Biomedical Signal Processing and Control 41 (2018) 152–160

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Feature fusion for imbalanced ECG data analysis

Wei Lu, Honghui Hou, Jinghui Chu*

School of Electronic Information Engineering, Tianjin University, Tianjin 300072, PR China

ARTICLE INFO

Article history:
Received 19 July 2017
Received in revised form 13 October 2017
Accepted 19 November 2017

Keywords:
Electrocardiogram signals
Imbalanced data set
Feature fusion
2D-convolutional neural network
Shallow characteristic
Random forest

ABSTRACT

World Health Organization (WHO) indicates that cardiovascular disease remains challenging in diagnosis and treatment. The electrocardiogram (ECG) is a very important diagnostic assistant for cardiac diseases. Traditionally, most of the ECG analysis methods are evaluated by their intra-patient performance, which however may not suitable for inter-patient cases. Here, we propose a complete classification system with excellent generalization ability. We first extract the 2D-convolutional and PQRST features of a single heartbeat after preliminary processing. We then balance the data with the Random Over Sampler algorithm after comparing several imbalanced algorithms. Finally, we use a Random Forest (RF) classifier to classify the data according to the Association for the Advancement of Medical Instrumentation (AAMI) standards (1988). Results show that Recall$_M$ (MR), Precision$_M$ (MP) and Fscore$_M$ (MF) of our proposal are all above 99%. In order to evaluate the performance of different methods, we designed inter-patient and intra-patient experiments separately. To further demonstrate the robust and adaptability of our model, we then transferred it to another data set and performed the experiment. In our experiments, the values of macro- and micro-metrics are up to 99%. All of the results are averages of five experiments, and the Average Accuracy (AA) of experiments applied here are above 99%, which illustrates that our proposal is a promising alternative and superior to most of the state-of-the-art methods.

1. Introduction

According to the data of the World Health Organization, 30% of global deaths are due to cardiovascular diseases (CVDs). Thus, the global burden of cardiovascular is still the most important healthcare issue around the world [1]. Arrhythmias are the most typical and important cardiovascular diseases, which may cause temporary shock and even sudden death. The best assistant to monitor and diagnose the arrhythmias is ECG, which is a visual signal captured or measured by placing electrodes on the surface of the body to detect voltage changes. Cardiologists often analyzed the ECG directly in the past, whereas computer-aided classification of arrhythmias has become popular more recently [2].

A fully automatic classification of arrhythmias includes four parts: ECG processing, heartbeat segmentation, feature extraction, and classification. The goal of ECG processing is to make the signal clearer and lay the foundation for subsequent experiments, i.e. ECG denoising [3–5], detection of characteristic points [6], etc. In the second phase, the ECG is divided into single or multiple periods of heartbeats by using the heartbeats frequency information [7,8]. In the last stage, common classification methods such as Artificial Neural Network (ANN) [9], Bayesian Network (BN) [10], Random Forest (RF) [11], Support Vector Machine (SVM) [12,13], etc. have been used to obtain the true class of the samples.

Feature extraction plays an important role throughout the process, and various methods have been proposed and validated. For classical methods, features are exacted from the time domain such as R-R intervals [14,15] and QRS width [16] as well as frequency domain for example S-transform [17,18], wavelet transform [9,15,19–24], Fourier transform [20,25,26], Modified Cosine Transform [19,20], etc. After feature exaction, there would be feature selection to remove related characteristics and reduce dimensions to improve the final accuracy, generally including principal component analysis (PCA) [9,21], linear discriminant analysis (LDA) [21], decision tree (DT) [22], independent component analysis (ICA) [14,15,21], etc. However, these methods are usually used for extracting handcrafted features from the ECG waveforms, and one of issues is incomplete use of information provided by the source data. Deep learning techniques can overcome this shortcoming. Previously [12], an electrocardiogram beat classification method was proposed based on Deep Belief Networks (DBN) with features extracted by DBN and timing interval. In [27], the authors used a labeled HRV data sets to train a Convolutional Neural Networks (CNN) model as a supervised approach, and used Stacked Autoencoders with Restricted Boltzmann Machines to obtain unlabeled features. According to the literature [28,29], 1D-CNN could

---

* Corresponding author.
E-mail addresses: luwei@tju.edu.cn (W. Lu), slyviahou@tju.edu.cn (H. Hou), cjh@tju.edu.cn (J. Chu).

https://doi.org/10.1016/j.bspc.2017.11.010
1746-8094/

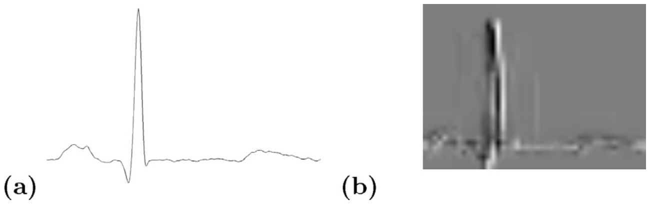

Fig. 1. The figures a and b stand for the original heartbeat and its feature map after convolution operation separately. The difference between each other is obvious.

be used for patient-specific classification with patient-specific and relatively small common data, which could save training time and be easily applied to the wearable device. An approach using Staked Denoising Autoencoders (SDAEs) with sparsity constraint for unsupervised feature representation and Active Learning (AL) to fine tuning Deep Neural Network (DNN) could efficiently match the statistical distribution of the data [30], etc.

