Study on the Recognition of Electrogastric Signals Combining Multiple Decomposition Features

Ruijin Ma, Huisheng Zhang, Xiaofei Mao

Department of Electronic and Information, Northwestern Polytechnical University, 127 West Youyi Road, Xi’an 710072, China

Abstract

After de-noising, electrogastric signals can be analyzed and the category of the signals can be obtained. It can be observed from Fig.1 that the amount of the raw data of the generated waveform is too large. For the convenience of calculation, appropriate processing is necessary. In this paper, a method of kernel principal component analysis is used due to the characteristics of the waveform data.

Keywords: EMD LMD RVM EGG.

1. INTRODUCTION

After obtaining the electrogastric signals after pretreatment, waveform data are further analyzed and methods of feature extraction and machine learning are taken for processing the electrogastric signals (Chey et al., 2001; Brown et al., 2011). Conducting a secondary decomposition on the electrogastric signals after de-noising, sub-signals are generated independently under the method of EMD and LMD (Tello et al., 2015). At the same time, feature are described for the corresponding sub-signals in each method. For each signal, in order to fully express the information it contains, fractal dimension and sampling entropy are used for waveform feature description (Jiang et al., 2001). The method of kernel principal component analysis are taken with the purpose of enhancing the processing speed and reducing the number of dimensions (Piancastelli et al., 2011). After analyzing the classifier of kernel learning, relevance vector machine is used for data learning and classifying after dimension reduction. Finally, for original signals and signals after de-noising, experiments are conducted to verify the features that are obtained by using single decomposition algorithm and multiple decomposition algorithm respectively. The results show that the accuracy of the proposed electrogastric signal recognition algorithm, which combines multiple decomposition features and kernel principal component analysis, is 17% higher than the single decomposition algorithm without denoising, and is 9% higher than the single decomposition algorithm after signal denoising.

2. FEATURE EXTRACTION OF ELECTROGASTRIC SIGNALS

2.1 Original data selection for classification

In this paper, empirical mode decomposition and local mean decomposition (LMD) are used to decompose the electrogastric signals. As it can be seen from Fig.1, there exist five wavelets by using each method, and features in the wavelets are evaluated afterwards, then all of the feature combinations are regarded as unsorted data. Once the data source for classification is determined, two descriptors can be obtained for each wavelet by applying fractal dimension and sampling entropy that are indicated in the following section.
In order to further describe the decomposed waveform, and for convenience of EGG classification, the algorithm of fractal dimension that proposed by Katz is used for obtaining the dimension of each wavelet (Golia and Hittinahalli, 2013).

\[
FD = \frac{\log n}{\log n + \log \frac{D}{L}}
\]  

(1)

The definition of fractal dimension is as follows, where \(L\) is the total length of the signal waveform curve, i.e. the sum of the distances between two successive points; \(d\) is the distance between the first point and another point in the sequence, and the specific point is the furthest point away from the first point, i.e. the expression of \(d\) is:

\[
d = \max(\text{dist}(s1-s2))
\]  

(2)

\(n\) is the number of the steps in the waveform and \(n=L/a\), where \(a\) is the average distance between two successive points.

Calculating the corresponding fractal dimension to each of the wavelets, and we can obtain five eigenvalues for each electrogastric signal sequence. At the same time, fractal dimension for the original signals are also calculated, hence, a six dimensional eigenvector can be acquired as \(X=[x1,x2,...,x6]\). 

Figure 1. Decomposition of electrogastric signals based on EMD

Figure 2. Decomposition of electrogastric signals based on LMD

2.2 Feature description of waveform signal based on fractal dimension
2.3 Feature description of waveform signal based on sampling entropy

For a time sequence containing N points \( \{u(j) : 1 \leq j \leq N\} \), there exists \( N \times m + 1 \) vectors \( X_m(i) \) \( \{i\} \leq i \leq N \times m + 1\), where \( X_m(i) = \{u(i + k) : 0 \leq k \leq m - 1\} \) represents the vectors of \( m \) data points from \( u(i) \) to \( u(i + m - 1) \).

Distance between two vectors is defined as
\[
d[u(i), u(j)] = \max \{ |x(i + k) - u(j + k)| \} 0 \leq k \leq m - 1\}
\]
i.e. the biggest balance between the corresponding scalar components of the vectors.

\( B \) is the distance vector, \( X_m(i) \) is the vector of \( r \) and is the quantity of \( X_m(j) \), \( A \) is the distance vector, and \( X_{m+1}(i) \) is the vector of \( r \) and is the quantity of \( X_{m+1}(j) \). Hence, the sampling entropy is defined as:
\[
SampEn = -\log(A/B)
\]

Where \( r \) is obtained by multiplying 0.15 and the standard deviation of the size of the time sequence \( u(j) \), i.e. \( r = 0.15 \times STD(u(j)) \).

