Online Neural Network Training for Automatic Ischemia Episode Detection

D.K. Tasoulis[1,⋆], L. Vladutu[2], V.P. Plagianakos[3,⋆], A. Bezerianos[2], and M.N. Vrahatis[1,⋆]

[1] Department of Mathematics and University of Patras Artificial Intelligence Research Center (UPAIRC), University of Patras, GR-26110 Patras, Greece, {dtas,vrahatis}@math.upatras.gr

[2] Department of Medical Physics, University of Patras, GR-26500 Patras, Greece, bezer@patreas.upatras.gr, LiviuVladutu@ieee.org

[3] Department of Information and Communication Systems Engineering, University of the Aegean, GR-83200 Samos, Greece, vpp@aegean.gr

Abstract. Myocardial ischemia is caused by a lack of oxygen and nutrients to the contractile cells and may lead to myocardial infarction with its severe consequence of heart failure and arrhythmia. An electrocardiogram (ECG) represents a recording of changes occurring in the electrical potentials between different sites on the skin as a result of the cardiac activity. Since the ECG is recorded easily and non–invasively, it becomes very important to provide means of reliable ischemia detection. Ischemic changes of the ECG frequently affect the entire repolarization wave shape. In this paper we propose a new classification methodology that draws from the disciplines of clustering and artificial neural networks, and apply it to the problem of myocardial ischemia detection. The results obtained are promising.

Keywords: Online training, ischemia episode detection.

1 Introduction

Myocardial ischemia is the most common cause of death in the industrialized countries and, as a consequence, its early diagnosis and treatment is of great importance. Myocardial ischemia diagnosis using long–duration electrocardiographic recordings is a simple and non–invasive method that needs further development before being used in everyday medical practice. The capability of accurate and early detection of an acute ischemic event is critical for the persuadation of the proper treatment. The Electrocardiogram (ECG) represents a recording of the changes occurring in the electrical potentials between different sites on the skin as a result of the cardiac activity. Since the ECG is recorded

⋆ This work was supported in part by the "Karatheodoris" research grant awarded by the Research Committee of the University of Patras, and the "Pythagoras" research grant awarded by the Greek Ministry of Education and the European Union.

L. Rutkowski et al. (Eds.): ICAISC 2004, LNAI 3070, pp. 1062–1068, 2004.

easily and noninvasively, it becomes very important to provide means for reliable ischemia detection from ECG analysis.

There are a few mandatory steps for automated detection of ischemic episodes. After the initial removal of noise it follows the second stage, when all the important ECG features (J point, isoelectric line, and T wave peak) are extracted. Using the above features, in the third stage each cardiac beat is classified as normal or ischemic. In the final stage, sequential ischemic beats are grouped properly and the ischemic episodes can be identified.

The ST–T Complex of the ECG represents the time period from the end of the ventricular depolarization to the end of the corresponding repolarization in the electrical cardiac cycle. Ischemic changes of the ECG frequently affect the entire repolarization wave shape and thus are inadequately described by isolated features, even if these are obtained as an average of several signal samples [1]. Additionally, in many cases the ST segment is sloped or is influenced by noise. The approach proposed at the current work avoids the utilization of local, isolated features by designing upon the Principal Component Analysis (PCA) technique for extracting PCA coefficients (features) that describe the global content of the ST–T Complex.

2 Preprocessing

A description of the European ST-T Database is provided in [2], explaining the rules for the localization of the ST and T episodes. The main aim of the ECG signal preprocessing is to prepare a compact description of the ST–T complex, composed from the ST Segment and the T–wave, for input to the classification methodology with the minimum loss of information. From the samples composing each beat, a window of 400 msec is selected (100 samples at the 250 Hz sampling frequency). This signal component will form the input to the PCA to describe most of its content within a few (i.e. five) coefficients. To have a reference for the extraction of the relevant segment, the position of the R–peak should be detected. The start of the ST–T Segment was selected at approximately 60 msec after the detected R peak. However, in the database, there are both patients with bradycardia and tachycardia. Therefore, a more flexible approach that accounts for heart rate variations is required. The selection of the distance between the S point and the previously detected R–peak is correlated with the heart rhythm of the patient. The distance between the R–peak and the J point is in the range of 45–80 msec. Due to the fact that the correction of the ST–T length using the Bazett's formula, yields to a similar PCA basis function set, an analysis approach is selected with a fixed time window of 400 msec. This assumption is valid for the set of first 5 Principal Components (PCs) we used for representation. During the ST–T segment extraction, we rejected a small number (less than 1%) of ST–T segments, considered as particularly noisy.

The PCA method transforms a set of correlated random variables of dimensionality m, to a set of $d \leqslant m$ uncorrelated (in terms of their second order statistics) variables according to the direction of maximum variance reduction in the training set. The uncorrelated variables correspond to the subspace decomposition based on the first principal components of the input data. This decomposition is in terms of second order statistics optimum, in the sense that it permits an optimal reconstruction of the original data in the mean–square error sense. In our case, PCA has performed well for the extraction of representative vectors with only five coefficients. Thus, at the particular dimensionality reduction problem there is not sufficient evidence that the successive samples of the ECG signal are correlated in complex nonlinear ways. The ST–T Segment can be reconstructed effectively with the first five PCA projections that represent about 98.1% of the total signal energy. The PCA projection coefficients are then fed to the Feedforward Neural Network (FNN) to perform the classification decision about the category pertaining to each analysis case (i.e. normal, abnormal, artifact). The first PC and the second one (but to a less extent) represent the dominant low frequency component of the ST–T segment; the third, fourth and fifth contain more high frequency energy.

