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.

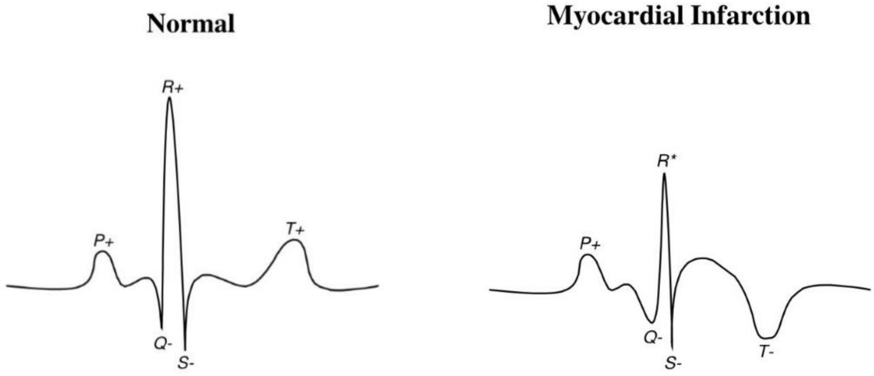

Fig.3 The normal and abnormal ECG during one cardiac cycle.

3. Models and Methods

3.1 Functional regression for classification

For regression-based functional classification models, functional generalized linear models or more specifically, functional binary regression, such as functional logistic regression, are popular approaches. For a random sample $\{(Y_i, x_i); i = 1, \dots, n\}$, where Y_i represents a class label, $Y_i \in \{1, \dots, L\}$ for L classes, for example Y_1 stands for summer months and Y_2 stands for winter months in Italy electricity power demand data, associated with functional observations x_i, a classification model for an observation X_0 based on functional logistic regression is

$$log\frac{Pr(Y = k \mid x_0)}{Pr(Y_i = L \mid x_0)} = \gamma_{0k} + \int_{\mathcal{T}} x_0\,(t)\xi_{1k}(t)dt, \quad k = 1, \dots, L-1 \tag{2.1}$$

Where γ_{0k} is an intercept term and $\xi_{1k}(t)$ the coefficient function of the predictor $x_0(t)$ and $Pr(Y_i = L \mid x_i) = 1 - \sum_{k=1}^{L\sum(Y_i=k|x_i)} Pr$. This is a functional extension of the baseline odds model in multinomial regression. Given a new observation x_0, the model-based Bayes classification rule is to choose the class label Y_0 with the maximal posterior probability among $Pr(Y_0 = k \mid x_0), k = 1, \dots, L$. More generally, used the generalized functional linear regression model based on the FPCA approach. When the logit link is used in the model, it becomes the functional logistic regression model, several variants of which have been studied.

Generalized functional linear models (GFLM) are extensions of classical linear models with the following three components: a random component where for the responses, $Y \sim$ exponential family, with means $E(Y) = \mu$; linear predictors, $\eta = \sum x_p\,\beta_p$, where x_p is the p-th predictor variable; and a monotone link function, $g(\mu) = \eta$. When Y is binomial, this is a binomial regression model. A special case is logistic regression where the link function $g(\cdot)$ is the logit function, i.e. $logit(x) = log\{x/(1-x)\}$, so that $g^{-1}(x) = e^x/(1+e^x)$.

In the framework of the classification problem, the response Y denotes membership in one of two groups, coded as a binary random variable, where $Y = 1$ if the observation comes from G_1 and $Y = 0$ if it comes from G_0. The predictor function $X(t), t \in [0, T]$ from now on is assumed to be a centered random curve, i.e. $\mu(t) \equiv 0$. For an i.i.d. sample $X_i(t)$, for $i = 1, \dots, n$ the linear predictors are defined by $\eta_i = \alpha + \int \beta\,(t)X_i(t)dt$, leading to the functional generalized linear model:

$$Y_i = g^{-1}(\eta_i) + e_i, \quad i = 1, \dots, n, \tag{2.2}$$

where $g(\cdot)$is a link function as before, α is a constant and $\beta(\cdot)$is the parameter function. The errors e_i are assumed to be independent,$E(e_i) = 0, var(e_i) < C < \infty$. The M-truncated model becomes

$$Y_i = g^{-1}\left(\alpha + \sum_{m=1}^{M} \beta_m\,\xi_{im}\right) + e_i, \quad i = 1,2, \dots, n, \tag{2.3}$$

For fixed M, $\beta^T = (\alpha, \beta_1, \beta_2, \dots, \beta_M)$, the unknown parameter vector, can be estimated by solving the estimating or score equation

$$U(\beta) = \sum_{i=1}^{n} (Y_i - \mu_i)\,g^{'}(\eta_i)\varepsilon_i/\sigma^2(\mu_i). \tag{2.4}$$

Denote the solution by $\hat{\beta}^T = (\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_M)$.

