Assessment of Electrocardiograms with Pretraining and Shallow Networks

Vicent J. Ribas Ripoll[1], Anna Wojdel [2], Pablo Ramos [3], Enrique Romero [4], Josep Brugada [3]
[1] Centre de Recerca Matemàtica, Barcelona, Spain
[2] Custom Software and Electronics, Barcelona, Spain
[3] Hospital Clínic, UB, Barcelona, Spain
[4] Departament Llenguatges i Sistemes Informàtics, UPC, Barcelona, Spain

Abstract

Objective: Clinical Decision Support Systems normally resort to annotated signals for the automatic assessment of ECG signals. In this paper we put forward a new method for the assessment of normal/abnormal heart function from raw ECG signals (i.e. signals without annotation) based on shallow neural networks with pretraining.

Methodology: this paper resorts to a prospective clinical study that took place at Hospital Clínic in Barcelona, Spain. This study took place in 2010-2012 and recruited 1390 patients. For each patient we recorded a 12-lead ECG and diagnosis was conducted by the Cardiology service at the same hospital. Two datasets were produced, the first contained the automatically annotated version of all input signals and the second contained the raw signals obtained from the ECG.

Results: The new method was tested through cross-validation with a cohort of 200 test patients. Performance was compared for both annotated and raw datasets. For the annotated dataset and a shallow network with pretraining we obtained an accuracy of 0.8639, a sensitivity of 0.9560 and specificity of 0.7143. The raw dataset yielded an accuracy of 0.8426, a sensitivity of 0.8977 and a specificity of 0.7785.

Conclusion: Shallow networks with pretraining automatically obtain a representation of the input data without resorting to any annotation and thus simplify the process of assessing normality of ECG signals. Despite the fact that sensitivity has decreased, accuracy is not much lower than that obtained with standard methods. Specificity is improved with the new method. These results open up a promising line of research for the automatic assessment of ECG signals.

1. Introduction

Cardiovascular diseases (CVD) remain the main cause of death worldwide. According to the most recent statistics of the World Health Organisation mortality rates are expected to range between 246 deaths for 100,000 population in 2015 to 264 for 100,000 population in 2030 [1]. The early detection of CVD through ambulatory care services may significantly decrease the admission time in a hospital cardiology service, and consequently, mortality rates for citizens.

In this regard, electrocardiography plays a very important role in the assessment of cardiac function. However, this assessment can be challenging and diagnosis almost always require confirmation from an expert cardiologist. Moreover, most of the decision support systems used today resort to annotated versions of ECG signals for screening different diseases. In this paper we put forward a new method for the assessment of normal/abnormal heart function from raw ECG signals (i.e. signals without annotation) based on *shallow* neural networks with *pretraining*.

This short paper is organised as follows: section 2 gives an overview of neural networks, restricted Boltzmann machines and the model that we used. In section 3, we present the datasets used in our experiments and show the experimental prediction results for each different dataset (i.e. annotated vs. raw). In chapter 4 we present the conclusions of this short study.

2. Methods

2.1. Neural Networks

Neural networks comprise a variety of well-known and widely used machine learning techniques for a number of tasks including classification or regression, among others [2]. A neural network is a set of simple interconnected processing units (neurons). Every unit is able to perform simple computations (activation function), taking as inputs the outputs of other units and the strengths of the connections (weights). The knowledge of the network is stored in its weights.

A neural network is trained to optimize a certain cost function which depends on the weights and the data, such as the cross-entropy or the squared error. Usually, the cost function is differentiable and the training process uses the derivatives with respect to the weights. The

most widely used algorithm to compute these derivatives is Back-propagation [3].

Typically, the units of a neural network are structured in layers: the input layer, one or more hidden layers and the output layer. Usually, two adjacent layers are fully connected, and there are no connections between non-adjacent layers. The computation starts in the input layer, which propagates the data through the hidden layers to the output layer. Every unit transforms the information that receives according to its weights.

Until recently, most successful neural networks were shallow, i.e., they had one or two hidden layers. In fact, deeper architectures were thought to be too difficult to train, and their empirical results were found to be similar, and often worse, than the results obtained by shallow networks [4]. In the last years, however, deep networks have emerged as a powerful model, outperforming shallow networks in different problems and studies [5].

