Diagnosis of Cardiac Arrhythmia Using Kernel Difference Weighted KNN Classifier

WM Zuo[1], WG Lu[1], KQ Wang[1], H Zhang[2]

[1]Harbin Institute of Technology, Harbin, China
[2]University of Manchester, Manchester, UK

Abstract

In this paper, we proposed a kernel difference-weighted k-nearest neighbor classifier (KDF-WKNN) for the diagnosis of cardiac arrhythmia based on the standard 12 lead ECG recordings. Different from classical KNN, KDF-WKNN defines the weighted KNN rule as the constrained least-squares optimization of sample reconstruction from its neighborhood, and then uses the Lagrangian multiplier method to compute the weights of different nearest neighbors by introducing the kernel Gram matrix G. In arrhythmia analysis, it is unavoidable that some attribute values of a person would be missing. Thus, this paper further introduces a modified PCA method to address this problem. To evaluate the performance of KDF-WKNN, Experimental results on the UCI cardiac arrhythmia database indicate that, KDF-WKNN is superior to the nearest neighbor classifier, and is very competitive while compared with several state-of-the-art methods in terms of classification accuracy.

1. Introduction

Cardiac arrhythmia, disorders of cardiac rhythm, may indicate the susceptibility of serious heart disease, stroke or sudden cardiac death. Early diagnosis of cardiac arrhythmia makes it possible to choose appropriate anti-arrhythmic drugs, and is thus very important for improving arrhythmia therapy.

Electrocardiogram records the electronic activities of the heart, and has been widely adapted for diagnosing cardiac arrhythmia. By far, a number of signal processing [1], pattern recognition [2, 3], and machine learning [4] methods had been proposed. The publications of several generally available arrhythmia datasets also played an important role in stimulating research on cardiac arrhythmia diagnosis [5, 6].

In this paper, we proposed a novel pattern recognition approach, kernel difference-weighted k-nearest neighbor classifier KDF-KNN, for the diagnosis of cardiac arrhythmia based on the ECG recordings. The proposed method first defines the weighted KNN rule as the constrained least-squares optimization of sample reconstruction from its neighborhood, and we then use the Lagrangian multiplier method to seek a simple and efficient solution to compute the weights of different nearest neighbors by introducing the Gram matrix G. Taking into account the nonlinear structure information, we further extend difference-weighted KNN to its kernel version, Kernel difference-weighted k-nearest neighbor (KDF-KNN), by defining the kernel distance and the kernel Gram matrix. KDF-KNN weighs the nearest neighbors by using both the norm and correlation of the differences between the unclassified sample and its nearest neighbors. Thus KDF-WKNN is expected to achieve better classification performance than KNN.

In arrhythmia analysis, it is the unavoidable that some attribute values of a person would be missing. Thus, this paper further introduces a modified KDF-KNN method to address this problem. To evaluate the performance of KDF-WKNN, we use the UCI cardiac arrhythmia database which contains 452 instances with 245 normal and 207 arrhythmia (15 subclasses) instances. Our experimental results indicate that KDF-WKNN is superior to the original KNN and distance-weighted KNN, and is very competitive while compared with Naïve Bayesian classifier, KNN, and VFI5 in terms of classification accuracy.

2. Methods

2.1. Description of data set

The Cardiac Arrhythmia Database from the UCI Machine Learning Repository is used [7]. This data set contains 452 instances of samples from 16 classes. The first class is "Normal", and the other 15 classes are 15 kinds of arrhythmia. For each sample, there are 279 attributes, where the first four, age, sex, height, and weight, are the general description of the participant, and the other 276 attributes are extracted from the standard 12 lead ECG recordings. For the details of the data set, please refer to [4,6].

There are two significant characteristics which should be noted for the UCI cardiac arrhythmia database. First,

the distribution of class labels is imbalanced. For the "Normal" class, here are 245 instances of samples. For three kinds of cardiac arrhythmia class (the first, second, and third degree AtrioVentricular block), there is none instance of sample. Second, there are several missing attribute values (about 0.33%).

