"Science Stays True Here"
Journal of Mathematics and Statistical Science, 353-363 | Science Signpost Publishing

Classification of ECG Signals with the Dimension Reduction Methods

Cigdem Bakir

Computer Engineering Department, Yildiz Technical University

Davutpasa Street, 34220, Istanbul, Turkey

Abstract

In this study, dimension reduction methods were applied to ECG signals and success of such dimension reduction techniques for the classification and segmentation of ECG signals were discussed. As classification method, such as bagging, boosting and random subspace, ensemble methods have been chosen. Because ensemble learning methods are powerful and popular classifiers. Also, segmentation of data through neighbourhood feature extraction (NFE) method were enabled by transiting from high dimensioned space to low dimension space by considering the longitudinal combination of ECG signals. Results classification results of NFE algorithm performed through longitudinal combination and as a newly developed method were compared with classification results of ECG signals obtained through dimension reduction by taking one ECG instance. Results of NFE dimension reduction technique performed by considering the neighbour ECG instances, advantage of effect on segmentation of ECG signals were presented at empirical results section and the success of suggested method was indicated. In addition, ensemble learning methods results were presented comparatively.Results obtained by performed study are promising for the studies to be conducted in further period.

Keywords: ensemble learning; ECG; classification.

Introduction

Electrocardiagram (ECG) means electrical signs, which show how the heart works in the human body and also show heart motions. Cardiac diseases arisen depending on cardiac arrhythmias, and cardiologic disorders and anomalies occurred in heart are diagnosed by ECG signals. Therefore, it is very important today that ECG signals can be analyzed properly [5].

Today most people die depending on cardiac diseases. Therefore, attention must be paid to heart health. ECG records any changes in the electrical signs generated in the heart and dispersing to all body and tries to diagnose any disorders occurred in the heart. ECG shows briefly the heartbeat and is used commonly in signal processing [6].

Most studies are conducted on classification of ECG signals. But, in the studies, it is not ensured exactly that ECG signals may be understood quickly and interpreted properly. The new method developed together with this study is applied successfully to all dimension reduction techniques. Furthermore, success to classify any ECG signals is increased and it is endeavored to guide any studies to be conducted in other biomedical fields as well.

Bortolan et.al tried to classify ECG data using neural networks on the Condo-database. The purpose of this study is to diagnose diseases present in the 3.253 ECG signal using the advanced feeder neural network. Seven classes of normal, left, right and biventricular hypertrophy, anterior, inferior and myocardial infarction were tried to be found in the Condo- database. Results were obtained with two statistical models [7].

Melgani and his colleagues have classified ECG data using Support Vector Machines (SVM) and Particle Swarm Optimization (PSO) algorithms. The purpose of the study made in the database of patients with heart failure from the Massachusetts Institute of Technology-Beth Israel Hospital (MIT-BIH) hospital is to find the best value of the discriminant function that can optimize the SVM classifier. Therefore, the classifier's success has been tried to be improved. However, it was done on 20 patients and it is very difficult to find the best common function value to increase the success of the system [13].

Zhang et al. have classified ECG signals by combining the decision trees with Wavelet transformation. Hence, they have tried to detect the heart diseases and abnormalities in the heart [14]. However, since heart diseases are quite important to humans, studies being carried out should be developed. In addition, there is a need for further study in this area.

In the work performed, band selection and feature extraction for size reduction was made first. In this study, it was tried to make class discrimination by using the class labels with training methods by using feature extraction in the ECG data. There are many methods of carrying out feature extraction. The oldest and most known among these methods is PCA, which is linear technique. Feature extraction techniques have also been used for our current data set in our work. [8].

As a second step, we tried to classify ECG data by considering 5 and 10 extensional neighborhoods for our data set which we created by these feature extraction methods. In the ECG data, each data has been expressed as a sample point. This sample point has been expressed as an ECG sample. The vectors created for these image elements are valid for all bands, and have a certain value. If the class of an image element is specified, the class of the neighbor just next to it is the same as itself, and this neighboring ECG example shows the extensional neighborhood. For a sample in this proximal neighborhood ECG signal, the neighboring nxn around it has been taken, and the ECG signal has been trained. The closeness of the extensional neighbor between two points has been measured by the Euclidean distance. Two data points which are close to one another in their extensional neighborhoods are similar and they are likely to be in the same class. As the distance increases for nxnom neighbor, the similarity decreases [3].