Calculating sampling entropy for each decomposed waveform datum, and a vector of six dimension can also be acquired as \( Y = [y_1, y_2, ..., y_6] \).

2.4 KPCA-based Dimension reduction

After using EMD and LMD for decomposition, a 24-dimansional eigenvector can be obtained by using the features combined by sampling entropy and fractal dimension. If conducting decomposition in sections for a set of collected data (the data are divided into N sections), then a vector of \( N \times 24 \) dimensions can be gained. In this case, a dimension reduction is necessary, and the KPCA algorithm is employed on the features (Yu et al., 2007).

KPCA is an improved method of Principal component analysis (PCA). PCA is to find the direction of the projections that may best represent the original data in the sense of the least mean square deviation, and the projections are orthogonal. The process can be simply described as: assuming that there are M samples of \( n \) dimensions \( x_1, x_2, ..., x_M \).

calculate the \( n \)-dimensional mean vector \( m \) and the covariance matrix \( \Sigma \) of \( n \times n \) rank:
\[
\Sigma = \frac{1}{M} \sum_{i=1}^{M} (x_i - m)(x_i - m)^T
\]  \( (3) \)

(1) calculate the eigenvalue of \( \Sigma \) and the eigenvector, and then select the eigenvector of the first \( k \) maximum eigenvalue as the direction of the principal component according to the variance contribution, and it is generally considered that the rest \( n-k \) directions are contributed by noise.
(2) A matrix \( A \) of \( n \times k \) can be comprised with \( k \) eigenvalues. The principal component representation can be obtained by projecting the original data to a \( k \)-dimensional space, and it can be expressed as

\[
y = A'(x - m)
\]

(4)

PCA is a linear method, and nonlinear structure in the data thus cannot be correctly represent, which limits the capacity of handling complex problems for PCA model. Scholkopf extended PCA to a non-linear area, the essential of which is to make PCA transformations in a high dimensional space. Such method uses kernel skills and transforms the non-linear problem into normal problem of eigenvalue, which is called kernel PCA method (KPCA). KPCA method can be widely applied in the field of face recognition, handwritten digit recognition, etc. and displays a better performance compared to the PCA algorithm. In astronomy, the structure of the acquired spectrum data collection cannot be expressed because of its classification, redshift coupling, and high dimensional features. Therefore, we tried to use KPCA algorithm for non-linear structural feature extraction. Differences in data structure extraction of the PCA and KPCA algorithms are shown in Fig. 3.

![Figure 3. Feature extraction when using PCA and KPCA](image)

Only when appropriate non-linear mapping \( \Phi \) is selected, will the proper direction in the feature space be found to represent the direction of highest variance for the data set. The key of KPCA is the selection of the kernel function and its related parameters. Different kernel functions and parameters represent different non-linear mapping, and this is the common problem that exists in the algorithms based on kernel skills. This is a N-P C problem and till now there is no perfect solutions to it. In the experiments, different kernel widths and number of the principal components will be taken and the influences of the values will be discussed.

3. SIGNAL CLASSIFICATION BASED ON RELEVANCE VECTOR MACHINES

Commonly used algorithms include ANN, KNN, SVM, etc. and the characteristics of the methods are shown in Fig.4-1. The SVM algorithm, with strong capacity of classifying for small training samples, has been widely applied in recent years’ research. However, it also has inherent weakness (Demir and Ertruk, 2007; Wu and Huang, 2009).

Relevance vector machine, as RVM in short, was brought up by Tipping, and RVM is able to overcome the shortcomings of SVM as listed below:

1. Probability predictions can be obtained since sometimes we expect to get the probability distribution of a certain classification with possible values.

2. Error parameter \( C \) is not necessary in RVM. In SVM, the value of \( C \) may influence the result to a great extent, and various possible values may need to be tried for a better performance.
(3) The speed of RVM is much faster than it of SVM when dealing with large scale data (Wu and Huang, 2009).

At the same time, in the publications of home scholars, the classification accuracy and the consuming time of SVM and RVM have been compared. Comparing to SVM, RVM has a faster speed with an approximate accuracy of SVM because of its sparsity. This characteristic is suitable for online learning and fast diagnose, where the working efficiency can be enhanced for the instruments, and classifier training can be conducted efficiently with data input of the new samples (Cheng et al, 2006). The theoretic model of M is constructed under the frame of Bayes and it is stated as follows,

\[
\{x_i\}_{i=1}^N \] is the eigenvalue of the training set, and \( t = [t_1, t_2, \ldots, t_n]^T \) is regarded as the target value. Assuming that \( t \) is the Gaussian distribution with a mean value of \( y \):

\[
p(t_i) = N(t_i \mid y(x_i; \omega), \sigma^2); y(x_i; \omega) = \sum \omega_j k(x_i, x_j) + \omega_0
\]

where \( K(x_i, x_j) \) is the kernel function, \( \omega_j \) is the weight. In order to obtain a sparse solution, \( \omega_j \) is a Gaussian distribution with a mean value of zero:

\[
p(\omega_j \mid a_j) = N(\omega_j \mid 0, \sigma^2) .
\]

