TELKOMNIKA, Vol.11, No.3, March 2013, pp. 1363 ~ 1370
ISSN: 2302-4046

ECG Signal Feature Selection for Emotion Recognition

Lichen Xun*[1], Gang Zheng[2]
[1]Key Laboratory of Tianjin Intelligent Computing and Software Technologies, Tianjin University of Technology, Tianjin, China
[2]Department of Computer and Communication Engineering, Tianjin University of Technology, Tianjin, 300384, China,
*Corresponding author, e-mail: xun0221@gmail.com, kenneth_zheng@vip.163.com

Abstract
This paper aims to study the selection of features based on ECG in emotion recognition. In the process of features selection, we start from existing feature selection algorithm, and pay special attention to some of the intuitive value on ECG waveform as well. Through the use of ANOVA and heuristic search, we picked out the different features to distinguish joy and pleasure these two emotions, then we combine this with pathological analysis of ECG signals by the view of the medical experts to discuss the logic corresponding relation between ECG waveform and emotion distinguish. Through experiment, using the method in this paper we only picked out five features and reached 92% of accuracy rate in the recognition of joy and pleasure.

Keywords: *Emotion recognition, ECG, Feature selection*

1. Introduction

In recent years the emotional calculation has become a focus and the main direction in human-computer interaction field, emotional recognition is especially important among them. People use the facial expressions, language, action and biological signal to distinguish emotions in various fields [1]. Due to the better objectivity and robustness of biological signal, its reaction is also more realistic. Therefore, through the use of biological signals such as ECG (Electrocardiogram) people can get a more convenient and accurate result in emotion recognition. MIT multimedia laboratories of American and Germany's Augsburg University of human-computer interaction group did a lot of research on extracting features from physiological signals to recognize emotion, and verified feasibility of some methods [2-5]. In China, there are many research institutions doing this respect of experiments, such as Tsinghua University, Jiangsu University and Southwest University. Several methods of feature selection based on Ant Colony System, Tabu search algorithm, Simulated Annealing, Particle Swarm Optimization, and using K-nearest neighbor for emotion classification, was introduced to obtain higher recognition rate and effective feature subset. But all methods are completely started from the algorithm, while ignoring some of the visual characteristics of the ECG waveform which may play an important role in emotion recognition so that increase the difficulty of the experiment or reduce the accuracy. In this paper, we start from comparison of different feature selection methods for the two states: joy and pleasure. By analyzing the ECG waveform and selected features, we explore the relationship between them, and finally verify that SVM as an emotion recognition classifier has a high accuracy rate.

2. Research Method

2.1. ECG Acquisition

The experimental data used in this paper consists of two parts:

The first part is the experimental data; we used the original physiological data provided by Germany's Augsburg University. The data were collected under four kinds of emotion states joy, pleasure, sadness and anger which were induced by different tone of the songs. They collected two minutes of the ECG signal in each emotion state, and the sampling frequency was

Received October 2, 2012; Revised January 14, 2013; Accepted January 24, 2013

set to 256Hz. Then 150 of ECG signal were chose to research. Each ECG signal was 30720 sampling points. We used signals of joy and pleasure as experimental data.

The second part is the test data; subjects were volunteers of Tianjin University of technology from 20 to 22. Their physical health statuses were good and they hadn't cardiac disease and mental disease. Experimental materials were some film clips which had higher social agreement. Compared with other emotional stimulus such as images, sounds, music and so on, film clips could arouse subjects' inner emotions more reliably and more factually. Our laboratory had carefully evaluated and picked out some film clips of joy, pleasure and sad emotion. Each movie clip was more or less 5 minutes [6]. 80 subjects were randomly asked to watch different clip as natural as possible and 3-lead ECG recorder was used to collect signals in each emotion state. After they watched the video, they were asked to fill a paper about what they felt. Each emotion got a there levels rank which were strong, normal and weak. If they chose strong, it was mean. So the data of these people were chosen to be test data they had full emotion of what we wanted they aroused. At last, after removing noises out of band and baseline drift, 400 ECG signals were chose to research, each ECG signal was 30720 sampling points. We used signals of joy and pleasure as test data.

