emotions. Mood swings can be expressed by the change of physiological signals. Because these performances are not controlled by the individual subjective consciousness, the conclusions are more objective and true. The method based on the statistical characteristics is very difficult to describe the complex changes in physiological signals, so the J48 decision tree was adopted to train and recognize the chaotic characteristics of physiological signals in the paper, which have many advantages in solving the multi-class classification problem, such as high accuracy, fast classification speed and simple classification rules. The chaos characteristic matrixes consist of the extracted chaotic characteristic parameters, which is combined J48 decision tree classifier to recognize four different emotions. The results showed emotion recognition of physiological signals based on chaotic theory is feasible.

Introduction

There are many types of emotion recognition, including emotion recognition of physiological signals, body, face and voice [1]. The results were subjective influenced by tested individuals susceptible, which can not observe real emotions. Physiological signal of tested individuals don't subjective control by individuals, but impact by the emotional that small mood fluctuations can cause changes in physiological signals. We can study the physiological signals in certain characteristic parameters to recognition under the various different emotions. Multiple physiological signals of sentiment analysis are by analyzing characteristic parameters or a combination of characteristic parameters, looking for whether there is a corresponding mapping relationship from them and different emotional states. Physiological signal contains abundant information. Through analysis of the information can further understand the internal mechanism of various life phenomenon. In clinical diagnosis disease treatment and people with disabilities of rehabilitation training can be applied。

Professor MIT Media Lab Picard[2]. HCM Germany Augsburg university laboratory[3], collected the same subjects in the music-induced four physiological signals under different emotional states（EMG signal、Skin electrical signals and Respiration signal）. In 2004, South Korea's K. H. Kim, S. W. Bang and S.R. Kim [4] developed a emotion recognition system based on physiological signals. The experimental data used in this paper derived from HCM Augsburg university laboratory in Germany

Selection of The Classifier

Classifier is used to change the chaos mapping to a given category, based on the emotion analysis problem of multiple physiological signals can be transformed into classification problems in pattern recognition.

Many classification problems in the actual data generally do not follow normal distribution, and is not linearly separable. Simple linear classifier will often bring larger classification error, then it need to use a non-linear classifier. J48 decision tree classifier is more suitable for the condition, which is training set larger amount of data. It has a good ability to deal with the default data and data with noise, and has higher classification accuracy. In addition, suitable for the condition that the judgment of factors is relatively less, in the same time, the relationship of logic combination is not complicated. Because the sampling data of the chaotic characteristic parameters used in this paper is more, and the individual data is default, judgment factors of character space about chaos characteristic vector is less, the relationship of logic combination is not complicated, therefore, J48 decision tree classifier is taken for different emotions classification in this paper.

Principle of Decision Tree Algorithm

Decision tree [5] divide the feature space into several regions, in each region, if a category of samples is dominant, such samples can be marked with the category labels.

C4.5 algorithm[6] can overcome the shortcoming when selected properties tend to the attribute selected value more， it can also be discrete processing for continuous attribute, it also can deal with data contained missing value. C4.5 algorithm is the core algorithm of the decision tree J48 used in this paper. It can be said to be the implementation of the Weka C4.5 decision tree algorithm [7]。

The Experimental Results and Analysis

A. Construct the format of decision tree data

Before using J48 decision tree to classification of different emotions, first of all, introducing the data format used J48 decision tree. The data format is composed of a total of 100 samples of the characteristic matrix. Each sample has four chaotic characteristic parameters, the data formats of structure of J48 decision tree is shown in figure 2.

```
0.0702,1.4913,0.4321,0.2272,joy
0.0484,1.3273,0.1568,0.1155,joy
0.0591,0.9822,0.1723,0.1733,joy
0.0566,1.4643,0.3642,0.1656,joy
0.0718,1.7681,0.5371,0.2387,anger
0.042,1.5071,0.1748,0.1078,anger
0.0161,1.0208,0.1625,0.1579,anger
0.0608,2.2107,0.1995,0.1155,anger
0.0418,1.6666,0.4053,0.154,sadness
0.0129,1.5898,0.1353,0.1001,sadness
0.0189,1.2049,0.1357,0.1001,sadness
0.0666,1.2028,0.3563,0.2079,pleasure
0.0169,1.4515,0.1912,0.1463,pleasure
0.0636,0.883,0.1223,0.1463,pleasure
0.0493,1.2179,0.2889,0.1579,pleasure
0.0177,1.8657,0.1771,0.1309,pleasure
```

Fig.2 The data format of ECG signal

```
 a  b  c  d   <-- classified as
25  0  0  0 |  a = joy
 0 24  0  1 |  b = anger
 1  0 22  2 |  c = sadness
 1  0  1 23 |  d = pleasure
```

Fig.3 Confusion matrix

A traverse is called an instance in this figure, the equivalent of the sample in statistical concepts.

Part of instances are given in this paper, there are 100 instances in the whole instance space, and there are 25 instances of each emotion（joy、anger、sadness、pleasure）. The vertical line is called properties, equivalent to variables in the statistics. There are five attributes in the figure, the first four are numeric attributes, respectively is maximum Lyapunov exponent , the correlation dimension, approximate entropy and the complexity of the four chaotic characteristic parameters. The last one is the label attribute, representative of each instance belongs to category，every 25 samples have a classification labels, respectively is joy, anger, sadness, pleasure. If a missing value is numeric attribute instead of "?". Sort out characteristic matrix, makes CSV format table in the first. Then, use the Weka software to convert it into an electronic table. Finally get the data table about ECG, EMG, SC and RSP signal.