2.2. KDF-WKNN

Let $\{(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_m, y_m)\}$ be a training set, where $\mathbf{x}_i$ is the ith training sample, and y_i is the corresponding class label. Given an unclassified sample $\mathbf{x}$, a distance metric is first used to obtain the first k nearest neighbors $\{\mathbf{x}_1^{NN}, \cdots, \mathbf{x}_k^{NN}\}$ and their corresponding class labels $\{y_1^{NN}, \cdots, y_k^{NN}\}$, $y_i^{NN} \in \{\omega_1, \cdots, \omega_c\}$. Sample-weighted KNN classifier is then operated for classification by running the following two steps:

(1) Assign each nearest neighbor $\mathbf{x}_i^{NN}$ a weight w_i using a weight algorithm;

(2) Assign the sample $\mathbf{x}$ a class label $\omega_{j_{\max}}$ using the following rule:

$$\omega_{j_{\max}} = \arg\min_{\omega_j} (\sum_{y_i^{NN} = \omega_j} w_i). \tag{1}$$

In the following, we introduce a kernel difference-weighted k-nearest neighbors (KDF-WKNN) classifier to assign the appropriate weights on nearest neighbors [7]. Given two samples $\mathbf{x}$ and $\mathbf{x}'$, a kernel distance in the feature space could be defined by introducing the kernel function $k(\mathbf{x}, \mathbf{x}') = (\Phi(\mathbf{x}) \cdot \Phi(\mathbf{x}'))$,

$$d(\mathbf{x}, \mathbf{x}') = \|\Phi(\mathbf{x}) - \Phi(\mathbf{x}')\|^2 = k(\mathbf{x}, \mathbf{x}) - 2k(\mathbf{x}, \mathbf{x}') + k(\mathbf{x}', \mathbf{x}'). \tag{2}$$

Two popular kernel functions are radial basis function (RBF) kernel $k(\mathbf{x}, \mathbf{x}') = \exp(-\|\mathbf{x}-\mathbf{x}'\|^2/2)$ and polynomial kernel $k(\mathbf{x}, \mathbf{x}') = (1 + \mathbf{x} \cdot \mathbf{x}')^d$.

After the definition of the kernel distance, we define the weight assignment as a constrained optimization problem of sample reconstruction from its neighborhood:

Problem. The weights of the nearest neighbors $\mathbf{w} = [w_1, \cdots, w_k]^T$ is defined as a vector corresponding to the constrained optimal reconstruction of $\Phi(\mathbf{x})$ using $\mathbf{X} = [\Phi(\mathbf{x}_1^{NN}), \cdots, \Phi(\mathbf{x}_k^{NN})]$,

$$\mathbf{w} = \arg\min_{\mathbf{w}} \frac{1}{2} \|\Phi(\mathbf{x}) - \mathbf{X}\mathbf{w}\|^2$$

$$\text{s. t. } \sum_i w_i = 1. \tag{3}$$

Obviously, the objective of the constrained least-squares optimization problem is a quadratic function while the constraint is a linear, thus is formulated as a quadratic programming (QP) problem. We then use the Lagrangian multiplier method to seek a simple and efficient solution to this QP problem. Let

$\mathbf{D} = [\Phi(\mathbf{x}) - \Phi(\mathbf{x}_1^{NN}), \cdots, \Phi(\mathbf{x}) - \Phi(\mathbf{x}_k^{NN})]^T$. The optimization problem can be rewritten as

$$\mathbf{w} = \arg\min_{\mathbf{w}} \frac{1}{2} \mathbf{w}^T \mathbf{D}\mathbf{D}^T \mathbf{w}$$

$$\text{s.t. } \sum_i w_i = 1 \tag{4}$$

From Eq. (4), the optimal weights $\mathbf{w}$ are only dependent on the difference and its k nearest neighbors, $\{\Phi(\mathbf{x}) - \Phi(\mathbf{x}_1^{NN}), \cdots, \Phi(\mathbf{x}) - \Phi(\mathbf{x}_k^{NN})\}$. Thus we name our method rule as kernel difference-weighted KNN.