For functional binomial regression, as in classical binomial regression for discriminant analysis, set $\pi_i = P(Y_i = 1)$ and prior probabilities p_I and p_0 for the groups G_1 and G_0, respectively. We estimate π_i by $\hat{\pi}_i = \hat{P}(Y_i = 1 \mid X_i(t)) = g^{-1}(\hat{\alpha} + \sum_{m=1}^{M} \hat{\beta}_m \hat{\varepsilon}_{im})$. Then we classify the i-th observation into G_1 if $\hat{\pi}_i \geq p_1$, otherwise into G_0.

3.2 Functional discriminant analysis for classification

In contrast to the regression-based functional classification approach, another popular approach is based on the classical linear discriminant analysis method. The basic idea is to classify according to the largest conditional probability of the class label variable given a new data object by applying the Bayes rule. Suppose that the k-th class has prior probability $\pi_k, \sum_{k=1}^{K} \pi_k = 1$. Given the density of the k th class, f_k , the posterior probability of a new data object X_0 is given by the Bayes formula,

$$Pr(Y = k \mid x_0) = \frac{\pi_k f_k(x_0)}{\sum_{j=1}^{K} \pi_j f_j(x_0)}. \tag{2.5}$$

Developments along these lines include a functional linear discriminant analysis approach to classify curves, a functional data-analytic approach to signal discrimination, using the FPCA method for dimension reduction and kernel functional classification rules for nonparametric curve discrimination. Theoretical support and a notion of "perfect classification" standing for asymptotically vanishing misclassification probabilities has been introduced for quadratic functional classification.

Different from classification methods based on regression, functional linear discriminant analysis(FLDA) based on classical linear discriminant analysis is also a common functional classification method. The basic idea is to classify new samples based on maximizing conditional probability by using Bayesian criterion. Also suppose that the prior probability is π_k, and $\pi_0 + \pi_1 = 1$. Given the density function f_k of the k-th sample, the conditional probability of the sample to be classified x^* can be obtained by Bayesian formula as follows：

$$(y^* = k|x^*) = \frac{\pi_k f_k(x^*)}{\pi_1 f_1(x^*) + \pi_0 f_0(x^*)}. \tag{2.6}$$

Classify x^* based on the maximum posterior probability. If it is further assumed that the k-th sample obeys a Gaussian distribution with mean μ_k and covariance matrix Σ, then maximizing the conditional probability is equivalent to

$$\arg\max_k (L_k)\,, \tag{2.7}$$

where $\boldsymbol{L_k}$ is discriminant function

$$L_k = x^T \Sigma^{-1} \mu_k - \frac{\mu_k^T \Sigma^{-1} \mu_k}{2} + \log \pi_k. \tag{2.8}$$

3.3 Machine learning methods for classification

3.3.1 Support vector machine (SVM)

Support vector classifier is a supervised learning technique best used for linear system classification. This algorithm is used for the analysis of database classification and regression techniques. This algorithm constructs a hyper plane or a set of hyper planes in the multi-dimensional space thereby linearly distinguishing the classes. We can treat the classification to be the best when the distance from the hyper plane to the nearest training set data point is large.

3.3.2 Neural network

Artificial neural network is the interconnection of neurons which produce output upon the variation of characteristics of input and the activation. Each input is mapped to its unique weight, so to modify the output we alter the weights and the activation function. This process of modification of output based on input and activation is called as the learning rate.

3.3.3 K-Nearest Neighbors (KNN)

The k-Nearest Neighbors (KNN) family of classification algorithms and regression algorithms is often referred to as memory-based learning or instance-based learning. Sometimes, it is also called lazy learning.

These terms correspond to the main concept of KNN. The concept is to replace model creation by memorizing the training data set and then use this data to make predictions.

4. Classification for ECG Data

We apply functional methods and machine learning methods to ECG database to determine the type of ECG signal. Figure 6 shows the training of a 96-second ECG signal in which blue represents normal ECG signals and red represents Myocardial Infarction ECG curves. This paper selects 100 ECG signals as training data and 100 ECG signals as test data. In this paper, the K of KNN equals 4 and the kernel of SVM is Gaussian kernel.

The classification result is evaluated using 10-fold cross-validation. Different from machine learning methods, Table 3 reports the classification accuracies for the normal and abnormal data sets at different time using functional data methods. The classification accuracy of FLDA is higher than that of GFLM in the first 50 seconds, and then the effect of GFLM is significantly increased between 50 and 75 seconds. When the training reached 96 seconds, both GFLM and FLDA achieved high classification accuracy. Table 4 shows the classification accuracy of functional methods and machine learning methods. The difference between the four methods is not large, but the functional data method can obtain a better classification accuracy with less known data information.

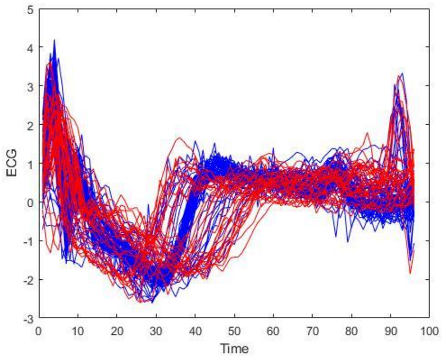

Fig. 6 Training data of ECG signal.

