Pay Attention and Watch Temporal Correlation: A Novel 1-D Convolutional Neural Network for ECG Record Classification

Yongchao Wang[1], Bin Xiao[1*], Xiuli Bi[1], Weisheng Li[1], Junhui Zhang[2*], Xu Ma[3]

[1] Chongqing Key Laboratory of Image Cognition, Chongqing University of Posts and Telecommunications, Chongqing, China
[2] The First Affiliated Hospital of Chongqing Medical University, Chongqing, China
[3] Human Genetics Resource Center, National Research Institute for Family Planning, Beijing, China

Abstract

Currently, many researches focus on the classification of heart beats. It is rare to study the classification of the entire ECG records. In fact, these two issues are completely different, record level classification is more difficult than beat level classification. In this paper, we propose a record level ECG classification method by combining 1-D deep convolutional neural network (CNN) and long short-term memory network (LSTM). The proposed method utilizes CNN to automatic extract features of ECG patches without complex data pre-processing or artificial feature scheme, and then utilizes two layers LSTM to model the underlying temporal correlation relation among ECG patches. In order to alleviate vanishing gradient problem and enhance representation power of features, we combine residual connection and attention mechanisms to build a residual attention module. Stacked residual attention modules not only deepen the network, but also refine the feature maps from both spatial and channel dimensions. We validate the proposed method through the dataset provide by The China Physiological Signal Challenge 2018 (6,877 ECG records, 9 classes), the cross-validation result of F1 score is 0.78. Experiments on PhysioNet/CinC 2017 (8,528 ECG records, 4 classes) reveal that the F1 score is 0.809.

1. Introduction

The mortality ration of cardiovascular disease (CVD) is much higher than other diseases including cancer and AIDS. Electrocardiogram (ECG) which records changes in electrical activity, is a non-invasive and widely used CVD diagnostic tool. Analysis of the ECG is the first step in the treatment of CVD, but ECG analysis relies heavily on the doctor's experience. Therefore, developing a computer-aided diagnostic technique to automatic analysis ECG is significant.

The process of most existing researches on beat level ECG classification can be divided into three steps: data pre-processing, feature extraction and classification. The pre-processing of ECG records mainly includes: ECG signal denoising, baseline wanders removing, frequency domain transformation, heart beats segmentation, and etc. In the step of feature extraction, researchers have proposed various types of features, such as 1) statistical features which use statistic to describe some characteristic of the data [1]; 2)ECG morphological features that include RR interval, heart rate variability, QRS complex wave, duration, amplitude, etc.; 3) transition features which mainly include Hermite transform coefficients, wavelet transform features [2,3,4], and 4) convolutional features which are extracted by convolutional neural networks [5,6]. Finally, the classifier mainly includes Nearest Neighbor Classifier, K Nearest Neighbor Classifier, Support Vector Machine, Multi-Layer Perception, Fully Convolutional Networks and etc..

Compared with the hot research of beat level ECG classification, the study of record level ECG classification is slightly silent. The main reasons are: Firstly, the benchmark database is too small to support record level ECG classification. Such a small number of ECG records is far from the requirement for deep learning approach which is data-driven. In recent years, these datasets even can't satisfy the requirements of beat-level ECG classification, and many methods have achieved a nearly perfect result. Secondly, an ECG record contains a lot of heart beats, but not all heart beats have the same type. In extreme cases, only a few heart beats are abnormal beats in a patient's ECG record. Therefore, record level ECG classification is a more difficult topic than that of beat level. In this paper, benefit from the datasets provided by PhysioNet/CinC Challenge 2017 denoted by CPSCDB [7] and the China Physiological Signal Challenge 2018 denoted by CINCDB [8], we implement the record level ECG classification on the larger dataset. In this paper, a novel 1-D convolutional neural network with attention mechanism and temporal correlation is proposed, and obtained excellent performance.