The Lagrangian function for the constrained optimization problem is

$$L(\mathbf{w}, \lambda) = \frac{1}{2} \mathbf{w}^T \mathbf{D}\mathbf{D}^T \mathbf{w} - \lambda(\mathbf{w}^T \mathbf{1}_k - 1), \tag{5}$$

where $\mathbf{1}_k$ is a $k \times 1$ vector with each element equal to 1. Let the kernel Gram matrix $\mathbf{G}^k = \mathbf{D}\mathbf{D}^T$, and $\nabla_{\mathbf{w}} L(\mathbf{w}, \lambda) = 0$, $\nabla_{\lambda} L(\mathbf{w}, \lambda) = 0$. We can derive an efficient and numerically stable approach to compute the weights $\mathbf{w}$. This approach first solves the system of linear equations

$$\mathbf{G}\mathbf{w}_0 = \mathbf{1}_k, \tag{6}$$

and then rescales the weights using

$$\mathbf{w} = \mathbf{w}_0 / (\mathbf{w}_0^T \mathbf{1}_k). \tag{7}$$

Actually, the kernel Gram matrix $\mathbf{G}^k$ is constructed by defining the element g_{ij}^k of the kernel gram matrix $\mathbf{G}^k$ as

$$g_{ij}^k = ((\Phi(\mathbf{x}) - \Phi(\mathbf{x}_i^{NN})) \cdot (\Phi(\mathbf{x}) - \Phi(\mathbf{x}_j^{NN}))). \tag{8}$$

Using the kernel trick, g_{ij}^k can be calculated explicitly

$$g_{ij}^k = k(\mathbf{x}, \mathbf{x}) - k(\mathbf{x}, \mathbf{x}_i^{NN}) - k(\mathbf{x}, \mathbf{x}_j^{NN}) + k(\mathbf{x}_i^{NN}, \mathbf{x}_j^{NN}). \tag{9}$$

We can further derive a more compact expression of the kernel matrix $\mathbf{G}^k$,

$$\mathbf{G}^k = \mathbf{K} + \mathbf{1}_{kk} k(\mathbf{x}, \mathbf{x}) - \mathbf{1}_k \mathbf{k}_c^T - \mathbf{k}_c \mathbf{1}_k^T, \tag{10}$$

where $\mathbf{K}$ is a $k \times k$ Gram matrix with the element $k_{ij} = k(\mathbf{x}_i^{NN}, \mathbf{x}_j^{NN})$, $\mathbf{1}_{kk}$ is a $k \times k$ matrix with each element equal to 1, $\mathbf{1}_k$ is a $k \times 1$ vector with each element equal to 1, $\mathbf{k}_c$ is a $k \times 1$ vector with the ith element equal to $k(\mathbf{x}, \mathbf{x}_i^{NN})$.

In some cases, the inverse of the matrix $\mathbf{G}$ is not unique if the matrix $\mathbf{G}$ is singular (e.g., the number of nearest neighbors, $k > N$, the dimension of the sample). To avoid this case, we adopt a regularization method by adding a small multiple of the identity matrix,

$$\mathbf{G}^k = \mathbf{G}^k + \eta \text{tr}(\mathbf{G}^k)/k, \tag{11}$$

where $\text{tr}(\mathbf{G})$ denotes the trace of the matrix $\mathbf{G}$, and $\eta = 10^{-0} \sim 10^{-3}$ is the regularization parameter. DF-WKNN can utilize the correlation of the samples.

Finally, we briefly summary the main steps of KDF-WKNN. Given an unclassified sample $\mathbf{x}$, KDF-WKNN first obtains the first k nearest neighbors $\{\mathbf{x}_1^{NN}, \cdots, \mathbf{x}_k^{NN}\}$ and their corresponding class labels $\{y_1^{NN}, \cdots, y_k^{NN}\}$, and then

calculates the kernel Gram matrix $\mathbf{G}^k$. The weights $\mathbf{w}$ of k nearest neighbors are then determined by solving the system of linear equations $[\mathbf{G}^k + \eta \mathrm{tr}(\mathbf{G}^k)/k]\mathbf{w} = \mathbf{1}_k$.

2.3. Addressing the missing attribute values

In this subsection, we presented a modified principal component analysis (PCA) approach to cope with the missing attribute values problem. Let $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$ be a training set with N samples, where sample $\mathbf{x}_j = [\mathbf{x}_j^{(1)}, \cdots, \mathbf{x}_j^{(d)}]^T$ is a d-dimensional vector. The total covariance matrix $\mathbf{S}_t$ of classical PCA is then defined as

$$\mathbf{S}_t = \frac{1}{N}\sum_{j=1}^{N}(\mathbf{x}_j - \overline{\mathbf{x}})(\mathbf{x}_j - \overline{\mathbf{x}})^T = \frac{1}{N}\sum_{j=1}^{N}\mathbf{x}_j\mathbf{x}_j^T - \overline{\mathbf{x}}\overline{\mathbf{x}}^T , \quad (12)$$

where $\overline{\mathbf{x}} = [\overline{x}^{(1)}, \cdots, \overline{x}^{(d)}]^T$ is the mean vector of all training images. If there are missing attribute values for some instances of samples, we introduce a new variable $z_j^{(i)}$ to denote whether the attribute value $x_j^{(i)}$ is missing,

$$z_j^{(i)} = \begin{cases} 0, & \text{if } x_j^{(i)} \text{ is missing} \\ 1, & \text{else} \end{cases} . \quad (13)$$

