A Restricted Boltzmann Machine Based Two-lead Electrocardiography Classification

Yan Yan
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
Email: yan.yan@siat.ac.cn

Xinbing Qin
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
Email: xb.qin@siat.ac.cn

Yige Wu, Nannan Zhang, Jianping Fan and Lei Wang*
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences

***Abstract*—An restricted Boltzmann machine learning algorithm were proposed in the two-lead heart beat classification problem. ECG classification is a complex pattern recognition problem. The unsupervised learning algorithm of restricted Boltzmann machine is ideal in mining the massive unlabelled ECG wave beats collected in the heart healthcare monitoring applications. A restricted Boltzmann machine (RBM) is a generative stochastic artificial neural network that can learn a probability distribution over its set of inputs. In this paper a deep belief network was constructed and the RBM based algorithm was used in the classification problem. Under the recommended twelve classes by the ANSI/AAMI EC57: 1998/(R)2008 standard as the waveform labels, the algorithm was evaluated on the two-lead ECG dataset of MIT-BIH and gets the performance with accuracy of 98.829% . The proposed algorithm performed well in the two-lead ECG classification problem, which could be generalized to multi-lead unsupervised ECG classification or detection problems.**

***Index Terms*—big data, electrocardiography classification, restricted Boltzmann machine, deep belief network.**

I. Introduction

The analysis of the electrocardiogram(ECG) has been a popular subject of research during the last three decades and provides valuable diagnostic information for many cardiac diseases. Nowadays, the tremendous amount of ambulatory ECG studies are underway because of it's abundant and rich information to the clinical point of view. At the same time, as the application of wearable devices and long-term monitoring equipment accumulate massive ECG data, the important works are efficient and high performance method for the ECG classification.

ECG classification has been regard as a complex pattern recognition problem for decades since ECG signals include non-stationary behaviours. Lots of digital signal processing methods like time domain and frequency domain methods had been well applied in ECG signal analysis. A cardiac arrhythmia beat is a rhythm of the heartbeat which might be irregular, faster, or slower than normal beats. The electrocardiogram signals were divided into vectors according to the R peaks of each waveform. High accuracy algorithms [1] had been proposed to locate the positions of the onset, peak and termination of individual components include the R peaks.

As well lots of methods had been used in the heartbeat features extraction and then classification, including time-domain features [2], frequency domain features [3], [4], morphological features, dynamic features [5]–[8]. Hidden Markov model(HMM) [9], support vector machine(SVM) [10], [11], independent component analysis(ICA) [12], artificial neural networks(ANN) [13], [14], fuzzy hybrid neural network(FNNs) [15] and extreme learning machine(ELM) [16] techniques are also employed in the classification problems. Some comparison works were also shown in [12], [17], [18]. Most of those classifiers rely on better features extracted from the ECG signals via complex algorithms. Some of the classification systems accomplished high performance due to the choice of some special features. The FNNs can be used to learn some features from the heartbeat, but not adapted to different heart rate ECG signals. The ANN method is tested on some features selected and the original heartbeat, it works well only when extracting useful features in simple classification. Majority of them classify the heartbeats according to ANSI/AAMI EC57 which groups the heartbeats into five classes. A drawback of those methods was that they depended on the supervised dataset and were weak in handling large amount of ambulatory ECG records.

In this paper, an unsupervised method of restricted Boltzmann machine(RBM) [19], [20] is proposed for the classification mission. Restricted Boltzmann machine is an energy based probability model which would be used to learn features from ambulatory electrocardiography dataset. Gradient descent algorithm [21] would be used to maximize the likelihood function. Then the weights learned from the RBM were used to construct a deep belief network(DBN) [22]–[26]. In the last layer, softmax model would be used to do multi-class classification. Fine-tuning the deep network by Hessian-free optimization [27] algorithm. Two-lead ECG records were used to train two classifiers with the optimization algorithm to combine two classifiers. Especially, the proposed method can be generalized to multi-lead ECG signals and achieve better results. The t-SNE [28] algorithm would be applied to visualize the high-dimensional heartbeats.

The MIT-BIH arrhythmia database [29] is the most widely used dataset to evaluate the performance for the classification and detection algorithm development. It is regard as the standard reference to fine-tune the deep network and assess the classifier performance. 400 electrocardiography records collected from the actual patients from hospitals would be used in the unsupervised learning approach, each containing about 24-hour long electrocardiography with three leads.

In the following sections, datasets, methods and optimization algorithm are shown in Section 2. The experiment results and the discussion are given in Section 3. And then the conclusion are illustrated in section 4.

TABLE I
MIT-BIH ARRHYTHMIA CLASSES

heartbeat classes Description	label	mapped label	symbol
normal beat	1	1	NORMAL
left bundle branch block beat	2	2	LBBB
right bundle branch block beat	3	3	RBBB
aberrated atrial premature beat	4	4	ABERR
premature ventricular contraction	5	5	PVC
fusion of ventricular&normal beat	6	6	FUSION
nodal (junctional) premature beat	7	7	NPC
atrial premature contraction	8	8	APC
ventricular flutter wave	31	9	FLWAV
ventricular escape beat	10	10	VESC
nodal (junctional) escape beat	11	11	NESC
atrial escape beat	34	12	AESC

[a] Recording 102, 104, 107, 217 are removed.
[b] The count lists in the table differ from other literatures due to the computation need.