2.   Method

The main processes of the proposed method are as follow: firstly, the raw ECG records are segmented into several 30s long patches. Secondly, feature maps are extracted by CNN with residual attention modules, and followed by sending to the two-layer LSTM to extract temporal correlation relation. Finally, fully connected layers are used to classify and obtain the record level results based on a majority voting decision strategy.
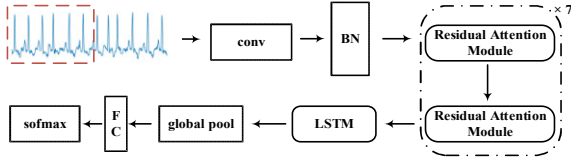

Figure 1: The proposed Neural Network

2.1.   Patch segment

There are no complicated pre-processing steps such as signal denoising, baseline wanders removing or spectrogram converting in the proposed method. We just only segment ECG record into several 30s long patches with 5s overlap. Note that 30 seconds is much longer than the duration of a heartbeat and if there are patches less than 30s, using 0 padding to 30s.

2.2.   Residual attention module

In this method, the neural network can be divided into a convolutional neural network part and a recurrent neural network part. The residual attention module is a main component of the convolutional neural network part, by stacking the residual attention module, it is easier to train deeper networks, and each module inside can reinforce or suppress information streams in both space and channel dimensions.

The attention mechanism is partly inspired by the CBAM [9], most of the attention mechanisms only consider spatial attention, and only use global max-pooling operations or global average-pooling operations when aggregate global information. On the contrary, as Figure 2 shows, we sequentially capture channel attention and spatial attention and use both types of pooling operations to refine the global information.

In the channel attention block, spatial information is aggregated by using both global average-pooling and global max-pooling operations from the input feature maps along spatial dimension. Then the pooled one-dimensional feature vector is sent to Multi-Layer Perception (MLP), and the number of neural units in the hidden layer is reduced by the R reduction rate in the input feature vector dimension, the number of neural units in the output layer is same as input dimension. Note that in order to decouple

the two types pooling operations, the MLP structure is constructed separately for the two pooling branches instead of the shared MLP, which is different from the CBMA. Finally, we acquire the raw channel attention vector by element-wise sum the two MLP's output, and the final attention vector is normalized by the sigmoid function.

In the spatial attention block, similar with the channel attention block, channel information is aggregated by using global average-pooling and global max-pooling operations from the input feature maps along channel dimension. Then the two pooled feature vectors are merged into a one-dimensional raw attention vector using a conventional convolution operation with convolution kernel size 7 and stride 1. Finally, the raw attention vector is normalized to between 0 and 1 by the sigmoid function.

Figure 2 generally depicts the structure of the residual attention module: the input data is convolved by a conventional convolution operation and then the representative features are extracted in the channel and spatial dimensions by sequentially through the channel attention block and the spatial attention block. Meanwhile, the dimensions of the input data are expanded by two times using a 1x1 convolution, then these expanded feature maps are added to feature maps which adjusted by the attention mechanism to establish a skip connection for deal the vanishing gradients problem. In summary, the stacked residual attention module progressively abstracts and recalibrates features in two dimensions in an iterative manner and also enhanced gradient propagation.
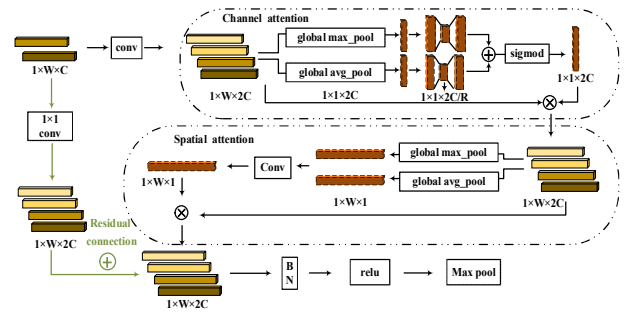

Figure 2: The residual attention module, the tensor shape is given in the format: height × width × channels.

2.3.   LSTM

The ECG record is a kind of time series data. The regular heartbeat makes the waveforms such as P wave, QRS wave and T wave appear periodically. These waveforms have important information for diagnosing CVD. LSTM is a kind of recurrent neural network that can model temporal relations. At each time step (i.e. each cycle), the recurrent connections allow a memory of previous inputs to persist in the network's internal state, which can be subsequently used to influence the final network output [10]. Therefore, it is reasonable to apply LSTM in ECG record classification due to the strong