With the variable $z_j^{(i)}$, the mean vector is redefined as

$$\overline{x}^{(i)} = \frac{1}{N}\sum_{j}^{N} x_j^{(i)} z_j^{(i)} . \quad (14)$$

Similarly, the (k,l)th element of the total scatter matrix $\mathbf{S}_t$ can be calculated by

$$\mathbf{S}_t(k,l) = \frac{1}{N}\sum_{j=1}^{N} x_j^{(k)} x_j^{(l)} z_j^{(k)} z_j^{(l)} - \overline{x}^{(k)}\overline{x}^{(l)} . \quad (15)$$

The PCA projector $\mathbf{T}_{pca} = [\varphi_1, \varphi_2, \cdots, \varphi_{d_{PCA}}]$ can be obtained by calculating the eigenvalues and vectors of the total scatter matrix $\mathbf{S}_t$, where φ_k is the kth eigenvector corresponding to the kth largest eigenvalue of $\mathbf{S}_t$, and d_{PCA} denotes the PCA dimension.

After obtaining the PCA projector, we can use it to cope with the missing attribute values. Given a sample $\mathbf{x}$, the classical PCA projection method calculates the feature vector $\mathbf{y}$ by

$$\mathbf{y} = \mathbf{W}^T\mathbf{x} , \quad (16)$$

The feature vector $\mathbf{y}$ can also be calculated by minimizing the 2-norm objective function

$$J_1(\mathbf{y}) = \|\mathbf{x} - \mathbf{W}\mathbf{y}\|_2^2 = \sum_{i=1}^{d}(x_i - \mathbf{W}_i\mathbf{y})^2 , \quad (17)$$

where $\mathbf{W}_i$ is the ith row of $\mathbf{W}$. From Eq. (17), the feature vector $\mathbf{y}$ corresponds to

$$\frac{\partial J_1(\mathbf{y})}{\partial \mathbf{y}} = 2\mathbf{W}^T\mathbf{W}\mathbf{y} - 2\mathbf{W}^T\mathbf{x} = 0 , \quad (18)$$

$\mathbf{W}$ is a set of orthogonal basis, $\mathbf{W}^T\mathbf{W} = \mathbf{I}_d$, where denotes a

$d \times d$ identity matrix. It is simple to see that $\mathbf{y} = \mathbf{W}^T\mathbf{x}$.

If there are missing attribute values for the sample $\mathbf{x}$, the objective function in Eq. (17) should be modified to

$$J(\mathbf{y}) = \sum_{i=1}^{d} z_i^2 (x_i - \mathbf{W}_i\mathbf{y})^2 . \quad (19)$$

The feature vector $\mathbf{y}$ is then calculated by

$$\frac{\partial J(\mathbf{y})}{\partial \mathbf{y}} = 2\sum_{i=1}^{d} z_i\mathbf{W}_i^T\mathbf{W}_i\mathbf{y} - 2\sum_{i=1}^{d} z_i\mathbf{W}_i^T x_i = 0 ,$$

$$\mathbf{y} = \left(\sum_{i=1}^{d} z_i\mathbf{W}_i^T\mathbf{W}_i\right)^{-1} \sum_{i}^{d} z_i\mathbf{W}_i^T x_i . \quad (20)$$

Using feature vector $\mathbf{y}$, the reconstructed sample $\mathbf{x}'$ is

$$\mathbf{x}' = \mathbf{W}\mathbf{y} . \quad (21)$$

Using the reconstructed sample $\mathbf{x}'$, we use the following strategy to recovery the missing attribute values. If the missing attribute is a real variable, we use the reconstructed value to represent it. If the missing attribute is a binary variable, we use the rounding operation on the reconstructed value to recovery the missing attribute.