TABLE 3. Classification accuracy for ECG data at different time.

Time	GFLM	FLDA
25	0.81	0.83
50	0.84	0.87
75	0.93	0.92
96	0.97	0.95

TABLE 4. Classification accuracy for ECG data.

Methods	GFLM	FLDA	SVM	NN	KNN
Accuracy	0.97	0.95	0.96	0.97	0.97

5. Conclusion and discussion

Classification of functional data is an important topic in statistics. This paper briefly reviews existing approaches to functional data classification. It has been proposed to divide these methods into three groups. In addition to algorithm-based classification methods, this paper mainly introduces model-based classification methods and probability density classification methods. The first group is regression-based functional classification models. For example, we introduce a generalized functional linear regression model with a logit link function for binary data. The second group is the probability density classification method. For example, we introduce functional discriminant analysis for classification, which can be used to generate classifications on new (test) curves. We also compare functional data methods with machine learning methods such as SVM, Neural Network, KNN.

We analyzed and compared the displayed ECG data at different times using GFLM and FLDA, respectively, neither method is very good when the data is less well known, but when classifying new individual predictions, both This method can obtain better classification accuracy. There is little difference between functional data methods compared to machine learning methods when it comes to predicting classes as a whole for new individuals. In the data, there are many factors that affect the classification accuracy. For example, there may be some outliers in the training data, which will become the main interference factors for classification prediction. How to reduce the influence of outliers deserves further study.

References:

1. G Kaur, G. Singh, V. Kumar, *A review on biometric recognition,* International Journal of BioScience and Bio-Technology 6(4), (2014) 69-7.6.
2. K. Polat, S. Güne, *Breast cancer diagnosis using least square support vector machine,* digital Signal Processing 17(4), (2007) 694-701.
3. T. Debnath, M. M. Hasan, T. Biswas, *Analysis of ecg signal and classification of heart abnormalities using artificial neural network,* 2016 9th International Conference on Electrical and Computer Engineering (ICECE), (2016) 353-356.
4. C. J. Stone, *Consistent nonparametric regression,* The Annals of Statistics 5(4), (1977) 595-620.
5. L. Devroye, G. Lugosi, *A universally acceptable smoothing factor for kernel density estimates,* The Annals of Statistics 24(6), (1996) 2499-2512.
6. G. M. James, T. J. Hastie, *Functional linear discriminant analysis for irregularly sampled curve,* Journal of the Royal Statistical Society Series B-Statistical Methodology 63(3), (2001) 533-550.
7. B. Li, Q. Yu, *Classification of functional data: a segmentation approach,* Computational Statistics & Data Analysis 52(10), (2008) 4790-4800.
8. A. Delaigle, P. Hall, *Achieving near perfect classification for functional data,* Journal of the Royal Statistical Society Series B-Statistical Methodology 74(2), (2012) 267-286.
9. F. Rossi, N. Villa, *Support vector machine for functional data classification,* Neurocomputing 69 (7/9), (2006) 730-742.
10. A. Rabaoui, H. Kadri, M. Davy, *Nonparametric Bayesian supervised classification of functional data,* 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), (2012) 3381-3384.
11. X. Y. Leng, H. G. Müller, *Classification using functional data analysis for temporal gene expression data,* Bioinformatics 22(1), (2006) 68-76.
12. C. Abraham, G. Biau, B. Cadre, *On the kernel rule for function classification,* Annals of the Institute of Statistical Mathematics 58(3), (2006) 619-633.
13. X. H. Wang, S. Ray, B. K. MALLICK, *Bayesian curve classification using wavelets*, Journal of the American Statistical Association 102(479), (2007) 962-973.
14. C. Preda, G. Saporta, C. Lévéder, *PLS classification of functional data,* Computational Statistics 22(2), (2007) 223-235.
15. V. Gomez-Verdejo, J. Arenas-Garcia, A. R Figueiras-Vidal, *A dynamically adjusted mixed emphasis method for building boosting ensembles,* IEEE Transactions on Neural Networks 19(1), (2008) 3-17.
16. G. Biau, F. Bunea, M. H. Wegkamp, *Functional classification in Hilbert spaces,* IEEE Transactions on Information Theory 51(6), (2005) 2163-2172.
17. C. Chang, Y. K. Chen, R. T. Ogden, *Functional data classification: a wavelet approach*, Computational Statistics 29(6), (2014) 1497-1513.
18. J. R. Berrendero, A. Cuevas, J. L. Torrecilla, *Variable selection in functional data classification: a maxima-hunting proposal,* Statistica Sinica 26(2), (2016) 619-638.
19. X. T. Dai, H. G. Müller, *Principal component analysis for functional data on Riemannian manifolds and spheres,* The Annals of Statistics 46(6B), (2018)3334-3361.
20. N. Darabi, S. M. E. Hosseini-Nasab, *Projection-based classification for functional data,* Statistics 54 (3), (2020) 544-558.