correlation among ECG signal. The data for each time step is sequentially input to the LSTM, It is note that, as show in Figure 3, the input data of each time step is the value of all channels in the current time point. LSTM fuses the current information with the internal state information through three different gating mechanisms when processing each time step data.
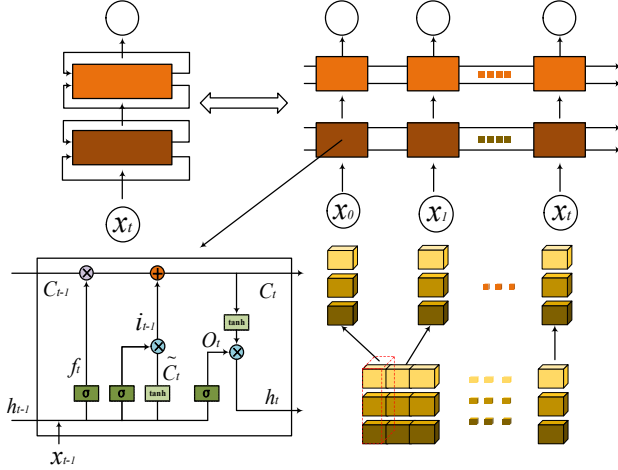

Figure 3: The LSTM (top left), the equivalent format (top right), the internal of an LSTM unit (bottom left), the input of LSTM (bottom right)

Figure 3 shows the LSTM network. In an LSTM unit, there are three gates that control the flow of information: forget gate, input gate and output gate. Forget gate determine how much the hidden state information of the prior moment will be used, forget gate determine how much candidate state $\tilde{c}_t$ will be used to update hidden state c_t, finally the output gate determine the output h_t of current unit.

2.4. Classification

We employ three fully-connected layers with softmax function to compute the classification probabilities.

In order to obtain the classification result of a record, the classification result of each ECG patch is counted, and the classification result of the record is obtained by using the voting mechanism. If a tie occurs, the patch with the highest probability determines the category of the record.

3. Experiments

3.1. Database

In order to fully prove the validity of the proposed method, we performed 5-fold cross-validation experiments on CPSCDB and CINCDB. Since the hidden data sets used

for the final score during the competition were unavailable to the public and still remain private, the data used in the experiments are from publicly training sets. Table 1 provides the detail description about these two datasets.

Table 1: Description of the CPSCDB and CINCDB.

Name	CPSCDB		CINCDB	
Total number	6,877		8,528	
Lead	12 leads		single lead	
Duration	6 s - 60 s		9 s - over 60 s	
Frequency	500 Hz		300 Hz	
Type And Number	Norma	918	Normal	5154
	AF	1098	AF	771
	I-AVB	704	Other rhythm	2557
	LBBB	207		
	RBBB	1695	Noisy	46
	PAC	556	-	-
	PVC	672	-	-
	STD	825	-	-
	STE	202	-	-

3.2. Module and Training

Table 2: The two columns refer to network on CPSCDB and CINCDB respectively. The brackets denote filter kernel size, stride, and growth rate of kernel. The square brackets denote the number of neural units.

Input: $1\times15000\times12$	Input: $1\times9000\times1$
conv, $(15,1,1)$	conv, $(15,1,1)$
$\begin{vmatrix} \text{conv, } (3,1,2\times) \\ \text{max pool } (5,2,1\times) \end{vmatrix} \times 6$	$\begin{vmatrix} \text{conv, } (7,1,2\times) \text{ R: 2} \\ \text{max pool } (5,2,1\times) \end{vmatrix} \times 6$
$\begin{vmatrix} \text{conv, } (3,1,2\times) \\ \text{max pool } (5,2,1\times) \end{vmatrix} \times 1$	$\begin{vmatrix} \text{conv, } (3,1,2\times) \\ \text{max pool } (5,2,1\times) \end{vmatrix} \times 1$
LSTM$\times2$, size:1536	LSTM$\times2$, size:128
FC$\times3$ [1024,768,9]	FC$\times3$ [1024,768,4]
Softmax	Softmax

Table 2 shows the detailed construction of our proposed network, ReLu nonlinear activation function and batch-normalization are ignored. Due to the different number of leads in the two datasets, the networks implemented in these two datasets have several different parameters. A weighted cross-entropy function measures the difference between the predicted probability distribution and the ground-truth probability distribution, because the unbalance of datasets. We employed the stochastic gradient descent (SGD) with 0.9 Nesterov momentum as optimizer. The batch size was set to 32, and for prevent over-fitting, weight decay of 10-4 and dropout layers with rate of 0.5 was also applied. Meanwhile, the initial learning rate is set to 0.01 for CPSCDB, and 0.1 for CINCDB, and divided by 5 after each 40 epochs.