2.2. Feature Extraction

After all kinds of pretreatment, 81 features extracted by Germany's Augsburg University were obtained to study, the Figure 1 below shows the main waveform of ECG and some key points of it , they are P,Q,R,S,T point and RR, PQ,ST,QS interval [7-8]. The X axis is time and the Y axis is voltage:

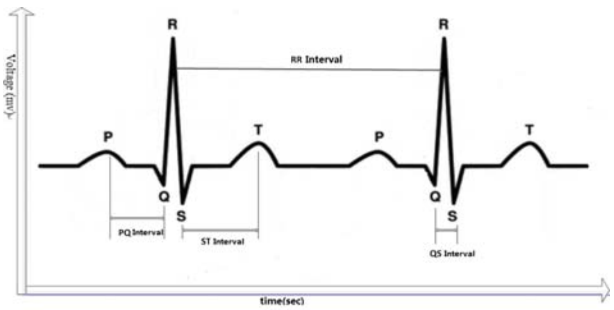

Figure 1. Key points of the ECG waveform

In these features, we got Max, Min, Mean, Median, standard deviation, Range of wave P, Q, R, S, T and interval PQ, QS, ST. At the same time, some of HRV values were calculated like SDNN (Standard Deviation of Normal to Normal), pNN50 (NN50 count divided by the total number of all NN intervals) which were got in time domain and PSD (Power spectral density) which was got in frequency domain. For testing data, we conducted the same processing with experimental data. After de-noising, baseline drift removing and normalization, two parts of data were constructed out of two original feature matrixes.

EcgR-mean is the average of two consecutive R-spacing, the median is the middle value; the STD is standard deviation; min is the minimum; max is the maximum value; range is the difference between the maximum and minimum; features of PQST point are similar.

EcgPQ-mean is the average spacing of adjacent P and Q point, the median is the middle value; the STD is standard deviation; min is the minimum; max is the maximum value; range is the difference between the maximum and minimum; features of QS and ST point are similar.

Heart rate variability (HRV) is a physiological phenomenon where the time interval between heart beats varies [9]. It is measured by the variation in the beat-to-beat interval. The term "NN" is used in place of RR to emphasize the fact that the processed beats are "normal" beats.HRV is related to emotional arousal. High-frequency (HF) activity has been found to decrease under conditions of acute time pressure and emotional strainand elevated state anxiety, presumably related to focused attention and motor inhibition.

EcgHrv is used to analyze the variation of sinus rhythm RR interval. According to the R peak position which has been determined, we obtained the number of sampling points n between two adjacent R peaks, then the sampling interval T is multiplied by n, to get the desired NN interval. The median is the middle value; the STD is standard deviation; min is the minimum; max is the maximum value; range is the difference between the maximum and minimum.

2.3 Feature Selection

Feature selection reduces the dimensionality of data by selecting only a subset of measured features (predictor variables) to create a model. Selection criteria usually involve the minimization of a specific measure of predictive error for models fit to different subsets. Algorithms search for a subset of predictors that optimally model measured responses, subject to constraints such as required or excluded features and the size of the subset [10].

2.3.1. ANOVA

Analysis of variance (ANOVA) is a collection of statistical models, and their associated procedures, in which the observed variance in a particular variable is partitioned into components attributable to different sources of variation. In its simplest form, ANOVA provides a statistical test of whether or not the means of several groups are all equal, and therefore generalizes t-test to more than two groups. Doing multiple two-sample t-tests would result in an increased chance of committing a type I error. For this reason, ANOVAs are useful in comparing two, three, or more means. ANOVA using the following formula 1 and formula 2 to evaluate:

$$SS_{between} = \sum_{i=1}^{a} n_i (\overline{Y_i} - \overline{Y})^2$$

$$SS_{within} = \sum_{i=1}^{a} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y_i})^2$$

