ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 17, No. 1, 2020, *pp*.075-080

Classification of ECG signals based on functional data analysis

Zhangxiao Miao, Chunzheng Cao[1]
School of Mathematics and Statistics, Nanjing University of Information Science & Technology,
Nanjing 210044, China

(Received February 15, 2022, accepted April 29, 2022)

Abstract: Electrocardiogram (ECG) signals are the impulses generated by the heart which are used to analyze the proper functioning of heart. This paper applies a functional data analysis method to classify ECG signals. The classification of functional data can be divided into regression model-based classification methods and density-based classification methods. In this paper, Generalized Functional Linear Models (GFLM) and Functional Linear Discriminant Analysis (FLDA) are introduced. Finally, we apply GFLM, FLDA and SVM, neural network, KNN in the actual data, and find that the functional data classification method performs better in the classification of ECG signals.

Keywords: functional classification, generalized functional linear model, functional linear discriminant analysis.

1. Introduction

Analysis of electrocardiogram (ECG) enables biometrics, activity recognition, and more importantly, patient screening and diagnosis activities [1]. At present, the most common diagnostic method of arrhythmia is to rely on the experience of doctors. Faced with a large amount of ECG signal data, doctors may miss diagnosis and misdiagnose due to fatigue caused by continuous work for a long time. ECG automatic classification improve the efficiency and accuracy of ECG diagnosis and avoid human caused error. It can help doctors diagnose and treat arrhythmias in a timely manner, thereby reducing the incidence and mortality of cardiovascular diseases. Polat et al. [2] classified ECG signals by using the support vector machine to effectively distinguish between normal and abnormal ECG signals. Debnath et al. [3] proposed a method based on artificial neural network to complete the Automatic classification and recognition of ECG signals. Meanwhile, functional data contains more perfect and sufficient information, which can well avoid the information loss. This kind of data is widely used in economics, finance, biological information, meteorology, medicine, industry and other fields. The infinite dimension of functional data conforms to the requirements of data information richness and structural complexity in the era of big data, making it a hot topic in statistical research in recent years, whether in practical application or theoretical exploration. It would be interesting to classify ECG signals using functional data analysis.

In the last years, researchers concentrated their efforts to solve functional data classification problems. Stone [4] and Devroye [5] had found that the K-Nearest Neighbor was significantly different from the finite-dimensional case when applied to functional data. James et al. [6] who extended the classical linear discriminant analysis to functional data, proposed the functional linear discriminant analysis (FLDA) with obvious effect when only part of the curve was observed, and also provided the quadratic discriminant analysis and regularized function discriminant analysis. Li et al. [7] proposed functional segment discriminant analysis (FSDA) method by combining classical linear discriminant analysis and support vector machine, which is especially suitable for sparse functional data. Delaigle et al. [8] proved that perfect asymptotic classification could be achieved using linear method in the classification of function data by taking advantage of the inherent high-dimensional nature of function data. Rossi et al. [9] studied the support vector of functional data and the kernel support vector machine to explore the consistency of classification. Rabaoui et al. [10] proposed a non-parametric method combining the generation model and the functional data analysis method to identify and

[1] Corresponding author. *E-mail address*: caochunzheng@163.com.

analyze phoneme signals based on the improved Bayesian classifier. Then, it was found that the non-parametric method has certain advantages after compared with the functional support vector machine method.

In finite samples of functional data, linear truncation by projection onto partial least squares or principal component basis can achieve good classification performance. Leng et al. [11] applied functional principal component and Logistic regression model to classify yeast cell cycle data, and found that Logistic regression model based on principal component basis had certain superiority by comparing it with discriminant analysis after B-spline expansion of prediction. Abraham et al. [12] studied the moving window classification method of functional data dichotomies and proved that moving window classification is convergent under certain conditions. Wang [13] applied wavelet basis to approximate the prediction function and proposed the Bayesian Logistic regression model. In order to overcome the defect that reaction variable information cannot be fully applied in the expansion of known bases and principal component bases, Preda et al. [14] proposed the function discriminant analysis method based on partial least squares bases. Gomez-Verdejo et al. [15] put forward an interactive information estimation method for classification tasks, considering that the classification of functional data needs to select a reduced subset of features in the initial set. However, it is difficult to estimate through a limited sample set. Biau et al. [16] established the weak consistency of the KNN method for the random curve valued in the Hilbert space. After the function prediction was expanded on the Fourier basis, KNN was applied to its coefficients and was applied to speech recognition.

Considering that the results of the functional data classification method are easily affected by the selection distance, Chang et al. [17] used a distance classification method based on wavelet threshold for image data, and the actual data analysis showed that this method had a good classification effect. Berrendero [18] proposed a method of variable selection based on distance correlation for function classification. To overcome the difficulty that functional data cannot be classified by Bayes, Dai et al. [19] changed the classification of functional data into a problem based on principal component score through principal component base dimensionality reduction, and adopted Bayes classification for principal component score. Functions can be classified by the distance from the class center. Darabi et al. [20] proposed a new classification method based on the projection distance of weighted functions, which can achieve optimal classification results by selecting projection functions with optimal classification results.

This article is organized as follows: Section 2 mainly introduces the models and methods and we briefly introduce two typical functional data classification methods. In section 3, we use them to classify two real examples. Finally, in Section 4, we analyze the research results and give a summary.

2. ECG Data

Electrocardiogram (ECG) signals are physiological signals of the human body, including a large number of physiological and pathological information of the human body. Physiological signals of different individuals have obvious differences. Therefore, ECG has been used to monitor and diagnose clinical heart diseases. Nowadays it has become one of the most important non-implantable tools for ECG monitoring. The analysis of the ECG signal is used for detecting cardiac diseases. The ECG signal mainly consists of PQRST waves. In a normal heart, each beat begins in the right atrium. The atrial depolarization is represented by the P-wave. Ventricular depolarization and atrial repolarization are represented by the QRS complex, while the ventricular diastole is represented by the T-wave. If any fatty material is present on the inner walls of the heart, the coronary arteries become narrow. It results in restricted blood supply to heart. The heart does not get sufficient oxygen. Consequently, it leads to Ischemia. If this continues for a long time cells may die resulting in damages to the heart muscles causing myocardial infarction.

Fig.3 shows the normal and ECG signal in different segments of PQRST wave form. The ECG data used in this work was formatted by R. Olszewski at Carnegie Mellon University, 2001 which is available from http://www.timeseriesclassification.com/dataset.php#. The dataset contains 200 samples, including 100 training set samples and 100 test set samples. Each series traces the electrical activity recorded during one heartbeat. The two classes are a normal heartbeat and a Myocardial Infarction. Fig.3 shows an example of electrocardiogram diagnosis: the leftmost represents data recorded during a normal heartbeat, and the rightmost represents data recorded during a heartbeat exhibiting behavior indicative of a cardiac condition called myocardial infarction.