II. ELECTROCARDIOGRAPHY DATASETS AND ARRHYTHMIA CLASSES

A. The datasets

An ambulatory electrocardiography dataset was used for training in the proposed model, which contains 400 records each consist of a 24-hours long electrocardiography data with three leads which were bandpass filtered at 0.1 100Hz and digitized at 128Hz.

The MIT-BIH arrhythmia database [29] contains 48 recordings. Each of which is slightly over 30 minutes long and digitized at 360Hz. In most records, the upper signal is a modified limb lead II(MLII), obtained by placing the electrodes on the chest. The lower signal is usually a modified lead VI(cccasionally V2 or V5, and in one instance V4), the electrodes are also placed on the chest. Most labels were placed at the R-wave peak, but manually inserted labels were not always located precisely at the peak. Recording 102, 104, 107, 217 containing paced beats are removed.

B. Arrhythmia classes and labels

The MIT-BIH Arrhythmia database [29] has 40 kinds of annotation pertained to twelve labels. There were twelve labels used in the 44 records so that the heartbeats were grouped into twelve classes: NORMAL(1)–normal beat, LBBB(2)–left bundle branch block beat, RBBB(3)–right bundle branch block beat, NESC(11)–nodal (junctional) escape beat, AESC(34)–atrial escape beat, ABERR(4)–aberrated atrial premature beat, NPC(7)–nodal (junctional) premature beat, APC(8)–atrial premature contraction, VESC(10)–ventricular escape beat, PVC(5)–premature ventricular contraction, FLWAV(31)–ventricular flutter wave, FUSION(6)–fusion of ventricular and normal beat. Details of annotations, labels and mapped labels are illustrated in Table I.

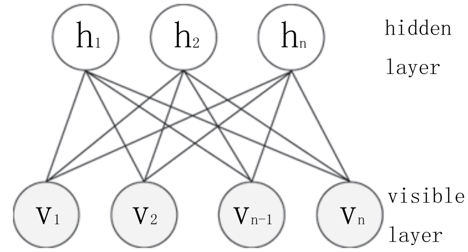

Fig. 1. A depiction of an two-layer RBM model, which included one layer with visible units and one layer with hidden units. Each visible unit is connected to the hidden units without visible-visible or hidden-hidden connections.

III. METHODOLOGY

In this section, the restricted Boltzmann machine theory, the model learning procedure, the deep belief network model, the softmax model and the optimization algorithm are proposed.

1) Restricted Boltzmann machine: Restricted Boltzmann Machine [19] is a stochastic neural network with strong unsupervised learning ability. Figure 1 depicted the RBM network structure. Each visible unit is connected to the hidden units without visible-visible or hidden-hidden connections. There were no connections between the visible and hidden layers. The visible units were independent, then the Gibbs sampling method could be used to approximate the probability distribution. It consists of one layer of visible units with input $X = (v_1, v_2, ..., v_n)$, one layer of hidden units with output $Y = (h_1, h_2, ..., h_m)$ and two bias units whose states were always on and a way to adjusting the value of each unit.

Boltzmann machine is based on statistical mechanics. The energy function $E(v, h)$ of an RBM was defined as:

$$E(v, h|\theta) = -\sum_{i=1}^{n} a_i v_i - \sum_{j=1}^{m} b_i h_i - \sum_{i=1}^{n}\sum_{j=1}^{m} v_i W_{ij} h_j \quad (1)$$

v and h present the state vectors of the visible and hidden layers.a_i, b_j and W_{ij} are parameters, $\theta = \{W_{ij}, a_i, b_j\}$. So based on the energy function, the distribution of v and h is:

$$P(v, h|\theta) = \frac{e^{-E(v,h|\theta)}}{Z(\theta)}, Z(\theta) = \sum_{v,h} e^{-E(v,h|\theta)} \quad (2)$$

The purpose of RBM is to learn the optimal θ. According to the probability distribution, the maximum likelihood function is defined as:

$$\theta^* = arg \max_{\theta} L(\theta) = arg \max_{\theta} \sum_{t=1}^{T} \log P(v^{(t)}|\theta) \quad (3)$$

$$L(\theta) = \sum_{t=1}^{T} \left(\log \sum_{h} exp[-E(v^{(t)}, h|\theta)] - \log \sum_{v} \sum_{h} [-E(v, h|\theta)] \right) \quad (4)$$

To get the optimal θ^*, stochastic gradient descent [21] method was used to maximum the likelihood function $L(\theta)$. The partial derivative of the parameters is shown below:

$$\begin{aligned} \frac{\partial \log P(v|\theta)}{\partial W_{ij}} &= \langle v_i h_i \rangle_{data} - \langle v_i h_i \rangle_{model}, \\ \frac{\partial \log P(v|\theta)}{\partial a_i} &= \langle v_i \rangle_{data} - \langle h_i \rangle_{model}, \\ \frac{\partial \log P(v|\theta)}{\partial b_j} &= \langle h_i \rangle_{data} - \langle h_i \rangle_{model}. \end{aligned} \quad (5)$$

$\langle . \rangle_P$ denotes the distribution about P. $\langle . \rangle_{data}$ is easy to be calculated when the training samples were defined. $\langle . \rangle_{model}$ could not be resolved directly, but approximated by Gibbs sampling. Here we use the contrastive divergence(CD) [30] algorithm proposed by Hinton in 2002, with which would achieve better results by only one step of Gibbs sampling .