An important issue in the success of deep networks is the initialization of weights. While shallow networks may obtain good results with random initializations, deep networks need more complex procedures to initialize these weights. The training of a deep network consists of two steps: an unsupervised pre-training and a supervised fine-tuning [6]. The pre-training step is used to find a good set of initial weights. The fine-tuning is equivalent to the standard training procedure of a neural network.

One of the most common approaches for the pre-training step (and the first breakthrough in deep learning) is described in [7,8] and is based on an unsupervised generative model called Restricted Boltzmann Machines (RBMs) [9]. Other pre-training algorithms are based on auto-encoders [4,10]. In the following we will describe some basic issues of RBMs.

2.2. Restricted Boltzmann Machines

RBMs are energy-based probabilistic models whose energy function is:

$$\text{Energy}(\vec{x},\vec{h}) = -\vec{b}^t\vec{x} - \vec{c}^t\vec{h} - \vec{h}^t\vec{W}\vec{x}\,. \qquad (1)$$

In energy-based probabilistic models, a probability distribution is defined from an energy function:

$$P(\vec{x},\vec{h}) = \frac{e^{-\text{Energy}(\vec{x},\vec{h})}}{Z}\,, \qquad (2)$$

where $\vec{x}$ are variables and $\vec{h}$ are hidden variables introduced to increase the expressive power of the model. The normalization factor Z is called the partition function:

$$Z = \sum_{\vec{x},\vec{h}} e^{-\text{Energy}(\vec{x},\vec{h})}\,. \qquad (3)$$

The consequence of that particular form of the energy function in (1) is that in RBMs both $P(\vec{h}|\vec{x})$ and $P(\vec{x}|\vec{h})$ factorize [11]. In this way it is possible to compute $P(\vec{h}|\vec{x})$ and $P(\vec{x}|\vec{h})$ in one step, making it possible to perform Gibbs sampling efficiently [12]. This is the basis of the computation of an approximation of the derivative of the log-likelihood, whose exact computation is computationally intractable. The most common learning algorithm for RBMs is called Contrastive Divergence (CD) [13].

3. Results

3.1. Dataset Description

The dataset used in our study consists of 12-lead digital ECGs recorded in resting position. All ECGs were obtained from patients admitted to the cardiology service of Hospital Clínic in Barcelona during years 2011-2012.

All subjects included in our study were selected at random from all cardiology patients with no patient-specific screening. No selection was done regarding age, gender, or base pathology. The only criteria used in our ECG selection was related to its readability in terms of noise levels in the recorded sample. This selection was performed by cardiologists from the same service. Our study included ECGs that were at least 10 seconds long.

In total, our dataset consists of 12-lead ECG signals from 1390 subjects (749 male and 641 female). The age of the subjects ranged between 1 month and 94 years old with a median of 63 years old and inter-quantile range (IQR) 23 years. Table 1 shows the different diagnosis in our dataset and the number of cases for each different condition.

Table 1. Number of ECG diagnoses belonging to each group making occurrence in our dataset (it is possible to have multiple diagnoses for each ECG)

diagnosis group	no.
sinus: arrhythmia, bradycardia, tachycardia	209
atrial: ectopic rhythm, tachycardia, premature complexes	36
atrial flutter and atrial fibrillation	120
ventricular escape or premature complexes	63
AV conduction abnormalities	89
intra-ventricular conduction disturbances	249
P wave abnormalities	56
QRS low voltage, QRS axis deviation	48
ventricular hypertrophy	31
myocardial infraction	126
ST-, T-, and U-wave abnormalities	235
pacemaker rhythm and malfunction	44
other (e.g. Brugada syndrome)	15

Before recruitment, each ECG recording was carefully

diagnosed and described by cardiologists from the hospital. The description provided by cardiologists contained information about all possible abnormalities visible on the ECG recording accompanied by the final diagnosis. In order to simplify and further automatise the processing of the recordings, the diagnosis was provided in two forms: as textual description and in the form of predefined list of codes. We chose deploy the codes from the list of ECG diagnosis used at Mayo Clinic, as presented in [14]. The list of all diagnostic codes which we detail in our database contains 61 different items. We show a summary of this list by dividing the diagnoses into several groups. Table 1 shows number of recordings with ECG diagnostic codes assigned to each of those groups. The most common codes for atypical patterns (each with more than 100 occurrences) in our dataset are: sinus bradycardia, atrial fibrillation, and complete right bundle branch block (RBBB).