For each feature, the output is the p-valve, if the value is close to 0, then there are significant differences in this characteristic values for the different categories. Through the above experiments, we selected 44 features from original feature matrixes: ecgR-mean, ecgR-median, ecgR-std, ecgR-min,ecgR-range, ecgP-mean, ecgP-median, ecgP-std, ecgP-min, ecgP-range, ecgQ-mean, ecgQ-median, ecgQ-std, ecgQ-min, ecgQ-range, ecgS-mean,ecgS-median, ecgS-std, ecgS-min, ecgS-range, ecgT-mean, ecgT-median, ecgT-std, ecgT-min, ecgT-range, ecgST-mean, ecgST-median, ecgST-min, ecgPampl-mean, ecgPampl-median, ecgPampl-min, ecgRampl-std, ecgRampl-min, ecgRampl-range, ecgHrv-median, ecgHrv-std, ecgHrv-min, ecgHrv-range, ecgHrvDistr-mean, ecgRampl-mean, ecgHrv-mean, ecgHrvDistr-median, ecgHrvDistr-min, ecgHrvDistr-triind.

2.3.2. Heuristic Search

A common method of feature selection is sequential feature selection. The method has two variants:

- Sequential forward selection (SFS), in which features are sequentially added to an empty candidate set until the addition of further features does not decrease the criterion.
- Sequential backward selection (SBS), in which features are sequentially removed from a full candidate set until the removal of further features increase the criterion

<u>Sequential Forward Selection (SFS)</u>

Algorithm description: Starting from the empty set, sequentially add the feature x+ that maximizes J (Yk+x+) when combined with the features Yk that have already been selected. SFS performs best when the optimal subset is small. When the search is near the empty set, a large number of states can be potentially evaluated. Towards the full set, the region examined by SFS is narrower since most features have already been selected.

Through the above experiments, we selected 37features from original feature matrixes: ecgR-mean, ecgR-median, ecgR-std, ecgR-min, ecgR-max, ecgR-range, ecgP-mean, ecgP-median, ecgP-std, ecgP-min, ecgP-max, ecgP-range, ecgQ-mean, ecgQ-median, ecgQ-max, ecgS-std, ecgS-range, ecgT-mean, ecgT-median, ecgT-std, ecgT-min, ecgT-max, ecgT-range, ecgST-max, ecgPampl-std, ecgPampl-min, ecgRampl-std, ecgSampl-mean, ecgSampl-std, ecgHrv-mean, ecgHrv-std, ecgHrv-min, ecgHrv-max, ecgHrv-range, ecgHrv-specRange1, ecgRampl-mean, ecgSampl-median.

Sequential Backward Selection (SBS)

Algorithm description: Starting from the full set, sequentially remove the feature x− that least reduces the value of the objective function J (Y−x−).

Through the above experiments, we selected 3features from original feature matrixes: ecgHrvDistr-range, ecgHrv-range, ecgHrv-max.

3. Results and Analysis

3.1. Classification Results with Single Method

In the experiment we used SVM as classifier and compared with LDA and Fisher classifier.

Support vector machines (SVM)

Support vector machines (SVM) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis [11]. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the input, making it a non-probabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

Linear discriminant analysis (LDA)

Linear discriminant analysis (LDA) is method used in statistics, pattern recognition and machine learning to find a linear combination of features which characterizes or separates two or more classes of objects or events [12]. The dependent variable of LDA is a categorical variable and LDA has continuous independent variables and a categorical dependent variable. Consider a set of observations X (also called features, attributes, variables or measurements) for each sample of an object or event with known class y. This set of samples is called the training set. The classification problem is then to find a good predictor for the class y of any sample of the same distribution (not necessarily from the training set) given only an observation X.

Fisher's linear discriminant

Fisher's linear discriminant is a classification method that projects high-dimensional data onto a line and performs classification in this one-dimensional space. The projection maximizes the distance between the means of the two classes while minimizing the variance within each class. This defines the Fisher criterion, which is maximized over all linear projections, w:

$$J(w) = \frac{|m_1 - m_2|^2}{s_1^2 + s_2^2}$$

Where m represents a mean, s2 represents a variance, and the subscripts denote the two classes. In signal theory, this criterion is also known as the signal-to-interference ratio. Maximizing this criterion yields a closed form solution that involves the inverse of a covariance-like matrix. This method has strong parallels to linear perceptrons. We learn the threshold by optimizing a cost function on the training set.