2) Softmax classifier: This model generalized logistic regression [31] in classification missions which would be useful in heartbeats arrhythmia classification problems. The softmax model is a kind of supervised learning method in conjunction with the deep belief network.

Supposing m samples in the training set :

$$\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(m)}, y^{(m)})\} \quad (6)$$

the inputs were vectors $x^{(i)}$ corresponding to the features space. The labels are denoted by $y^{(i)}$ corresponding to the arrhythmia classes of the inputs. The cost function of softmax regression with a weight decay term was defined as:

$$J(\theta) = -\frac{1}{m}[\sum_{i=1}^{m} \sum_{j=1}^{k} 1\{y^{(i)} = j\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{k} e^{\theta_l^T x^{(i)}}}] + \frac{\lambda}{2} \sum_{i=1}^{k} \sum_{j=0}^{n} \theta_{jk}^2 (\lambda > 0) \quad (7)$$

and the partial derivative of the parameters were:

$$\triangledown_{\theta_j} J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} [x^{(i)}(1\{y^{(i)} = j\} - p(y^{(i)} = j|x^{(i)}; \theta))] + \lambda \theta_j (\lambda > 0) \quad (8)$$

To train an optimum classifier, an optimization gradient descent algorithm called L-BFGS was used.

3) Deep belief network: Figure 2 shows that RBMs can be stacked and trained with a greedy manner to form a deep belief network(DBN) [22], [26]. In the last layer, a softmax classifier is connected with the DBN. DBN is the graphical model of a hierarchical architecture. The five layers network shows in Figure2 are used to the heartbeat classification. The procedures were:

1) Train the first layer as an RBM which models the raw input X as a visible layer.
2) After training the RBM, representations of the input were obtained.
3) Train the next layer as an RBM which models the transformed data as a visible layer.
4) Iterate step 2 and 3 for the desired number of layers.

Finally the RBMs are combined to a DBN with the softmax model. Fine-tuning then was used as the supervised method to minimize the likelihood function and improve the adaptability. Here also the L-BFGS algorithm was used.

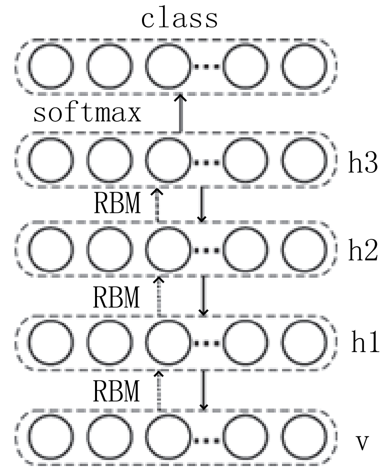

Fig. 2. The RBMs are stacked to from a deep belief network(DBN). The RBM can be trained layer by layer. It is easy to construct a DBN with the trained RBMs. Also, a softmax model to fine-tune all parameters behind the last layer

A. Combined optimization algorithm for multi-lead classifiers

The ECG wavelet transform performs differently in different channels by the waveforms. Each heartbeat channel shows diversely due to the P,QRS-complex and T wave constituent. Taking those into consideration, multi-lead ECG classification is significant improved by sample voting method. So a weight optimization method is proposed in two leads ECG signal classification. The method can be generally used in multi-lead ECG data classification.

For each classifier trained with distinct lead, a reliability value is denoted as the accurate rate of the classifier. let γ represent the reliability value. Then the classifiers' reliability is defined as $\gamma_1, \gamma_2, \cdots$

Using the testing samples to assess the result. The statistical matrix:

$$ClassSTST = \begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{12} & \cdots & C_{2n} \\ \ldots & \ldots & \ddots & \ldots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{pmatrix} \quad (9)$$

In the statistical matrix, there are n classes and C_{11} represents the class I which is classified as class I, C_{12} represents class I is classified as II, etc. The diagonal values are the correct classification. The purpose is to increase the diagonal values, so we adopt weights of the outputs for each classifier. The weights of the first classifier is $W_1 = (w_{11}, w_{12}, \cdots, w_{1n})$, the second were $W_2 = (w_{21}, w_{22}, \cdots, w_{2n})$. The constraint condition is $\sum_{i=1}^{2} w_{ik} = 1$.

First initial the weight to a mean value. Each sample has an output vector $O = o_1, o_2, \cdots, o_n$, the class is decided by the maximum value. Adding the weight of the value, the output is $(o_1 w_1, o_2 w_2, \cdots, o_n w_n)$. If the label of the sample is l, we would like to maximize $o_1 w_{11} + o_1 w_{21}$ while correspondingly minimize others. Through this, the l class accurate is promoted while the false negative rate is also increased. The optimization algorithm would find a balance between the two weights of all testing samples. Accordingly, the optimal function is defined as :

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} x \exp^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{10}$$

To find the maximum value of x, the derivative of the equation is:

$$F^{'}(x) = 0, x = \frac{\mu}{2} + \sqrt{\frac{\mu^2}{4} + \sigma^2} \tag{11}$$

Here, μ is denoted as initial weights. On the basis of the statistical matrix , counting the difference of the correct (diffcort) and false negative (difffln eg) quantities of the two classifiers. If the difference of the difference of the correct and false negative greater than zero, $\sigma^2 = \sqrt{\frac{diffcort - difffl sneg}{totalnumber}}$, so updating the corresponding class weight of the first classifier as $w_1 = \frac{\mu}{2} + \sqrt{\frac{\mu^2}{4} + \sigma^2}$. Else, $\sigma^2 = \sqrt{\frac{difffl sneg - diffcort}{totalnumber}}$, updating the weight of the second classifier as $w_2 = \frac{\mu}{2} + \sqrt{\frac{\mu^2}{4} + \sigma^2}$. Finally normalize the weights, the optimal combined value is $\gamma_1 o_1 w_1 + \gamma_2 o_2 w_2$. Generally, every two classifiers can be used to optimize the multi-lead ECG classification.