The likelihood function of the training set is

\[
p(t \mid \omega, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp \left\{ \frac{1}{2\sigma^2} \| t - \Phi_\omega \|^2 \right\}
\]

where \( t = (t_1, \ldots, t_N)^T \) ; \( \omega = (\omega_1, \ldots, \omega_N)^T \); \( \Phi \) is a matrix with all its rows correspond to the response \( (\Phi)_i = [1, \varphi_1(x_i), \ldots, \varphi_p(x_i)] \) of the kernel functions to the input \( x_i \). With prior probability and the likelihood distribution theory, and according to the Bayes formula, the weighted posteriori probability distribution is

\[
p(\omega \mid t, \sigma, \alpha) = \frac{p(t \mid \omega, \sigma^2) p(\omega \mid \alpha)}{p(t \mid \alpha, \sigma^2)}
\]

But this distribution is a Gaussian distribution with multiple variables, i.e.

\[
p(\omega \mid t, \alpha, \sigma^2) = N(\mu, \Sigma)
\]

where \( \Sigma = (\sigma^{-2}\Phi^T \Phi + A)^{-1} \) is the covariance, \( A \) is the diagonal matrix of \((a_0, \ldots, a_n)\) and \( \mu = \sigma^{-2}\Sigma \Phi^T i \) is the mean value. The likelihood distribution of the target value is acquired by integrating over the weighting variables \( p(\omega \mid t, \alpha, \sigma^2) = \int p(t \mid \omega, \sigma^2) p(\omega \mid \alpha) d\omega \):

Solving the marginal likelihood of the super-parameter: \( p(t \mid \alpha, \sigma^2) = N(0, C) \) where \( C = \sigma^{-2}I + \Phi A^{-1} \Phi^T \). The Maximum A Posteriori (MAP) of the weight is determined by super-parameter \( \alpha \) and the noise variance \( \sigma^2 \). But the estimated value \( \hat{\alpha} \) and \( \hat{\sigma}^2 \) are calculated according to the maximum marginal likelihood distribution. Posterior distribution expresses the uncertainty of the optimal value of the weight, which can also be used to display the prediction uncertainty of the model. If given an input \( x^* \), the corresponding probability distribution is:

\[
p(t^* \mid x^*, \hat{\alpha}, \hat{\sigma}^2) = \int p(t^* \mid x^*, \omega, \hat{\sigma}^2) p(\omega \mid t^*, \hat{\alpha}, \hat{\sigma}^2) d\omega
\]

which obeys the Gaussian distribution:
\[ p(t^* | x^*, \hat{a}, \hat{\sigma}^2) = N(y^*, \sigma^2) \]  
\[ y^* = \mu^* \Phi(x^*) \quad \sigma^2 = \hat{\sigma}^2 + \Phi^*(x^*) \sum \Phi(x^*) \]  

**Table 1** Comparison among SVM, Bayesian and KNN classifiers

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Amount of allowed training data</th>
<th>Prediction accuracy</th>
<th>System structure</th>
<th>Calculating speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayesian classifier</td>
<td>small</td>
<td>normal</td>
<td>normal</td>
<td>fast</td>
</tr>
<tr>
<td>KNN classifier</td>
<td>middle</td>
<td>normal</td>
<td>simple</td>
<td>normal</td>
</tr>
<tr>
<td>SVM</td>
<td>large</td>
<td>high</td>
<td>complex</td>
<td>slow</td>
</tr>
</tbody>
</table>

4. EXPERIMENT ANALYSIS

Collecting electrogastric signals with EGG-E series gastrointestinal analyzer and the embedded electrogastric signal collecting instrument that co-developed by our institute, 240 groups of data can be collected (with six people into two class: abnormal (slow, fast, disorder) and normal) after the signal classification recognizing by more than three doctors. Each class contains 20 waveform data with 1500 sampling points in each long waveform. Seven of the EGG records in each class of every person are put into the sample training pool, and the left three records are taken as the testing data, from which the features are extracted with following methods respectively:

1. Decompose the original electrogastric signal through the Morlet wavelet, and solve the fractal dimension of the decomposed waveform and obtain an eigenvector of 16 dimensions.

2. Decompose the original electrogastric signal through EMD, and solve the fractal dimension of the decomposed waveform and obtain an eigenvector of 12 dimensions.