Model and Analysis

In the real world, a big amount of data is handled. For this reason, it is necessary to calculate too much memory and calculation to correctly classify the data. By reducing the data, the data is used effectively, and the cost is reduced. To be able to do this without data loss, size reduction methods are needed which can reduce the number of bands, which can provide size reduction. Feature extraction is a very important step in the processing of high-dimensional data. Feature extraction is carried out on linear or non-linear methods on bands. In this study, first of all, feature extraction was done to be able to make educational work using classroom information, and to increase classification success and reduce transaction cost. In the work performed, first of all, signatures were obtained that provided specific properties for the processing of the whole data. These spectral signatures have been defined as the spectral signatures that best represent the spectral distribution. Linear feature extraction techniques have been applied to these spectral signatures instead of all data. Using the size reduction techniques employed in the study, the existing data set was obtained by extracting 5 and 10 dimensional feature vectors. In addition, ensemble techniques such as bagging, bootstrap and random forest have been used to train the entire data set. The purpose here is; to find the best intermediate evaluation to be achieved by non-linear projection methods. This method has been applied to all spectral signatures and the size has been reduced. In this study, ensembles techniques such as bagging, bootstrap and random forest have been used.

Bagging

The bagging algorithm works with the small dimensions of the training data sets. The original training set is divided into N sub-clusters. Each of these sub-clusters is used as a training set. Bootstrap samples are created by randomly changing locations of the samples. Each subset also forms a classifier at the same time. Bootstrap samples are used to train different classifier components. These classifiers are combined by a compound classifier. That's why it was given the name Bagging.

Given a training dataset T of size N, standard batch bagging creates M base models. Each model is trained by calling the batch learning algorithm L_b on a bootstrap sample of size N created by drawing random samples with replacement from the original training set. The next algorithm gives the pseudocode for bagging[4].

Bagging (T, L_b, M)
for each m {1,2,K,M}
T_m = Sample_With_Replacement(T, T)
h_m $=L_b T_m$ ( )
return h_1,h_2,K,h_m { }

Bagging can also increase the predictive validity of an inconsistent estimator variable. Using variables with high variance, it makes them more usable. The bagging method gives better results compared to the individual trees. It is not a simple method. The only reason for the generation of different trees in this method is the use of different bootstrap samples.

Boosting

The basic idea of the Boosting method is to make inferences from the group of trees obtained as a result of giving different weights to the data set. It combines methods known as weak-learners. At the beginning, all observations have equal weight. As the tree community begins to grow, the weighting is organized as installed on the problem information. By combining these weightings, they form stronger learners [11]. While increasing the weight of incorrectly classified observations, the weight of rarely incorrectly classified observations is reduced. In this way, the trees gain the ability to organize themselves in the face of difficult situations. More than one algorithm has been developed in the Boosting method. AdaBoost is among the most commonly used boosting algorithms. The so-called code of this algorithm is given below [10].

D_k(i): Example i weight after learner k

α_k: Learner k weight

$\forall_i$: D_0 (i) ← 1 N

for k=1 to K do

D ← data sampled with D_{k-1}

h_k ← base learner trained on D

$\varepsilon_k \leftarrow \sum_{i=1}^{N} D_k\,(i)[h_k(x_i \neq y_i)$

$\alpha_k \leftarrow \frac{1}{2}\log\frac{1-\varepsilon_k}{\varepsilon_k}$

D_k (i) ← $\frac{D_{k-1}(i) - \alpha_k y_i\, h_k(x_i)}{Z_k}$

end for

Random Forest

Random Forest (RF) is a community learning method. Classification and regression decision trees established differently from one another constitute the decision forest community. The results obtained during the decision forest formation are combined and the latest estimate is made. In the RF method, the trees are generated with the selected bootstrap samples, and n randomly selected estimators in the each node separation. Each decision tree created is left to the widest extent and is not pruned. For classification; trees are generated in a way that each leaf node will contain only members of one class. For the regression; the trees continue to be divided until fewer units remain on the leaf node. The so-called code for the RF method is shown below [12].

A training set S := (x_1, y_1), . . . ,(x_n, y_n), features F, and number of trees in forest B.

function RandomForest(S , F)

H ← ∅

for iε1, . . . , B do

$S^{(i)}$ ← A bootstrap sample from S

h_i← RandomizedTreeLearn($S^{(i)}$, F)

H ←H∪{h_i}

end for

return H

end function

function RandomizedTreeLearn(S , F)

At each node:

f ← very small subset of F

Split on best feature in f

return The learned tree

end function

Principal Component Analysis (PCA)