The following table and pictures show the classification accuracy rate of these three methods. AR means Acceptance Rate, RR means Rejection Rate.

Table1. Classification results with single method

Feature selection methods		SVM			LDA			Fisher		
		AR	RR	Accuracy	AR	RR	Accuracy	AR	RR	Accuracy
ANOVA	Joy	84%	28%	78%	68%	32%	68%	68%	32%	68%
	Pleasure	72%	16%		68%	32%		68%	32%	
SFS	Joy	88%	16%	86%	80%	28%	76%	88%	24%	82%
	Pleasure	84%	12%		72%	20%		76%	12%	
SBS	Joy	88%	12%	88%	84%	28%	78%	86%	16%	82%
	Pleasure	88%	12%		72%	16%		84%	14%	

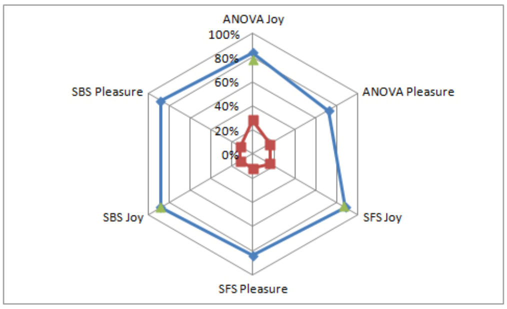

Figure 2. Classification accuracy rate with SVM

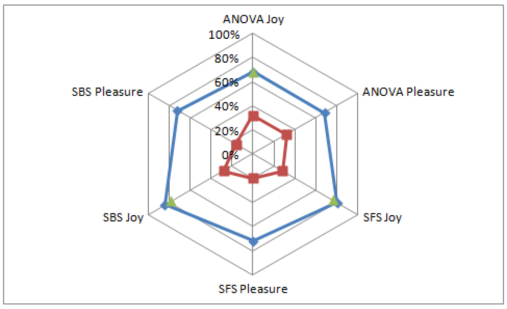

Figure 3. Classification accuracy rate with LDA

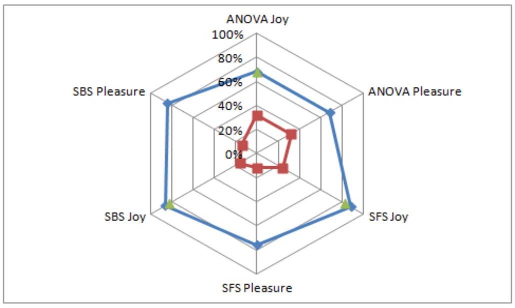

Figure 4. Classification accuracy rate with Fisher

3.2. Combination Algorithm with the SVM

Each time we used ANOVA, SFS, SBS all these three methods as filter one by one to select features which means every single selection was based on the result of the previous selection. Take ANOVA-SFS-SBS as an example, firstly we got 44 features from ANOVA method, then we used SFS to select features from these 44 features, at last we used SBS to process the result of the SFS and got ecgR-range, ecgRampl-std, ecgHrvDistr-median these three features. Similarly, ANOVA-SBS-SFS got 4 features: ecgRampl-range, ecgHrv-std, ecgHrv-range, ecgHrvDistr-min. SFS-SBS-ANOVA got 9 features: ecgP-mean, ecgP-range, ecgQ-mean, ecgS-std, ecgT-range, ecgPampl-min, ecgRampl-mean, ecgHrv-std, ecgHrv-range. SFS-ANOVA-SBS got 15 features: ecgP-mean, ecgP-range, ecgQ-mean, ecgS-std, ecgT-range, ecgST-max, ecgPampl-std, ecgPampl-min, ecgRampl-mean, ecgSampl-mean, ecgSampl-median, ecgSampl-std, ecgHrv-std, ecgHrv-max, and ecgHrv-range.