3.3.   Results

To evaluate the performance of the proposed algorithm, we followed the challenge guidelines and metrics. As shown in Table 3, for CPSCDB, the F1 score is 0.782. The value of F19 is significantly lower than other values. The reason maybe the number of this type is too small to train network well.

Table 3: mean F1 scores (in %) in CPSCDB. F11 to F19 correspond 9 types respectively, Faf, Fblock, Fpc, Fst denote atrial fibrillation, branch block, premature contraction, ST-segment respectively. F1 is the mean of F11 to F19

F11	F12	F13	F14	F15	F16	F17
80.0	84.5	83.3	81.0	87.2	73.1	81.8
F18	F19	Faf	Fblock	Fpc	Fst	F1
79.0	55.3	84.5	85.5	77.6	74.5	78.2

Table 4: F1 scores (in %) for the proposed method in comparison with existing CNN-based methods.

methods	N	A	O	Overall
CNN [5]	87.8	79.0	70.1	79.0
CRNN [5]	88.8	76.4	72.6	79.2
Proposed	90.6	78.1	74.2	80.9

The 5-fold cross validation results shown in Table 4. Martin Zihlmann et al. [5] built a CNN and a Convolutional Recurrent Neural Networks (CRNN) to extract features from ECG logarithmic spectrograms for classification. We used the raw ECG data, and achieves 1.7 percentage higher in F1 score than this method. Since the hidden data set remains private, the comparison method results are from corresponding paper.

4.   Conclusion

The excellent results achieved in two datasets proves that the proposed method is competitive, and has good adaptability to different classification problems. It is potential to be extended to other ECG classification problems. Moreover, since both datasets have serious data imbalance, the experimental results also show that the proposed method is robust to data imbalance problems.

The main reason for those superiority of the proposed method is that our approach utilizes the stacked residual attention module to deepen the network, and extracts more hierarchical abstract features. The novel attentional mechanism in the proposed method enhances representative features and suppresses useless features in both spatial and channel dimensions, and continually optimizes refine feature during stacking. At the same time, since the ECG is time series data, the time correlation hidden in the ECG record is also extracted using LSTM in this proposed method.

References

[1] Hong, Shenda, et al. "Encase: An ensemble classifier for ecg classification using expert features and deep neural networks." 2017 Computing in Cardiology (CinC). IEEE, 2017.

[2] Güler, İnan, and Elif Derya Übeylı. "Ecg beat classifier designed by combined neural network model." Pattern recognition 38.2 (2005): 199-208.

[3] Ince, Turker, Serkan Kiranyaz, and Moncef Gabbouj. "A generic and robust system for automated patient-specific classification of ecg signals." IEEE Transactions on Biomedical Engineering 56.5 (2009): 1415-1426.

[4] Li, Hongqiang, et al. "Genetic algorithm for the optimization of features and neural networks in ecg signals classification." Scientific reports 7 (2017): 41011.

[5] Zihlmann, Martin, Dmytro Perekrestenko, and Michael Tschannen. "Convolutional recurrent neural networks for electrocardiogram classification." 2017 Computing in Cardiology (CinC). IEEE, 2017.

[6] Kiranyaz, Serkan, Turker Ince, and Moncef Gabbouj. "Real-time patient-specific ecg classification by 1-D convolutional neural networks." IEEE Transactions on Biomedical Engineering 63.3 (2015): 664-675.

[7] Liu, Feifei, et al. "An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection." Journal of Medical Imaging and Health Informatics 8.7 (2018): 1368-1373.

[8] Clifford, Gari D., et al. "Af classification from a short single lead ecg recording: the physioNet/computing in cardiology challenge 2017." 2017 Computing in Cardiology (CinC). IEEE, 2017.

[9] Woo, Sanghyun, et al. "Cbam: convolutional block attention module." Proceedings of the European Conference on Computer Vision (ECCV). 2018.

[10] Wang, Guijin, et al. "A global and updatable ecg beat classification system based on recurrent neural networks and active learning." Information Sciences (2018).

Address for correspondence:
Bin Xiao
School of Computer Science
No.2, Chongwen Road, Nan'an district, Chongqing, China
xiaobin@cqupt.edu.cn
Junhui Zhang
The First Affiliated Hospital of Chongqing Medical University
No. 1 Yixueyuan Road, Yuzhong District, Chongqing, China
2275610878@qq.com