Following the extraction of principal components a noise reduction approach is used to improve these coefficients. The utilization of an advanced wavelet denoising technique has improved the classification results. The selected noise reduction approach was based on soft thresholding [3]. We have chosen five levels of wavelet decomposition and Daubechies–type wavelets.

3 The Classification Methodology

In this paper instead of constructing a global model for the pattern classification, we construct several local models, for neighborhoods of the state space. For this task, we use the novel k-windows clustering algorithm [4], to automatically detect neighborhoods in the state space. This algorithm, with a slight modification (unsupervised k-windows algorithm) has the ability to endogenously determine the number of clusters present in the dataset during the clustering process. Once the clustering process is complete, a trained FNN acts as the local predictor for each cluster. In synopsis, the proposed methodology consists of the following four steps:

1. Identify the clusters present in the training set.
2. For each cluster, train a different FNN using for training patterns, patterns from this cluster solely.
3. Assign the patterns of the test set to the clusters according to their distance from the center of the cluster.
4. Use the trained FNNs to obtain the classification scores on the test set.

The unsupervised k-windows algorithm generalizes the original algorithm [4]. Intuitively, the k-windows algorithm tries to place a d-dimensional window (box) containing all patterns that belong to a single cluster; for all clusters present in the dataset. At first, k points are selected (possibly in a random manner). The k initial d-ranges (windows), of size a, have as centers these points. Subsequently, the patterns that lie within each d-range are identified. Next, the mean of the

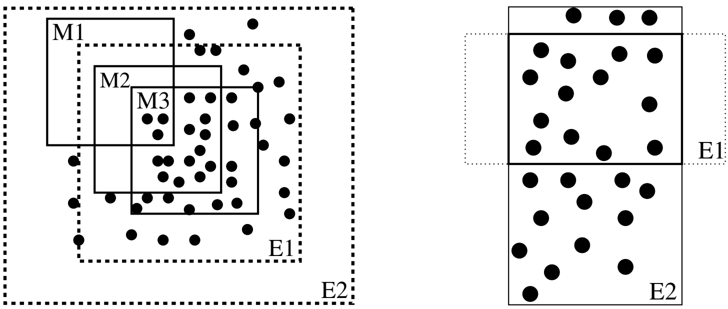

Fig. 1. Left: Movements (solid lines) and enlargements (dashed lines). Right: The enlargement process

patterns that lie within each d-range is calculated. The new position of the d-range is such that its center coincides with the previously computed mean value. The last two steps are repeatedly executed as long as the increase in the number of patterns included in the d-range that results from this motion satisfies a stopping criterion. The stopping criterion is determined by a variability threshold θ_v that corresponds to the least change in the center of a d-range that is acceptable to recenter the d-range (Figure 1, left).

Once movement is terminated, the d-ranges are enlarged in order to capture as many patterns as possible from the cluster. Enlargement takes place at each dimension separately. The d-ranges are enlarged by θ_e/l percent at each dimension, where θ_e is user defined, and l stands for the number of previous successful enlargements. After the enlargement in one dimension is performed, the window is moved, as described above. Once movement terminates, the proportional increase in the number of patterns included in the window is calculated. If this proportion does not exceed the user–defined coverage threshold, θ_c, the enlargement and movement steps are rejected and the position and size of the d-range are reverted to their prior to enlargement values. Otherwise, the new size and position are accepted. If enlargement is accepted for dimension $d' \geqslant 2$, then for all dimensions d'', such that $d'' < d'$, the enlargement process is performed again assuming as initial position the current position of the window.

This process terminates if enlargement in any dimension does not result in a proportional increase in the number of patterns included in the window beyond the threshold θ_c (Figure 1, right). In the figure the window is initially enlarged horizontally ($E1$). This enlargement is rejected since it does not produce an increase in the number of patterns included. Next the window is enlarged vertically, this enlargement is accepted, and the result of the subsequent movements and enlargements is the initial window to become $E2$. The key idea to automatically determine the number of clusters, is to apply the k-windows algorithm using a sufficiently large number of initial windows. The windowing technique of the k-windows algorithm allows for a large number of initial windows to be examined, without any significant overhead in time complexity. Once all the processes of movement and enlargement for all windows terminate, all overlapping windows are considered for merging. The merge operation is

Table 1. Percentages of correct classification on the test sets over 100 iterations