It is important to note that most of the diagnoses contain more than one diagnostic code. This applies not only to ECGs with their patterns classified as abnormal, but also to the ECGs considered as normal. The most common diagnoses that co-occur in normal ECG are: sinus arrhythmias, sinus bradycardia and tachycardia, atrial or ventricular premature complexes, and some ST-, and T-wave abnormalities like early repolarization or non-specific abnormalities. Classification of each ECG to *normal* or *abnormal* group was based on the final diagnosis provided by the cardiologists. All ECGs with diagnostic codes containing 1A (Normal ECG) or 1B (Borderline normal ECG or normal variant) were classified as *normal* and the rest of ECGs as *abnormal*. In total, our dataset contained 772 *normal* and 618 *abnormal* recordings, with 18% of *normal* ECGs additionally marked with some other diagnostic code(s).

3.1.1. Raw dataset

The digital ECG recordings that we obtained from the hospital presented different lengths due to different sampling frequencies (500Hz or 1000Hz). ECG formats have been unified by taking the first 10 seconds of each recording and down-sampling all 1000 Hz ECGs to 500 Hz with the Matlab function decimate. Baseline was further unified by applying a low-pass Butterworth filter with a cut-off frequency of 150 Hz to all signals. It is this homogeneous input dataset that we have used throughout our study.

3.1.2. Annotated dataset

Our annotated dataset consists of the set of 150 parameters belonging to three groups as described below. First group, the smallest one, consists of **demographic parameters** such as age and gender of the patient. The second group, **global measurements**, contains information about the ECG that is universal to all leads. The examples of such information are: average duration and axis of QRS complex, P-, and T-wave, average duration of RR, PR and QTc intervals, Sokolov index. Finally, the last group, **local measurements**, comprises parameters representing significant features of the ECG signal in specific leads. This group includes parameters like: codes representing morphology of QRS complex, P-, and T-wave, and specific amplitudes (of P-, T-waves, decomposed QRS complex, J point, ST segment) in every single lead.

All of the parameters above were obtained by processing the *raw dataset* with Hannover ECG System HES® software library.

3.2. ECG Assessment

In our experiments, we used shallow neural networks with unsupervised pre-training based on RBMs to obtain the initial set of weights. RBMs with Gaussian inputs and 200 hidden units were trained for 100 epochs with CD-1. In the fine-tuning step, the network was trained 50 epochs with a procedure based on the non-linear Conjugate Gradients method. The data were scaled to zero mean and variance one. The data set was split (randomly with balanced classes) into a training set with 973 subjects (70% of data), a validation set with 208 patients (15% of data) and a test set with another 208 patients (15% of the data). This procedure was repeated 10 times. For this dataset, we observed that deep networks obtained similar results to shallow networks. Therefore, we decided to select the most simple models. It is worth saying that both shallow and deep networks benefited from the pre-training procedure.

In our experiments we compared the performance of our classifier between the annotated and raw datasets. For the annotated dataset, the shallow network yielded an accuracy of 0.8639, a sensitivity of 0.9560 and specificity of 0.7143. The raw dataset yielded an accuracy of 0.8426, a sensitivity of 0.8977 and a specificity of 0.7785.

The results show that our method, starting from scratch (i.e. raw ECG signals) yields similar results to state-of-the-art hand-engineered features.

As we have described above, the weights learned by a network store the knowledge that it has learned about a system/model. In figure 1, we show the weights associated to one hidden unit in the network pretrained with the raw data. As it can be seen we have achieved ECG sensitive neurons capable of discriminating pathological traits in the ECG signal. For sake of comparison, we also present in figure 2 the same weights of the neural network removing the pretraining. It is apparent that the 'learned' traits are much noisier than those that we have obtained with pretraining.