The following table and picture shows the classification accuracy rate of these four methods:

Table2. Classification results with combined method

Method	Emotion	AR	RR	Accuracy
ANOVA-SFS-SBS	Joy	88%	12%	88%
	Pleasure	88%	12%	
ANOVA-SBS-SFS	Joy	88%	24%	82%
	Pleasure	76%	12%	
SFS-SBS-ANOVA	Joy	92%	8%	92%
	Pleasure	92%	8%	
SFS-ANOVA-SBS	Joy	92%	12%	90%
	Pleasure	88%	8%	

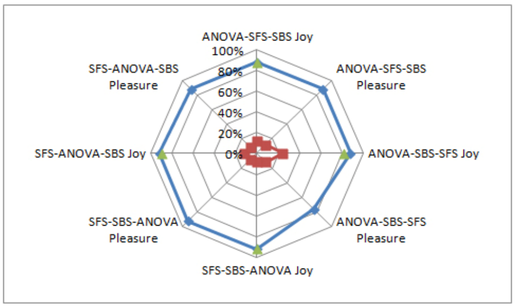

Figure 5. Classification accuracy rate with combined method

3.3. Analysis of the Feature Selection Strategy

It can be seen through the comparison of classification accuracy that the precision of this combined method was improved Compared with the single method and dimensions decreased significantly: the highest classification accuracy rate of single method was 88% which increased to 92% with combined method and features that selected with combined method were less than the single method. It also shown that SVM as a classifier is better than LDA and Fisher.

3.4. Analysis of the Feature Subset

It can be seen through the result of each selection strategy that features about R peak and HRV had obvious differences between joy and pleasure.

ecgRampl-std , ecgRampl-mean , ecgRampl-range are featureslitude and the difference between the maximum and minimum of R wave amplitude about amplitude of R peak, each representing standard deviation of R wave amplitude, average of R wave amp. ecgRampl-mean was selected twice in SFS-ANOVA-SBS anS-Sd SFBS-ANOVA.

EcgHrv-std , ecgHrv-range, ecgHrv-max are features about HRV, each representing standard deviation of HRV, the difference between the maximum and minimum of HRV and the maximum value of HRV. ecgHrv-std, ecgHrv-range were selected three times respectively in the ANOVA-SBS-the SFS, SFS-SBS-ANOVA and SFS-the ANOVA-SBS.

This point we can also intuitively see from the ECG waveform of testers. The following figure shows the waveform of the same person under two emotions. In each picture, the upper one is emotion of joy and the lower one is pleasure. It is clear that the heart rate in the two states is very different: the heart beats faster in the emotion of joy than pleasure. While the R wave amplitude also has differences: the R wave amplitude is higher when we are in the emotion of joy