PCA is the most popular orthogonal linear transformation. PCA is display of the data having the biggest variance in low dimensional space. High variance characteristics are preferred to low variance. By calculating to covariance matrix of X data matrix samples, it is possible to find the M linear mapping increasing the cost function. It removes the eigenvectors having the biggest eigenvalue. PCA used the euclidean distance between x_i and x_j data points. PCA transformation is as $\mu^T=X^T W$. W orthogonal matrix, μ^T linear transformation; whereas W displays the eigenvector corresponding to covariance matrix [1].

$$x_i = \sum_{j=1}^{p} w_{ij}\, Q_j \tag{1}$$

Kernel PCA

Kernel PCA-KPCA is the expanded feature of PCA method. However, while PCA is a linear method KPCA is a non-linear technique that improves linear techniques [15]. Data dimension is reduced by using Kernel matrix, so as K kernel matrix of the data points x_i in the form of $K(x_i, x_j)$. By changing K kernel matrix inputs, it is possible to find the centers and d eigenvectors. The essential in KPCA is to select the kernel function. Kernel function can be linear kernel, Gauss kernel and polynomial kernel. KPCA ensures quite successful results in face detection, speech recognition.

Linear discriminant analysis (LDA)

PCA states a high variety of data by a minimum number of components. But, these components never require best variety to classify them differently to ensure highest separation. To classify the data properly together with LDA, highest separation between different classes of data is ensured. It finds the dimension reduction required to provide this separation, calculating the covariance matrix. Total covariance matrix is calculated, utilizing the covariance matrix within the classes (S_W) and the covariance matrix between the classes (S_B) [2].

$$S_T = S_W + S_B \tag{2}$$

Local Linear Embedding (LLE)

It is a method similar to Isomap. However as distinct from Isomap, it only protects the local characteristics on data point graphic image. In LLE x_i data point converging to local neighbors and k w_i weights are matching a hyper smoothing being a linear union of their closest neighbors by preserving the neighbor relations of each data point. w_i weights are stable against reconfiguration, transformation, slewing, scaling. Various coefficients are created. $W=(W_{ij})_{i,j}$ minimizes the cost function. After calculating the weights, it is passed to low dimensioned space by preserving local neighbors.

Hessian LLE

As LLE discrete matrix techniques are used. It calculates the data graphic by using k the closest neighborhood. It performs preliminary analysis by measuring manifold curves of H matrix and displays in low dimensioned space. It is calculated as Tangent Hessians:

$$H_{l,m} = \sum_i \sum_j ((H_i)_{jl} * (H_i)_{jm}) \tag{3}$$

For each data point, approximate local tangent coordinates are calculated for manifold Hessians by calculating eigenvectors of covariance matrix. The H matrix is performed minimizes curviness of the

Afterwards, the matrix created in order to determined eigenvectors of covariance matrix is orthonormalized. Compared to LLE and Laplacian it is slowest and gives worst performance in discrete ones. However it is successful in convergence problems.

Analysis and Discussion

In this study, PTB (Physikalish-Techniseche Bundesantalt) ECG dataset is used. This dataset is generated by taking 549 ECG signals by the German National Metrology Institute from 289 people. 9 different diagnoses are made on the patients depending on heart diseases. Furthermore, each entry contains 15 signals measured simultaneously. Each signal is digitalized approximately by 1000 samples per second [9].

Whereas table 1 shows the results of classification accuracy made with ensemble methods for 5 dimensions; Table 2 shows the classification results for ensemble methods for 10 dimensions. Table 3 shows the results of the classification accuracy of the proposed NFE technique with ensemble methods by providing an extended combination for 5 dimensions and the results of classification accuracy for 10 dimensions have been shown in Table 4. When we look at the study made in Table 1 and Table 2, and the results of the method proposed in Table 3 and Table 4, it is observed that the proposed method greatly improves the success of the classification accuracy. In particular, the neighborhood value 10 has the highest classification success. However, the increase in the neighborhood value to a great extent reduces the success of classification. In this study, the advantages of size reduction on classification have been presented by providing an extended combination. It has been observed that the NEF technique increases the success of classification of ECG signals in both classification algorithms from about 3% to 9 % of the accuracy of classification.