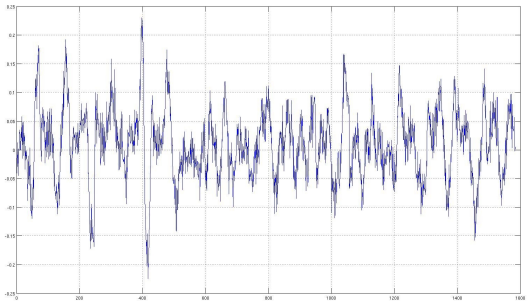

Figure 1. Shallow network weights with pretraining, the patterns obtained can be matched to ECG patterns.

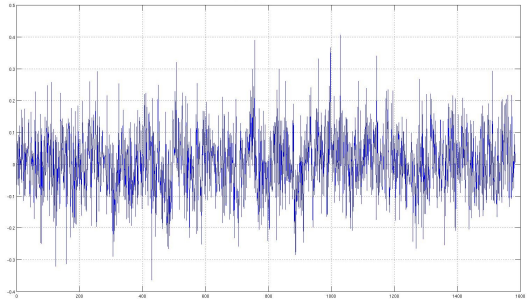

Figure 2. Shallow network weights with no pretraining.

4. Conclusion

Shallow networks with pretraining automatically obtain a representation of the input data without resorting to any annotation and thus simplify the process of assessing normality of ECG signals. Despite the fact that sensitivity has decreased, accuracy is not much lower than that obtained with standard methods. Specificity is improved with the new method. These results open up a promising line of research for the automatic assessment of ECG signals.

As we have shown in figure 1, the shallow network has learned an ECG representation with physiologic relevance. From this results obtained we hypothesize that the proposed method with a raw signal may outperform the accuracy obtained with the annotated dataset. This study in a larger dataset with a much richer pathology base and a higher number of cases for each pathology is proposed as further work.

Acknowledgements

This project has been partially funded by the Shockomics project, under the EU FP7 framework.

References

[1] Global health observatory data repository. `http://www.who.int`.

[2] Haykin S. Neural Networks. A Comprehensive Foundation (2nd edition). Prentice Hall International, 1999.

[3] Rumelhart DE, Hinton GE, Williams RJ. Learning Representations by Back-propagating Errors. Nature 1986; 323(9):533–536.

[4] Bengio Y, Lamblin P, Popovici D, Larochelle H. Greedy Layer-wise Training of Deep Networks. In Advances in Neural Information Processing Systems, volume 19. MIT Press, 2007; 153–160.

[5] Bengio Y. Deep Learning of Representations: Looking Forward. In Statistical Language and Speech Processing. Springer, 2013; 1–137.

[6] Erhan D, Bengio Y, Courville A, Manzagol PA, Vincent P. Why Does Unsupervised Pre-training Help Deep Learning? Journal of Machine Learning Research 2010;11:625–660.

[7] Hinton GE, Osindero S, Teh Y. A Fast Learning Algorithm for Deep Belief Nets. Neural Computation 2006; 18(7):1527–1554.

[8] Hinton GE, Salakhutdinov RR. Reducing the Dimensionality of Data with Neural Networks. Science 2006; 313(5786):504–507.

[9] Smolensky P. Chapter 6: Information Processing in Dynamical Systems: Foundations of Harmony Theory. In Rumelhart DE, McClelland JL (eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition (vol. 1). MIT Press, 1986; 194–281.

[10] Vincent P, Larochelle H, Bengio Y, Manzagol PA. Extracting and Composing Robust Features with Denoising Autoencoders. In 25th International Conference on Machine Learning. 2008; 1096–1103.

[11] Bengio Y. Learning Deep Architectures for AI. Foundations and Trends in Machine Learning 2009;2(1):1–127.

[12] Geman S, Geman D. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. IEEE Transactions on Pattern Analysis and Machine Intelligence 1984; 6:721–741.

[13] Hinton GE. Training Products of Experts by Minimizing Contrastive Divergence. Neural Computation 2002; 14:1771–1800.

[14] Murphy JG, Lloyd MA. Mayo Clinic Cardiology: Concise Textbook. Taylor & Francis, 2006. ISBN 9780849390579.

Address for correspondence:

Vicent J. Ribas Ripoll

Campus de Bellaterra, Edifici C, 08193 Bellaterra

vribas@crm.cat