FNN Classification Performance						
Test set	E103	E104	E106	E107	E108	E111
			RPROP			
mean	75.60	86.63	**59.78**	79.07	73.95	92.19
std	0.39	0.96	2.20	2.48	1.13	2.28
max	76.21	88.87	66.04	85.76	75.85	94.55
min	74.76	84.34	57.58	73.67	71.57	87.51
			iRPROP			
mean	75.64	87.07	59.87	**78.48**	73.91	**91.84**
std	0.54	0.73	2.88	1.62	1.10	4.00
max	76.53	88.29	69.17	81.71	75.75	94.92
min	74.36	84.57	57.48	75.24	71.97	80.41
			SCG			
mean	77.46	85.77	62.51	79.07	73.57	92.45
std	0.77	0.80	4.74	2.13	1.61	1.49
max	79.05	87.09	71.15	84.38	76.16	94.77
min	75.76	83.64	58.54	72.61	69.62	89.82
			BPVS			
mean	**72.50**	84.29	67.94	83.28	70.00	93.91
std	0.02	0.45	0.17	0.07	0.007	0.03
max	72.57	85.27	68.08	83.34	70.01	93.99
min	72.49	84.06	67.72	83.17	69.99	93.86
			AOBP			
mean	73.69	**83.83**	71.68	83.10	**69.67**	93.21
std	0.59	0.09	0.00	0.01	0.005	0.30
max	73.88	83.91	71.68	83.14	69.67	93.86
min	71.93	83.72	71.68	83.10	69.64	93.05

guided by a merge threshold θ_m. Having identified two overlapping windows, the number of patterns that lie in their intersection is calculated. Next the proportion of this number to the total patterns included in each window is calculated. If the mean of these two proportions exceeds θ_m, then the windows are considered to belong to a single cluster and are merged.

4 Numerical Results

Numerical experiments were performed using a Clustering, and a Neural Network, C++ Interface built under the Red Hat Linux 7.3 operating system using the GNU compiler collection (gcc) version 3.2. The efficient supervised training of FNNs is a subject of considerable ongoing research and numerous algorithms have been proposed to this end. In this work, we consider the following neural network training methods:

– Resilient Back Propagation (RPROP),
– Improved Resilient Back Propagation (iRPROP) [5],
– Scaled Conjugate Gradient (SCG),
– Adaptive On–Line Back Propagation (AOBP) [6],
– Back Propagation with Variable Stepsize(BPVS) [7],

After extensive experimentation the network architecture selected consisted of 8 nodes in the first hidden layer, 7 nodes in the second hidden layer, and two output nodes (5–8–7–2). All FNNs were trained for 300 epochs on the patterns of the training set and subsequently their performance was evaluated on the test sets. This process was repeated 100 times for all the training algorithms considered. The classification capability of the trained FNNs with respect to the accurate pattern classification in the test sets are reported in Table 1.

In the datasets E103, E104, and E108, FNNs trained with RPROP and iRPROP outperformed all other methods. The drawback of these two methods is the relatively high standard deviation. SCG also suffers from the same drawback. The performance of RPROP and iRPROP on dataset E106 is discouraging. For the remaining datasets FNNs trained with BPVS and AOBP, produced the best results. A significant advantage of AOBP, and to an extent BPVS, is the fact that the standard deviation of their performance is negligible. Overall, for the datasets E104, E107 and E111, the classification ability of the proposed methodology is very good.

5 Conclusions

This paper presents a methodology for automatic recognition of ischemic episodes, which draws from the disciplines of clustering and artificial neural networks. The methodology consists of four stages. To effectively partition the state space, the training patterns are subjected to clustering through the unsupervised k-windows algorithm. Subsequently, a different FNN is trained on each cluster. At the third stage, the patterns in the test set are assigned to the clusters identified in the training set. Finally, the trained FNNs are used to classify each test pattern. This methodology was applied to classify several test cases of the European ST–T database and the obtained results were promising.

References

1. Maglaveras, N., Stamkopoulos, T., Pappas, C., Strintzis, M.: An adaptive backpropagation neural network for real-time ischemia episodes detection. IEEE Transactions on Biomedical Engineering **45** (1998) 805–813
2. Vladutu, L., Bezerianos, A., Papadimitriou, S.: Hierarchical state space partitioning with a network self-organizing map for the recognition of $st - t$ segment changes. Medical & Biological Engineering & Computing **38** (2000) 406–415
3. Donoho, D., Johnstone, I.: Ideal spatial adaptation by wavelet shrinkage. Biometrika **81** (1994) 425–455
4. Vrahatis, M., Boutsinas, B., Alevizos, P., Pavlides, G.: The new k-windows algorithm for improving the k-means clustering algorithm. Journal of Complexity **18** (2002) 375–391
5. Igel, C., Hüsken, M.: Improving the Rprop learning algorithm. In Bothe, H., Rojas, R., eds.: Proceedings of the Second International ICSC Symposium on Neural Computation (NC 2000), ICSC Academic Press (2000) 115–121

6. Magoulas, G., Plagianakos, V., Vrahatis, M.: Adaptive stepsize algorithms for on-line training of neural networks. Nonlinear Analysis, T.M.A. **47** (2001) 3425–3430
7. Magoulas, G., Vrahatis, M., Androulakis, G.: Effective backpropagation training with variable stepsize. Neural Networks **10** (1997) 69–82