Table 1

Feature Extraction Methods	**Classification Methods**		
	Bagging	**Boosting**	**Random Forest**
PCA	68,74	70,24	71,65
Kernel PCA	75,51	76,45	77,33
LDA	65,28	71,34	72,56
LLE	70,14	70,93	73,72
Hessian LLE	72,82	73,60	75,74

Table 2

Feature Extraction Methods	**Classification Methods**		
	Bagging	**Boosting**	**Random Forest**
PCA	72,27	75,96	77,80
Kernel PCA	78,09	79,73	80,94
LDA	71,64	72,01	73,96
LLE	72,11	72,84	73,21
Hessian LLE	74,55	75,27	76,00

Table 3

Feature Extraction Methods	**Classification Methods**					
	Bagging		**Boosting**		**Random Forest**	
	5	**10**	**5**	**10**	**5**	**10**
PCA	72,02	75,65	73,17	79,38	75,29	80,66
Kernel PCA	77,72	79,40	77,73	80,70	80,55	84,15
LDA	68,31	72,71	72,08	75,34	75,55	79,32
LLE	71,34	75,37	73,79	77,97	77,06	78,85
Hessian LLE	76,88	80,24	74,67	79,55	78,71	81,73

Table 4

Feature Extraction Methods	Classification Methods					
	Bagging		Boosting		Random Forest	
	5	10	5	10	5	10
PCA	77,35	83,02	77,63	81,36	81,37	85,99
Kernel PCA	80,99	86,83	83,45	86,37	87,40	90,95
LDA	76,04	81,24	80,37	85,94	77,09	82,31
LLE	77,74	80,30	78,85	81,31	79,35	81,96
Hessian LLE	79,58	81,16	82,00	82,64	80,75	84,37

Conclusions

In this study, dimension reduction methods were applied to ECG signals. Also, segmentation of data through neighbourhood feature extraction (NFE) method by transiting from high dimensioned space to low dimension space by considering the spatial combination of ECG signals with results classification of traditional nonlinear feature extraction were compared. Morover, in this study were implemented ensemble methods and were compared of results the ensemble methods In this study, an extensional combination is provided and advantages of the dimension reduction on classification are submitted. Success of NFE technique suggested in this study on classification of ECG signals is shown. In any studies to be done in the future, it shall be ensured that NFE technique shall be improved further and applied in different practice fields.

References

[1] Maglaveras N., Stamkopolous T. etc, ECG pattern recognition and classification using nonlinear transformations and neural networks: A review, Medical Informatics, vol.52, issue 1-3, pp.191-208, 1998.

[2] Kiranyaz S., Ince T. etc, Personalized long-term ECG classification: A systematic approach, Expert Systems with Applications, issue 4, pp.3220-3226, 2011, vol.38.

[3] Bakir C., ECG Signals Classification with Neighborhood Feature Extraction Method, Medical Technologies National Conference (TIPTEKNO), PP.1-4, 2015.

[4] Oza N.C, Online Bagging and Boosting, 2005 IEEE Conference on Man and Cybernetics, 2005.

[5] Kiranyaz S., Ince T. etc, Personalized long-term ECG classification: A systematic approach, Expert Systems

with Applications, vol.38, issue 4, pp.3220-3226, 2011.

[6] Erdoğmuş P., Peşçaker A., Dalgacık Dönüşümü ile EKG Sinyallerinin Özellik Çıkarımı ve Yapay Sinir Ağları ile Sınıflandırılması, 5.Uluslararası İleri Teknolojiler Sempozyumu (IATS'09), 2009.

[7] Bontolan G. and Willems J.L., Diagnostic ECG Classsification based on neural networks, Journal of Electrocardiology, pp.75-79, 1993.

[8] Ceylan R.,Ozbay Y., Comparison of FCM, PCA and WT techniques for classification ECG arrhytmias using neural network, Expert Systems with Applications, vol.33,issue 2, pp.286-295, 2007.

[9] Vozda M ., C erny M ., Methods f or d erivation of orthogonal leads from 12-lead electrocardiogram:A review, Biomedical Signal Processing and Control, vol.19, pp.23-34, 2015.

[10] Introduction to Boosting, http://www.cs.man.ac.uk/~stapenr5/boosting.pdf, 2012.

[11] Boosting , http://cs.nju.edu.cn/zhouzh/zhouzh.files/publication/SpringerEDBS09b.pdf

[12] Random Forest, http://pages.cs.wisc.edu/~matthewb/pages/notes/pdf/ensembles/RandomForests.pdf.

[13] Melgani F. and Bazi Y., Classification of Electrocardiogram Signals With Support Vector Machines and Particle Swarm Optimization, IEEE Engineering in Medicine and Biology Society IEEE Computer Society, pp.667-677, vol.12, issue5, 2008.

[14] Zhang L., Peng H. etc, An approach for ECG classification based on wavelet feature extraction and decision tree, Wireless Communications and Signal Processing (WCSP), 2010 International Conference on, 2010.

[15] L. Macter, E. Postma, and J. Heik, Dimensionality reduction:A comparative review, Ticc Tilburg University, October 26, 2009.

Published: Volume 2017, Issue 12 / December 25, 2017