All of the methods deal with 1D-ECG signals, and we put forward an idea of switching 1D-ECG signals into 2D-images for extracting convolutional features for the first time since others [28] proved that the simple convolutional neural network could improve the training accuracy compared with other methods. However, we found that the 2D-convolutional feature maps are blurred (Fig. 1), especially at the key points (i.e., PQRST positions) which can affect experimental results. Meanwhile, ECG data sets are extremely imbalanced, and this can reduce the recognition accuracy of the minority class. To reconcile those shortcomings, we proposed a classification system (Fig. 2). Firstly, single heartbeat signals are obtained by preprocessing and segmentation. Secondly, PQRST features are exacted from single heartbeat signals for making up the blurred issue of convolutional maps and 2D images are created by connecting the dots of the 1D single heartbeat signals. Then, the highly abstract features of a simple CNN are extracted and fused with PQRST features. After imbalanced processing, the fused features of balanced data set are classified by a simple RF classifier. This is the first time that deep features (CNN) are fused with shallow features (PQRST) for extracting representative features to the best of our knowledge although feature fusion has been reported in shallow-with-shallow [31] and deep-with-shallow [12]. To prove the generalization of the model, we use another data set to evaluate the performance of our model.

The rest of the paper is organized as follows: Section 2 introduces the main methods used in this paper involving feature extraction, imbalanced data processing, the classifier and learning rate. Section 3 explains the publicly available ECG data sets and the standards we used as well as performance metrics for experiments. Section 4 is results and discussion. Finally, conclusions and future directions are detailed in Section 5.

2. Methods

Our research concentrates on dealing with imbalanced data processing, feature fusion, parameter adjustment as well as classifier design. All of the algorithms in this section expound around these four parts.

For clearer presentation, we establish here some of the basic notions used in most of the subsections. Considering the original data set after selecting and heartbeat segmentation OS has m examples, we define: $OS = \{(x_i, y_i)\}$, $i = 1, \ldots, m$, where $x_i \in X = \{f_1, f_2, \ldots f_n\}$ is a n-dimensional instance. $y_i \in Y = \{1, 2, \ldots z\}$ is the label of x_i. Balanced data set based on original data $BOS = \{(x_{newi}, y_{newi})\}$, $i = 1, \ldots, M$ has N dimensions where M is the sum of balanced data set.

2.1. Feature fusion

2.1.1. Convolutional features

CNN is a typical method for deep learning based on its characteristics that automatically highlights and extracts the most valuable high dimensional features through convolutional operation from the input data. It pays more attention to the local features and their positions, i.e., the position among other features is determined when the local feature is extracted. Otherwise, on the same feature maps of CNN, the weights of neurons are the same resulting in network learning in parallel, which markedly saves learning time. In this paper, considering that samples belong to simple line types, we proposed a simple CNN that contains two convolutional layers, two Pooling layers, three Full-Connected (FC) layers and two Dropout layers (Fig. 3).

As explained in [27], the output of convolutional layers $C_{x,y}$ can be computed according to Eq. (1):

$$C_{(x,y)} = h\left(\sum_{i=1}^{k_m}\sum_{j=1}^{k_n} OS_{(x'+i,y'+i)} \times w_{(i,j)} + b_{(i,j)}\right), \tag{1}$$

where the two-dimensional input data is $OS_{(x,y)}$, the kernel size is (k_m, k_n) and the steps of convolution are (s_x, s_y), $w_{(i,j)}$, $b_{(i,j)}$ are weights and bias of the kernel and $x' = x \cdot s_x - 2$, $y' = y \cdot s_y - 2$.

Applying to our designed network, we used 96 * 72 two-dimensional gray images to simulate digital heartbeats obtained from MIT-BIH data set and utilized four convolutional kernels whose sizes are all $\left\{\{1, 1, 1\}, \{1, -7, 1\}, \{1, 1, 1\}\right\}$ to emphasize edge information in first layer, and eight convolutional kernels with $(5, 5)$ size for second convolutional layer to detect the edges of the blurred feature maps of first layer. Because the texture information is more valuable than the background, we used a max-pooling method (selecting the max-value of neighborhood characteristics corresponding to the pooling window) and set $P_x = 2$, $P_y = 2$, which reduces the number of features in each feature map by four. In the fully connected part, we assembled a network with a hidden layer and set 200, 100, 5 neurons per layer. In that case, two dropout layers [32] are embedded into every two FC layers for the purpose of enhancing the network generation ability and preventing overfitting. The training and testing process of the dropout layers is presented in Eq. (2):

$$\begin{aligned} Train: D_{(i,j)} &= w_{(i,j)}\big|_p \times fc_{(i,j)} + b_{(i,j)}, \\ Test: D_{(i,j)} &= w_{(i,j)} \times fc_{(i,j)} \times P + b_{(i,j)}. \end{aligned} \tag{2}$$

Here, $fc_{(i,j)}$, $D_{(i,j)}$ are the input and output of the layer, $w_{(i,j)}\big|_p$ is the weight of selecting specific number of neurons according to retaining probability P, $w_{(i,j)}$ is the weight of all of neurons. Relu is adopted as the activation function of each layer except that softmax is used in the final classification.

The determination of the 2D-CNN consists of data feed forward pass and error back-propagation pass [35]. We used a cross-entropy cost function to adjust the network parameters, which were expressed as Algorithm 1:

Algorithm 1. Cross-entropy cost function for network parameter adjustment.

```
Require: $(x_1, y_1), \ldots, (x_m, y_m), \ldots, (x_n, y_n)$
Ensure: $w'_{i,j}, b'_{i,j}$
Randomly select $k$ train samples from original data set;
Initialization: $w_{(i,j)} \to 0$, $b_{(i,j)} \to 0$, $lr \to r$;
for i=1 to $k$ do
  Storage target output $S_i$;
  Calculate output vector $I_{(i,j)}$ of intermediate layer and actual output $A_i$ of last layer;
  Cost calculation:
```