B. ECG preprocessing

The preprocessing works include two mainly parts, the ECG data filter and heartbeat segmentation process. The filter task is adapted to remove the artifact signal from the ECG signal. The artifact signal includes baseline wander, power line interference, and high-frequency noise. According to our pretraining works, the noise of the signal has little influence to the classifiers except for the baseline wander. So only baseline wander is removed. The massive unlabelled data we collected is extracted and a resample from 128Hz to 360Hz procedure is adopted for data consistency. In heartbeat segmentation process, average samples of each beat is 277 samples. To get more information , we allow partially overlap and a window with a length of 340 data points in one beat was defined, the R peak of the wave is located at 141th point. Most annotations of the MIT-BIH arrhythmia database is lied the R-wave. For the dataset we collect, a high accurate algorithm has been explored to determine the R pick and then divide into the heartbeat segments according to the R pick.

C. RBM training

The goal of RBM learning is to maximize the product of the probabilities. The parameters of the network can be initialized by the constructor. This option is useful when an RBM is used as the building block of the deep belief network, in which case the weight matrix and the hidden layer bias is shared with the corresponding layer of the network. The active function of the nodes is sigmoid function. The data is batched to train the RBM layer by layer. A single-step contrastive divergence(CD-1) is used in the gradient descent procedure. After calculating the partial derivative, the weights and bias is updated.

D. DBN fine-tuning

After the learning process, the RMBs can be used to initialize a deep belief network. Standard backpropagation algorithm can be applied to fine-tune the model. That can significantly improve the performance of the DBN. The fine-turning process is a supervised learning procedure, so at the last layer, a multi-class model called softmax is connected to classify the ECG data. Then using the fine-turning method to minimize the cost function.

E. Performance assessment

When finishing the fine-tuning process, the combining optimization algorithm was used to optimize the two-lead classifiers. The criterias employed to assess performance of beat classification approaches were sensitivity (TPR), specificity (SPC) and overall accuracy (ACC). Sensitivity is related to the correctly classified normal beats of normal beats

$$TPR = \frac{TP}{TP + FN} \tag{12}$$

where TP(true positive) is the number of the correctly classified normal beats, FN(false negative) is the number of mistakenly classified normal beats. Specificity is the correctly classified abnormal beats

$$SPC = \frac{FN}{FP + TN} \tag{13}$$

Where FP(false positive)is the number of the correctly classified abnormal beats, TN(true negative)is the number of the mistakenly classified corresponding abnormal beats Overall accuracy is the correctly classified beats

$$ACC = \frac{TP + TN}{TP + FP + NP + FN} \tag{14}$$

IV. Experimental results

In this section, the experimental results for the two leads ECG signal and combining algorithm are presented to evaluate the accuracy of the methods. The detail results are shown in Table II, III and IV.