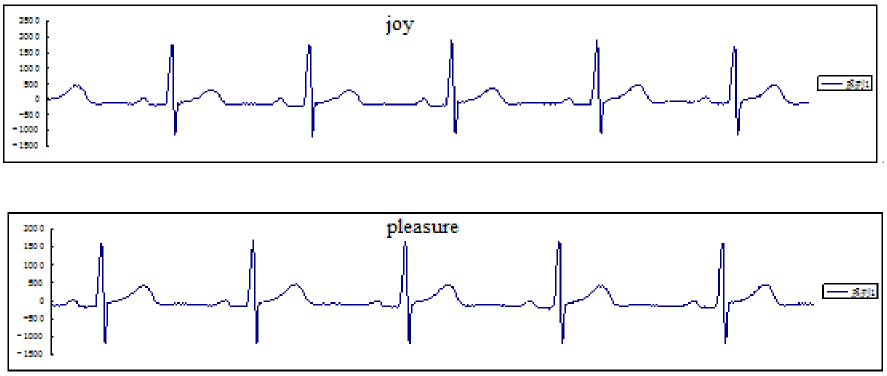

Figure 6. ECG waveforms in the state of joy and pleasure

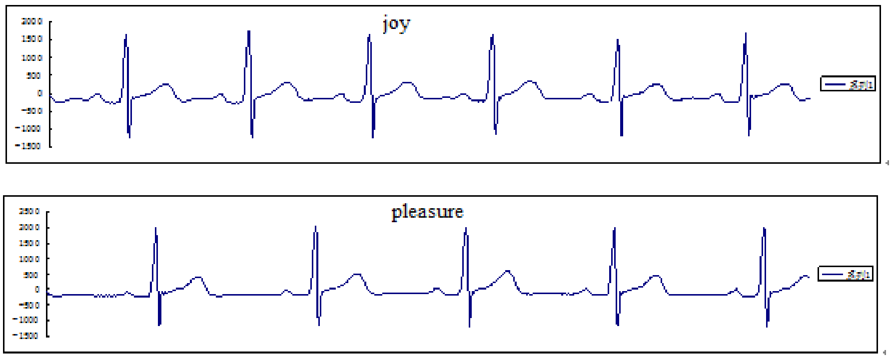

Figure 7. ECG waveforms in the state of joy and pleasure

So far the highest classification accuracy rate to distinguish joy and pleasure is 92%. The features are ecgP-mean, ecgP-range, ecgQ-mean, ecgS-std, ecgT-range, ecgPampl-min, ecgRampl-mean, ecgHrv-std, ecgHrv-range from SFS-SBS-ANOVA strategy. In the process of continued experiment, we found another features combination which could achieve this high accuracy rate. They are R-range, ecgRampl-std , ecgHrv-max , ecgHrv-range, ecgHrvDistr-range. This features subset make powerful validation to the conclusion that features about R point and HRV have important impact to distinguish joy and pleasure.

4. Conclusion

The experiments prove that the proposed strategy is feasible, and give a new way of thinking about emotion recognition experiments on the ECG waveform which is starting from the shape of the wave form to find features that have significantly different waveform between different emotions and giving them priority of selection. In this case, we can reduce complexity and improve the recognition accuracy. Next experiment is to use this method on more emotions such as anger, grief and make further improvements.

Acknowledgment

The paper is supported by Tianjin Natural Science Foundation (10JCYBJC00700) and Tianjin Key Foundation on Science Supporting Plan (10ZCKFSF00800).

References

[1] RW Picard. Affective Computing. The MIT Press, Massachusetts 1997.
[2] Vyzas E. Picard RW Affective pattern classification. Florida, MIT Media Laboratory Perceptual Computing Section Technical Report. 1998: 473.
[3] Picard R w, Vyzas E Heai, EY J. "Toward machine emotional intelligence analysis of affective physiological state". *IEEE Transactions Pattern Analysis and Machine Intelligence*. 2001; 1175—1191.
[4] Johannes Wagner, Jonghwa Kim, Elisabeth André. *"From Physiological Signals to Emotions: Implementing and Comparing Selected Methods for Feature Extraction and Classification"*. In IEEE International Conference on Multimedia & Expo (ICME2005). 2005; 940-943.
[5] Lan Li, Ji-hua Chen. *"Emotion Recognition Using Physiological Signals from Multiple Subjects".* Proceedings of the 2006 International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP'06). 355-358.
[6] Wen Wan-Hui, Qiu Yu-Hui, Liu Guang-Yuan. "Electrocardiography Recording, Feature Extraction and Classification for motion Recognition". 2009 World Congress on Computer Science and Information Engineering. 4: 168-172.
[7] Cai Jing, Liu Guangyuan, Hao Min. "The Research on Emotion Recognition from ECG Signal". 2009 International Conference on Information Technology and Computer Science. 1: 497-500.
[8] Chen Defu, Cai Jing, Liu Guangyuan. "Toward Recognizing Two Emotion States from ECG Signals". *2009 International Conference on Computational Intelligence and Natural Computing*. 2009; 1: 210-213.
[9] I Gede Pasek Suta Wijaya, Keiichi Uchimura, Gou Koutaki. Face Recognition Using Holistic Features and Simplified Linear Discriminant Analysis. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2012; 10(4): 775-787;
[10] Lu Rongxiu, Yang, Hui, Zhang, Kunpeng, Component Content Soft-Sensor of SVM Based on Ions Color Characteristics. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2012; 10(6): 1445-1452;
[11] Zhongzhe Xiao, Dellandrea E, Weibei Dou, et al."Featrues extraction and selection for emotional speech classification". *IEEE conference on Advanced Video and Signal Based Surveillance*. 2005; 411-416.
[12] HoT, and Basu M. "Complexity measures of supervised classification problems". *IEEE transactions on pattern and machine intelligence*. 2004; 24(3): 289 – 300.