3. Decompose the original electrogastric signal by using Hilbert transform, and obtain an eigenvector of 12 dimensions by using the method in (2).

4. Combine the eigenvectors of EMD and LMD, and obtain a 24 dimensional vector.

5. Divide the data into three sections, hence the eigenvector of the wavelet is of 48 dimensions while the eigenvector of LMD+EMD is of 72 dimensions.

Training and testing the three aforementioned feature data respectively with the RVM algorithm, and the algorithms that are applied are Morlet+RVM, EMD+RVM, EMD+LMD+RVM, correspondingly. The input is the waveform data, and the final output classification result is the prediction of the electrogastric signal waveform.

4.1 Influence of the kernel function parameters selection

First, 30 groups of data in the sample training pool are selected for RVM training, and the Gaussian kernel is used in the KPCA kernel function. Then the influence of the kernel width is taken into consideration. The influence of the kernel width to classification is shown in Fig. 4 (take the feature of LMD+EMD into consideration). The figure displays the classification result by using the RVM classifier when the dimension of the feature space is set as 20, 30 and 40, respectively. It can be observed that, when the kernel
width is set as 2, the classification accuracies in all the three conditions are the highest, among which the peak value of accuracy appears when the number of the principal components are set as 20, and the accuracy is relatively the lowest when the number set to 40.

![Comparison based on KPCA on different dimensions](image.jpg)

**Figure 4.** Influence of kernel width to KPCA classification

At the same time, how classification accuracy is influenced by the extracted dimensions of the subspace in the feature space is also investigated. Referring to the Fig.5, which indicates the accuracies of the KPCA and PCA feature extraction algorithms when the dimension varies from 1 to 10. It can be seen that, when the number of dimensions increases, better results can be obtained from both KPCA and PCA algorithms. In this condition, the classification result by using KPCA is slightly better than using PCA. However, in this experiment, it is not to say that when using KPCA and PCA, the higher the fractal dimension, the better. Because dimension reduction becomes meaningless if the dimension is too high. Practically, in the experiment, the accuracy starts to decrease when the value of the dimension is over 20. In this paper, as the KPCA method is taken for dimension reduction, so the dimension is not needed to be such high. The condition where the dimension is higher than 20 is not discussed in the paper, but intuitively speaking, such condition is caused by the noise in the electrogastric signals. Even EMD method is taken for de-noising during the measurement, noise cannot be completely avoided when the dimension of KPCA is too high and approaches to its original value.

![Influence of the extracted subspace dimensions on the classification accuracy](image.jpg)

**Figure 5.** Influence of the extracted subspace dimensions on the classification accuracy

4.2 Result comparison of different feature selection

163
Based on the aforementioned analysis, the recognition accuracy of the three algorithms (Morlet+RVM, EMD+RVM, EMD+LMD+RVM, KPCA are used for dimension reduction in all three algorithms) are shown in Fig.6 when the feature space dimension is 20, the extracted subspace dimension is six, the classifier is SVM, and the training samples are selected as 20, 30, 40, and 50.

The horizontal coordinate of the figure is the number of the training samples, and the vertical coordinate is the corresponding classification accuracy. It can be observed that when using Morlet alone for feature extraction, the classification rate is the lowest. However, if EMD+LMD algorithm is combined, the rate is the highest. The algorithm of using EMD alone performs mediocre for feature extraction, which indicates that the EMD’s effect for de-noising is better than wavelet transformation. Because the feature data amount of LMD is smaller than that of Morlet wavelet and EMD, thus LMD algorithm has a higher running speed. It can also be noticed that a highest classification rate (The average rates for each method in the following figure are: Morlet+RVM: 91.2%, EMD+RVM: 93.4%, EMD+LMD+RVM: 94.2%) can be obtained with least data (six dimension) when combing the KPCA method after using EMD for de-noising and reconstructing the original electrogastric signals and integrating the LMD feature extraction algorithm. Meanwhile, the characteristic of high accuracy stands out when using a small scale sample data. This indicates that the method proposed in the paper is effective and practical.

![Classification results of three algorithms](image)

**Figure 6.** Classification results of three algorithms (horizontal coordinate represents the number of the training samples and the vertical coordinate represents the classification rate)

5. SUMMARY

Since multiple methods were taken and a good price has been paid during the signal filtering process, we propose to employ decomposing algorithm in the period of de-noising. We further bring up an automatic analysis method for electrogastric signals by using PCA-based mixed features and the RVM classifier. From the results of different parameter values in the kernel function, and by using different feature extraction algorithms, it can be concluded that the proposed electrogastric signal recognition method based on the integration of multiple decomposed features has a very good performance.

6. REFERENCES


Tello R.M., Muller S.M., Bastosfilho T.F. (2014). Comparison of new techniques based on EMD for control of a SSVEP-BCI.