3. Experimental results

In our experiments, we randomly split the UCI cardiac arrhythmia database into 10 folds, and use a 10-fold cross validation (CV) method to determine the classification accuracy. To reduce bias in evaluating the performance, we calculate the average of the classification accuracy of the 10 runs of 10-fold CV.

Table 1 Classification accuracy of five methods

Method	NN [6]	NB [6]	VF15 [6]	VF15-FW [6]	KDF-WKNN
Accuracy (%)	53	50	62	68	70.66

Table 1 lists the classification accuracy of five methods, nearest neighbor (NN), naïve Bayesian classifier (NB), VF15, feature weighted VF15 (VF15-FW), and KDF-WKNN. The classification accuracy of KDF-WKNN is 70.66, which is higher than that of the other four methods.

To further discuss the performance of the proposed cardiac arrhythmia diagnosis method, we also provide the confusion matrix of KDF-WKNN, as shown in Table 2. As stated in Section 2.1, there is none instance of sample for class 11, 12 and 13, and thus the listed confusion matrix is a 13×13 matrix. From the confusion matrix, KDF-WKNN is effective to distinguish "normal", "Old Anterior Myocardial Infarction", "Left bundle branch block", and "Right bundle branch block". But for the other nine classes of cardiac arrhythmia, more samples and further investigation would be required.

Table 2 Confusion matrix of KDF-WKNN

	01	02	03	04	05	06	07	08	09	10	14	15	16
01	**235**	8	0	0	0	0	0	0	0	2	0	0	0
02	22	**21**	0	0	0	0	0	0	0	0	0	1	0
03	0	3	**12**	0	0	0	0	0	0	0	0	0	0
04	7	1	0	**7**	0	0	0	0	0	0	0	0	0
05	9	0	0	0	**2**	0	0	0	0	2	0	0	0
06	23	0	0	0	0	**1**	0	0	0	1	0	0	0
07	2	0	1	0	0	0	**0**	0	0	0	0	0	0
08	2	0	0	0	0	0	0	**0**	0	0	0	0	0
09	0	0	1	0	0	0	0	0	**8**	0	0	0	0
10	15	0	1	0	0	0	0	0	0	**34**	0	0	0
14	2	1	0	0	0	0	0	0	0	0	**0**	0	1
15	1	2	0	0	0	0	0	0	0	0	0	**2**	0
16	18	0	1	0	0	0	0	0	0	2	0	1	**0**

4.   Conclusions

In this paper, we proposed a kernel difference-weighted classifier (KDF-WKNN) for the diagnosis of cardiac arrhythmia from standard 12 lead ECG recordings. Our Experimental results on the UCI cardiac arrhythmia database show that KDF-WKNN outperforms NN, NB and VF15 in terms of classification accuracy.

Acknowledgements

References

[1] Thakor NV, Zhu YS. Applications of adaptive filtering to ECG analysis: Noise cancellation and arrhythmia detection. *Biomedical Engineering*. 1991; 38(8): 785-794.

[2] Coast DA, Stern RN, Cano GG, Briller SA. An approach to cardiac arrhythmia analysis using hidden Markov models. *Biomedical Engineering*. 1990; 37(9): 826-836.

[3] Lima CS, Cardoso MJ. Cardiac arrhythmia detection by parameters sharing and MMIE training of hidden Markov models. *Engineering in Medicine and Biology Society*. 2007 Aug; 3836 – 3839.

[4] Guvenir HA, Acar B, Demiroz G, Cekin A. A supervised machine learning algorithm for arrhythmia analysis. *Computers in Cardiology*. 1997 Sep; 433 – 436.

[5] Moody GB, Mark RG. The impact of the MIT-BIH arrhythmia database. *Engineering in Medicine and Biology Magazine*. 2001; 20(3): 45 – 50.

[6] Blake CL, Merz CJ. UCI Repository of Machine Learning Databases. 1998. Available from: http://www.ics.uci.edu/~mlearn/MLRepository.html

[7] Zuo W, Zhang D, Wang K. On kernel difference-weighted k-nearest neighbor classification. *Pattern Analysis and Applications*. 2008; 11: 247-257.

Address for correspondence

Wangmeng Zuo
Mailbox 332, Harbin Institute of Technology
Harbin 150001, China
E-mail: cswmzuo@gmail.com