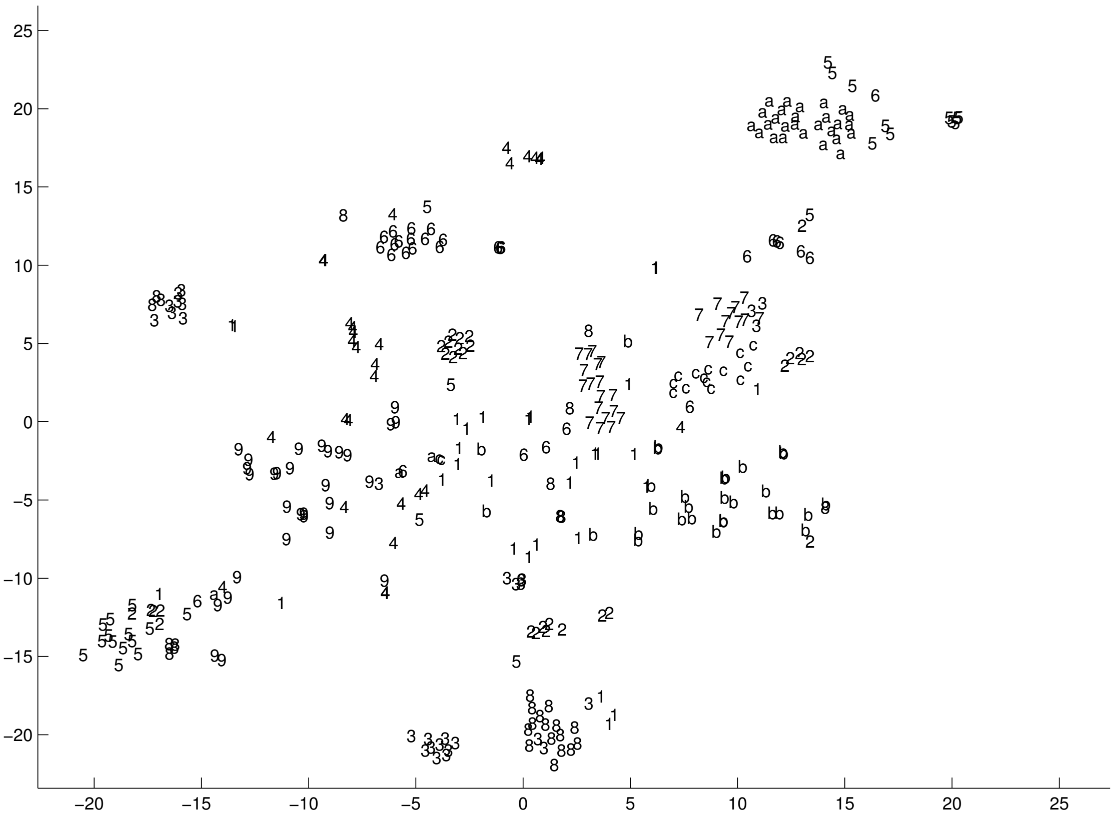

Fig. 3. The original heartbeats distribution visualizing with t-SNE [28].

A. The Segmentation of Heartbeat and Structure of DBN

Each of the 44 records is slightly over 30 minutes long and 650000 samples. According to our count, each beat has about 277 samples. To get more information ,We allow part of overlap and define a window with length of 340 data points(the R peak of the wave is located at 141th point). Considering the change of the heart rate, the window is adapted to avoid the mixture of two adjacent heart beats [1]. The segmentation of the heartbeat depends on the R-peak detection, so a high precision algorithm based on the local maximum is used to detect the R peaks. The Mexican-Hat wavelet [1] is also a choice to the R-peak detection.

The structure of the deep belief network we use is shown in Figure2. The input layer is a vector of the heartbeat with 340 points and the nodes of hidden layer 1 is 200, hidden layer 2 is 200 and hidden layer 3 is 100. The output layer is a softmax model. The original ECG data is used to classify the heartbeat without features extraction process. Figure 5 displays the error of the two channels using gradient descent algorithm when training the RBM. The t-SNE [28] algorithm is applied to visualize the distribution of heartbeat. 346 heartbeats are selected randomly from the dataset. It contains 30 heartbeats of each label from 1 to 11 and 16 heartbeats of all labeled 12 in the dataset. Figure 5 clearly displays the distribution of the primal heartbeats. Figure 3 shows the transformed features space. Character a represents label 10, b denotes label 11 and c indicates label 12. By comparing, the deep belief network is valid to heartbeats classification.

B. The classification performance

In the experiments, we adopted multi-lead ECG signal based on the restricted Boltzmann machine for the classification task. The MIT-BIH arrhythmia database is divided into three parts. Half of the beat is added to training the RBMs, one-third is applied to fine-tune the network and the left is used to test the model. In the 3 hidden layers deep belief network, the classifier outperforms in terms of sensitivity (SPR) 99.35%, specificity (SPC) 95.18% and accuracy rate (ACC) 98.25% using the first channels. Using the second channel we get the accuracy (ACC) of 97.43%, sensitivity (SPR) 98.90% and specificity (SPC) 93.36%. At convergence of the optmization process, the combining method achieves the accuracy (ACC) of 98.83%, sensitivity (SPR) 99.83% and specificity (SPC) 96.05%.