For multiple physiological signals to identify different emotions, the confusion matrix is as shown in Fig.3. For anger emotion, correct samples of classification is 24, fault sample divided into pleasure emotion is one. For sadness emotion, correct samples of classification is 22, fault sample divided into joy emotion is one. For pleasure emotion, correct sample of classification is 23, fault sample divided into joy emotion is one, and fault sample divided into sadness emotion is one. For joy emotion, correct sample of classification is 25，there is nothing wrong samples.

Figure 4 is the decision tree of emotion recognition for multi- physiological signals under different emotions (joy, anger, sadness, pleasure).

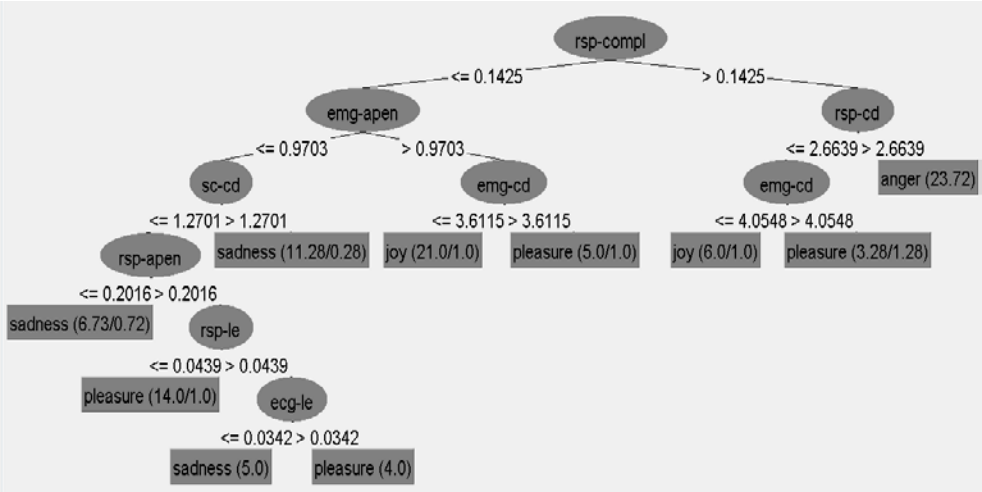

Fig.4 The decision tree of multi- physiological signals

As shown in figure 4, the root node of the decision tree is this chaos characteristics which are complexity of RSP signal. It shows that information gain rat of the chaotic characteristics is maximum. Based on different values of the property splitting into two, after comparing the information gain rate of the rest size for the chaotic characteristics, selected the child nodes, in turn, until it is completed for the classification of all samples, so leaf nodes obtained so far are pure.

B. The Results Contrast with Augsburg University

In the literature [3], Johannes Eagner et used the same data in their article. Applying the ANOVA, SFS, SBS and PCA for feature selection, then to identify the emotion through the LDF, KNN, MLP algorithm. The test results to different emotions such as table 2.

Table 2 The recognition rate contrasting with signal emotion recognition

Method	Joy	Anger	Sadness	Pleasure
LDF	77.27%	100%	72.73%	68.18%
15NN	72.73%	100%	77.27%	68.18%
MLP8	86.36%	100%	59.09%	68.18%
J48 decision tree	98%	94%	87%	86%

It can be seen in table 2, that multi-physiologic signal emotion recognition based on chaos theory has achieved good results comparing single emotion recognition rate.

Conclusion

The experimental results show that the decision tree shows a good classification effect for multi-kind of classification,.

1. Multi-physiological signals to identify different emotions of joy, anger, and recognition rate were respectively 98% and 94%. For the sadness and pleasure recognition rate is 87% and 86%. Comparing recognition rate with Augsburg university laboratory, we find the recognition of sadness and joy has achieved good results，The recognition rate of sadness Augsburg university laboratory is 77.27%, but is 87% in this paper. The recognition rate of pleasure Augsburg university laboratory is 68.18%, but is 86% in this paper.

2. Through the simulation experiment, we found that accuracy of multi-physiological signals on different emotion recognition is much better than a single physiological signal. Recognition rate of high arousal emotions (joy, anger) is higher than that of low arousal emotions (sadness, pleasure). Especially the recognition rate of joy emotion can achieve 98%.

Acknowledgment

This work is partially supported by The Ministry of education "Chunhui plan" Grant No. Z2014136 and Jilin Province Science Foundation Grant No. 201215110.

Reference

[1] Ge Chen. The pulse signal in the emotional state recognition research: [A master's degree thesis].Chong Qing：Southwest university computer application technology，2010.

[2] Picard R.W., Healey J. Affective Wearable, In Proceedings of the First International Symposium on Wearable Computers. Cambridge, 1997(l) : 231-240.

[3] Johannes Wagner, Jonghwa Kim, Elisabeth Andre. From Physiological Signals to Emotions: Implementing and Comparing Selected Methods for Feature Extraction and Classification. In IEEE International Conference on Multimedia & Expo (ICME 2005), 2005, 940-943.

[4] Kim K H， Bang S W，Kim S R. Emotion recognition system using short-term monitoring of physiological signals. Med Biol Eng Compute，2004，42:419-427.

[5] Bian Zhao Qi，Zhang Xue Gong. Pattern recognition. Bei Jing：Tsinghua Press，2000.

[6] Quinlan J R．Induction of decision tree．Machine Learning，1986, 1: 81-106.

[7] Quinlan J R．C4.5：Programs for Machine Learning．Morgan Kauffman，1993.