Table II, III and IV illustrates the detail statistics information

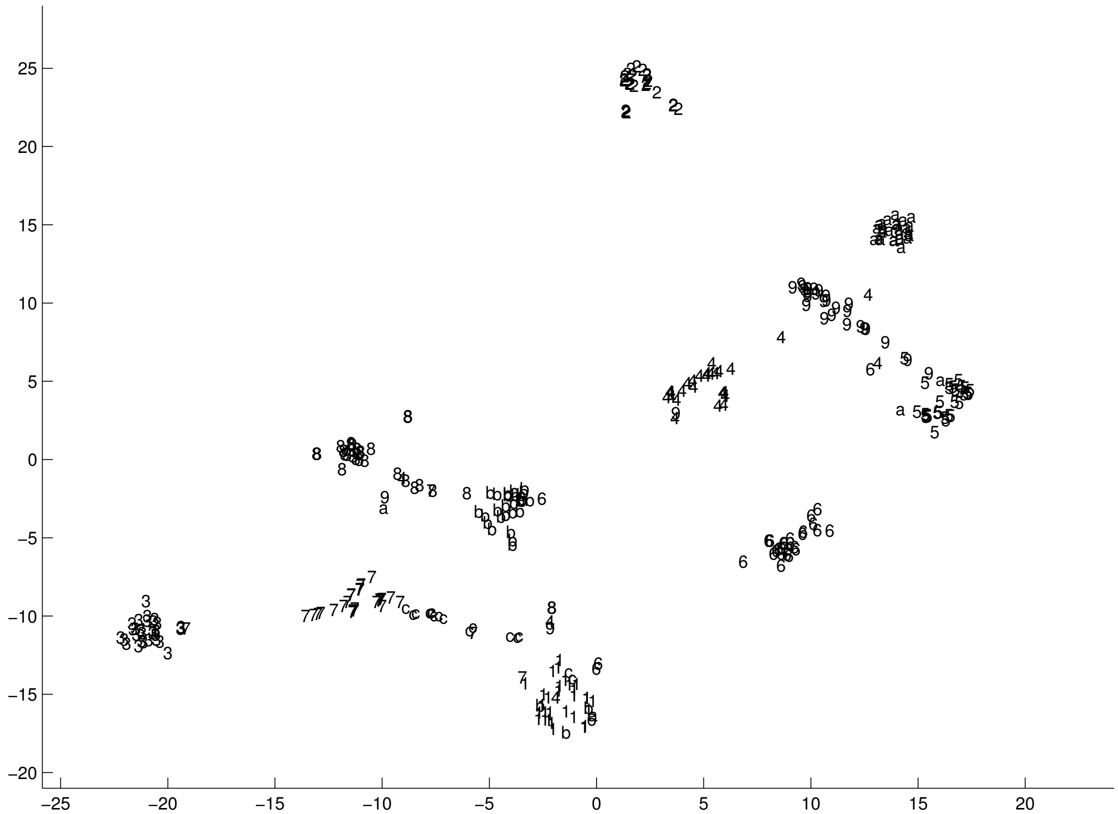

Fig. 4. The output layer's corresponding heartbeats distribution visualized with t-SNE [28].

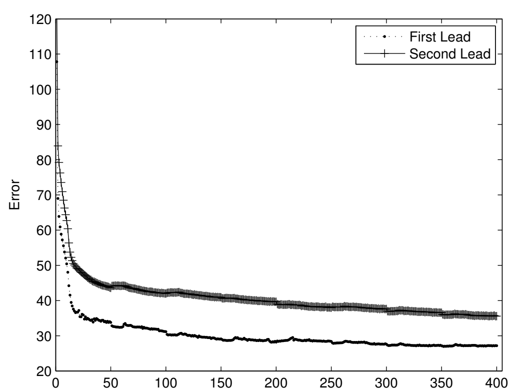

Fig. 5. The cost of the two leads using gradient descent algorithm [21] by RBM.

of the classifiers. The accurate rate is remarkably improved with the combining algorithm. It can be proved that multi-lead heartbeats classification is more reliable and higher in accuracy.

V. Discussion

One new aspect of this work is the new algorithm of combination in two ECG channels with deep belief network based on RBM. In previous study, most of the heartbeats classification used features extracted manually. In this research, the self learning method of heartbeat characteristics is illustrated. The detail performance is discussed below:

A. *Multilead fusion performance*

As the figures and tables in the experimental results showing, a better classification result had been achieved by the proposed method. Most of the accuracy of classes are improved by the combining algorithm in multi-lead signals.

For we use the limited labeled data from MIT database of only 51 persons, The network may be not fine-tuning to the best result state. As the classes statistics showed in Table II, III and IV, the normal beats is very large, other's beat is less, so the most incorrect heartbeats are the abnormal beats classified to the normal. Properly augmenting the proportion of the abnormal heartbeats will significantly improve the accurate rate.

TABLE II
TEST RESULT OF THREE HIDDEN LAYERS DEEP BELIEF NETWORK USING FIRST LEAD

		Algorithm classified label											
		NORMAL	LBBB	RBBB	ABERR	PVC	FUSION	NPC	APC	FLWAV	VESC	NESC	AESC
	NORMAL	12,289	3	2	3	15	9	0	37	2	0	0	0
	LBBB	7	1,369	0	0	6	0	0	1	0	0	0	0
	RBBB	4	0	1,179	0	3	0	0	4	0	0	0	0
	ABERR	7	1	0	12	5	0	0	0	1	0	0	0
	PVC	18	2	0	2	1,104	8	0	2	2	0	0	0
Original label	FUSION	19	0	0	1	6	97	0	0	0	0	2	0
	NPC	5	0	1	0	0	0	4	1	0	0	1	0
	APC	58	2	9	0	2	0	0	382	0	0	2	0
	FLWAV	10	0	0	1	8	0	0	0	58	0	0	0
	VESC	2	0	0	0	1	0	0	0	0	24	0	0
	NESC	6	0	1	0	0	0	0	1	0	0	15	0
	AESC	3	0	0	0	0	0	0	0	0	0	0	0

The test accuracy of the first lead is 98.247%.

TABLE III
TEST RESULT OF THREE HIDDEN LAYERS DEEP BELIEF NETWORK USING SECOND LEAD

		Algorithm classified label											
		NORMAL	LBBB	RBBB	ABERR	PVC	FUSION	NPC	APC	FLWAV	VESC	NESC	AESC
	NORMAL	12,233	7	3	3	57	13	1	34	12	0	6	0
	LBBB	12	1,366	0	0	5	0	0	0	0	0	0	0
	RBBB	5	0	1,169	0	5	0	0	9	2	0	0	0
	ABERR	7	0	1	10	6	0	1	0	1	0	0	0
	PVC	56	3	1	0	1,048	15	0	4	11	0	0	0
Original label	FUSION	22	0	0	0	5	97	0	0	0	0	1	0
	NPC	1	0	1	0	0	0	10	0	0	0	0	0
	APC	60	4	8	0	11	2	0	368	1	0	0	1
	FLWAV	7	0	0	0	8	0	0	1	61	0	0	0
	VESC	2	0	1	0	2	0	0	0	2	20	0	0
	NESC	6	0	1	0	1	0	0	0	0	0	15	0
	AESC	2	0	0	0	0	0	0	0	0	0	0	1

The test accuracy of the second is 97.433%.

TABLE IV
TEST RESULT OF COMBINATION OPTIMAL ALGORITHM WITH TWO LEADS

		Algorithm classified label											
		NORMAL	LBBB	RBBB	ABERR	PVC	FUSION	NPC	APC	FLWAV	VESC	NESC	AESC
	NORMAL	12,348	0	0	0	11	2	0	7	1	0	0	0
	LBBB	3	1,377	0	0	3	0	0	0	0	0	0	0
	RBBB	1	0	1,182	0	2	0	0	5	0	0	0	0
	ABERR	8	0	0	16	2	0	0	0	0	0	0	0
	PVC	16	0	0	0	1,108	10	0	1	3	0	0	0
Original label	FUSION	22	0	0	0	4	99	0	0	0	0	0	0
	NPC	3	0	0	0	0	0	9	0	0	0	0	0
	APC	58	2	5	0	1	0	0	389	0	0	2	0
	FLWAV	2	0	0	0	6	0	0	0	69	0	0	0
	VESC	3	0	0	0	1	0	0	0	0	23	0	0
	NESC	11	0	1	0	0	0	0	1	0	0	11	0
	AESC	3	0	0	0	0	0	0	0	0	0	0	0

The test accuracy of the combining leads is 98.829%.

B. The influence of filtering

The artifact signal includes baseline wander, power line interference, and high-frequency noise. In this experiment, only the baseline wander is removed. According to our training progress, the power line interference and high-frequency noise has little effect on the deep belief network. From the testing result that we can state that the baseline removal filter affects the heartbeat classification in our proposed approach. The results confirm that the deep network works well on data with noisy. We adopt the same network structure and train all the classifiers with only baseline wander removed in the preprocessing progress. So the key difference between our work and others' is that we use simplex preprocessing and original ECG signal without complex features extraction.

C. Compare with Others' works

For we collect large amount of ECG data from the hospital which contains lots of normal beats and abnormal beats, the learning method of restricted Boltzmann machine is used to learning the features from the massive data by an unsupervised way. Then the RBMs is adapted to build a deep belief network. The optimization algorithm we propose improves the accuracy with multilead ECG signal. The heartbeat classification also has been studied by other algorithm, including hidden Markov model(HMM) [9], support vector machine(SVM) [10] [11], independent component analysis(ICA) [12], Artificial neural networks(ANN) that use varied features extracted from the discreet signal data. Different performance assessment criteria has been adopted. In this comparison, we evaluate the performance on the indicators which put forward in the above sections: Sensitivity(TPR), specificity(SPC) and overall Accuracy(ACC). The $*NR$ symbol represents the result is not reported.

TABLE V
COMPARISONS WITH OTHERS' WORKS

Approaches	Accuracy(ACC)	Sensitivity(TPR)	Specificity(SPC)
Proposed	98.83%	99.83%	96.05%
Tadejko [2]	97.82%	99.70%	93.10%
Banerjee [4]	97.60%	97.30%	98.80%
Can [5]	99.71%	*NR	*NR
Osowski [15]	96.06%	98.10%	95.53%

[a] *NR means the results were not reported.
[b] The listed percentages are based on the assessment rules.

The Table V shows the performance of different models that used for the heartbeat classification, the proposed method offers a high accuracy of classification.

All the annotations in the MIT-BIH arrhythmia database is use in our study and Osowski [15] only selects 7 types. Can [5] get the highest accuracy but with a price of rejecting 2054 heartbeats(2.4% rejections). By comparing with others' experience, our approach provided higher performance in heartbeat classification without complex wavelet transform algorithms.

VI. CONCLUSIONS

This study shows that how a restricted Boltzmann machine works with large scale of ECG data. Training a deep belief network is time-consuming job, but it works fast to predict. Using the MIT-BIH Arrhythmia Database [29] and massive data we collected from the hospital, the batch progressing method is applied to speed up the training progress. At last, a fine-tuning method is used to adapt the whole network. From the obtained experimental results, it can be strongly recommended that the deep belief network is suitable to ECG classification. We test 16,831 samples randomly chosen from MIT-BIH and provide them with high accuracy of 98.247% from the first channel ECG signal and 97.433% from the second channel. By learning features from massive data the model can get a more accurate result. Another advantage of the multi-lead RBMs training progress can be found in its parallel. The training of different channel does not affect each other. So this method enhance the classification accuracy and does not increase the training time.

An optimization method is proposed for multi-lead heartbeat classification. The system shows better result with this method. Also the method is acting on the output of the deep belief network. The combined result is 98.829% in two channels which is better than with only one channel. For different channel shows diverse degrees of the components, the result confirmed that optimal algorithm substantially boosts the accuracy of the heartbeat classification with multi-lead. From the filtering process, the deep belief network is well robust with the original data has noise.

ACKNOWLEDGMENT

This study was financed partially by the National 863 Program of China (Grant No. 2012AA02A604), the Next generation communication technology Major project of National S&T (Grant No. 2013ZX03005013), the Key Research Program of the Chinese Academy of Sciences, and the Guangdong Innovation Research Team Funds for Image-Guided Therapy and Low-cost Healthcare. The Service Standardization of Medical Institutions in the Village and Town (Grant No. 201230119), funded by XinJiang Department of Science and Technology. This study was financed partially by the National Natural Science Foundation of China (Grant No.61403366) Shenzhen Basic Research Project Fund(JCYJ20140417113430695).

REFERENCES

[1] M. Burke and M. Nasor, "Ecg analysis using the mexican-hat wavelet," in *Int. Conf. Multirate Systems & Wavelet Analysis*. Citeseer, 2012, pp. 1–6.

[2] P. Tadejko and W. Rakowski, "Hybrid wavelet-mathematical morphology feature extraction for heartbeat classification," in *EUROCON, 2007. The International Conference on &# 34; Computer as a Tool&# 34;*. IEEE, 2007, pp. 127–132.

[3] S. Palreddy, W. J. Tompkins, and Y. H. Hu, "Customization of ecg beat classifiers developed using som and lvq," in *Engineering in Medicine and Biology Society, 1995., IEEE 17th Annual Conference*, vol. 1. IEEE, 1995, pp. 813–814.

[4] S. Banerjee and M. Mitra, "Application of cross wavelet transform for ecg pattern analysis and classification," *Instrumentation and Measurement, IEEE Transactions on*, vol. 63, no. 2, pp. 326–333, 2014.

[5] C. Ye, B. V. Kumar, and M. T. Coimbra, "Heartbeat classification using morphological and dynamic features of ecg signals," *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 10, pp. 2930–2941, 2012.

[6] T. Mar, S. Zaunseder, J. P. Martinez, M. Llamedo, and R. Poll, "Optimization of ecg classification by means of feature selection," *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 8, pp. 2168–2177, 2011.

[7] P. De Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ecg morphology and heartbeat interval features," *Biomedical Engineering, IEEE Transactions on*, vol. 51, no. 7, pp. 1196–1206, 2004.

[8] A. E. Zadeh, A. Khazaee, and V. Ranaee, "Classification of the electrocardiogram signals using supervised classifiers and efficient features," *computer methods and programs in biomedicine*, vol. 99, no. 2, pp. 179–194, 2010.

[9] R. V. Andreão, B. Dorizzi, and J. Boudy, "Ecg signal analysis through hidden markov models," *Biomedical Engineering, IEEE Transactions on*, vol. 53, no. 8, pp. 1541–1549, 2006.

[10] B. M. Asl, S. K. Setarehdan, and M. Mohebbi, "Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal," *Artificial intelligence in medicine*, vol. 44, no. 1, pp. 51–64, 2008.

[11] F. Melgani and Y. Bazi, "Classification of electrocardiogram signals with support vector machines and particle swarm optimization," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 12, no. 5, pp. 667–677, 2008.

[12] S.-N. Yu and K.-T. Chou, "Integration of independent component analysis and neural networks for ecg beat classification," *Expert Systems with Applications*, vol. 34, no. 4, pp. 2841–2846, 2008.

[13] Y. H. Hu, W. J. Tompkins, J. L. Urrusti, V. X. Afonso *et al.*, "Applications of artificial neural networks for ecg signal detection and classification," *Journal of electrocardiology*, vol. 26, pp. 66–73, 1993.

[14] G. K. Prasad and J. Sahambi, "Classification of ecg arrhythmias using multi-resolution analysis and neural networks," in *TENCON 2003. Conference on Convergent Technologies for the Asia-Pacific Region*, vol. 1. IEEE, 2003, pp. 227–231.

[15] S. Osowski and T. H. Linh, "Ecg beat recognition using fuzzy hybrid neural network," *Biomedical Engineering, IEEE Transactions on*, vol. 48, no. 11, pp. 1265–1271, 2001.

[16] S. Karpagachelvi, M. Arthanari, and M. Sivakumar, "Classification of ecg signals using extreme learning machine," *computer and information science*, vol. 4, no. 1, p. p42, 2011.

[17] M. Moavenian and H. Khorrami, "A qualitative comparison of artificial neural networks and support vector machines in ecg arrhythmias classification," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3088–3093, 2010.

[18] V. Chudacek, M. Petrík, G. Georgoulas, M. Cepek, L. Lhotská, and C. Stylios, "Comparison of seven approaches for holter ecg clustering and classification," in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*. IEEE, 2007, pp. 3844–3847.

[19] G. H. Ruslan Salakhutdinov, "Deep boltzmann machines," *The Journal of Machine Learning Research*, vol. 5, pp. 448–455, 2009.

[20] G. E. Hinton, "A practical guide to training restricted boltzmann machines," in *Neural Networks: Tricks of the Trade*. Springer, 2012, pp. 599–619.

[21] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'2010*. Springer, 2010, pp. 177–186.

[22] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[23] J. Martens and I. Sutskever, "Training deep and recurrent networks with hessian-free optimization," in *Neural Networks: Tricks of the Trade*. Springer, 2012, pp. 479–535.

[24] X. Chen and X. Lin, "Big data deep learning: Challenges and perspectives," 2014.

[25] Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle *et al.*, "Greedy layer-wise training of deep networks," *Advances in neural information processing systems*, vol. 19, p. 153, 2007.

[26] Y. Bengio, "Learning deep architectures for ai," *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[27] I. Basheer and M. Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application," *Journal of Microbiological Methods*, vol. 43, no. 1, pp. 3 – 31, 2000, neural Computting in Micrbiology. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167701200002013

[28] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, no. 2579-2605, p. 85, 2008.

[29] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000 (June 13).

[30] G. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural computation*, vol. 14, no. 8, pp. 1771–1800, 2002.

[31] C.-Y. J. Peng, K. L. Lee, and G. M. Ingersoll, "An introduction to logistic regression analysis and reporting," *The Journal of Educational Research*, vol. 96, no. 1, pp. 3–14, 2002.