

A MULTI-LEAD ECG CLASSIFICATION BASED ON RANDOM PROJECTION FEATURES

Iva Bogdanova*, Francisco Rincón*[†] and David Atienza**

*Embedded Systems Laboratory (ESL), Ecole Polytechnique Fédérale de Lausanne (EPFL)

[†]Department of Computer Architecture and Automation, Complutense University of Madrid (UCM)

email: {iva.bogdanova; francisco.rincon; david.atienza}@epfl.ch

ABSTRACT

This paper presents a novel method for classification of multi-lead electrocardiogram (ECG) signals. The feature extraction is based on the random projection (RP) concept for dimensionality reduction. Furthermore, the classification is performed by a neuro-fuzzy classifier. Such a model can be easily implemented on portable systems for practical applications in both health monitoring and diagnostic purposes. Moreover, the RP implementation on portable systems is very challenging featuring both energy efficiency and feasibility. The proposed method is tested on a 12-lead ECG database consisting of 20 beats during normal sinus rhythm, 20 beats with myocardial infarction and 20 beats showing cardiomyopathy for 60 different subjects. The experiments give a recognition rate of 100% for a small number of RP coefficients (only 25), i.e. after a considerable dimensionality reduction of the input ECG signal. The results are very promising, not only from the classification performance point of view, but also while targeting a low-complexity feature extraction in terms of computation requirements and memory usage for real-time operation on a wireless wearable sensor platform.

Index Terms— random projections, adaptive neuro-fuzzy classification, automatic multi-lead ECG, ECG dimensionality reduction,

1. MOTIVATION

Many approaches involving techniques for automatic classification of multi lead ECG signals have been proposed recently. Some of them are based on morphological analysis. Other approaches use statistical methods. Further, neural methods become a powerful tool for ECG classification. An overview of ECG feature extraction techniques has been given in [1]. Among these techniques, we can point wavelet decomposition [2], Discrete Cosine Transform (DCT) or Principal Component Analysis (PCA) [3].

*This work has been supported by Spanish Government grant TIN2008-00508, and partially supported by the Swiss National Science Foundation (SNF) grant 200021-127282.

Many classification methods have been studied and experimented on ECG signals but the neuro-fuzzy classifier shows particular advantages [4, 3]. The concept of adaptive networks and a supervised learning procedure based on a gradient descent algorithm to update the parameters has been introduced in [5]. The herein used neuro-fuzzy system harness the power of two paradigms: fuzzy logic and neural networks, by utilizing the mathematical properties of neural networks to tune rule-based fuzzy systems that approximate the way a human processes the information [6].

Different works that bring together feature extraction and ECG classification can be found in the literature. A model for ECG classification based on a fuzzy-gaussian neural network was presented in [3]. It incorporates feature extraction using PCA and a neuro-fuzzy classifier to obtain a recognition rate of 100% for ischemic heart diagnosis. Another interesting ECG classification model is presented in [2]. In that work, the authors use wavelet transforms to extract the features and further classify them, but their experiments are only conducted on 2-lead signals. Finally, a multi-lead ECG classification based on independent component analysis and support vector machines has been presented in [7].

In order to consider a real-time implementation of ECG classification on portable devices for monitoring, a simple and reliable methods for dimensionality reduction of the input signals are needed. This is required due to the limited resources of portable devices architectures. The random projection (RP) method fits this requirement, since it involves very low computational complexity. Moreover, it is highly important to consider multi-lead ECG signals, since different symptoms may appear on different leads. Therefore, feature combination allows to keep the characteristics of each of the leads in the classification procedure. This paper proposes a novel method for ECG feature extraction using RPs, to reduce the dimensionality of each of the 12 considered leads. Then, the extracted features are combined using a root-mean-squared (RMS) approach. Finally, the resultant signal is classified using a neuro-fuzzy approach. The experimental results show a recognition rate of 100% with a dimensionality reduction from 700 to only 25 coefficients. In addition, this system can be implemented on a wireless wearable sensor platform, due to its low complexity.

The rest of the paper is organized as follows. In Section 2, the RP concept is introduced. Then, we continue by presenting the neuro-fuzzy classifier in Section 3. The experimental results are shown in Section 4. Finally, in Section 5 we conclude and give some future work directions.

2. RANDOM PROJECTIONS

The use of RP as a tool for dimensionality reduction dates from [8], where the foundation of the dimensionality reduction was laid and its feasibility was proofed. They showed that N points in N dimensional space can almost always be projected onto a space of dimension $C \log N$ while controlling the ratio of distances and the distortion. Later, it was shown that any metric space with N points can be embedded by a bi-Lipschitz map into an Euclidean space of $\log N$ dimensions with a bi-Lipschitz constraint of $\log N$. Consequently, RPs reduce the dimensionality of a data-set while preserving its geometrical structure. Achlioptas formulated in [9] a special type of RPs, namely matrices with entries of ± 1 in random location in a matrix.

In this paper, we consider maps from $\mathbb{R}^d \mapsto \mathbb{R}^k$, expressed as $k \times d$ binary matrices, as given in [9]. Normally, in applications $k \ll d$. Such matrices are denoted by \mathbf{P} and defined by

$$\mathbf{P} = \sqrt{3} \begin{cases} +1, & \text{with probability } \frac{1}{6} \\ -1, & \text{with probability } \frac{1}{6} \\ 0, & \text{with probability } \frac{2}{3} \end{cases} \quad (1)$$

Achlioptas uses matrices of this type to prove in [9]:

Theorem 2.1 *Let \mathcal{S} be an arbitrary set of $S = |\mathcal{S}|$ points in \mathbb{R}^d . Given $\epsilon, \beta > 0$, let*

$$k_\beta = \frac{4 + 2\beta}{\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3}} \log(S) \quad (2)$$

for any integer $k \geq k_\beta$, let \mathbf{P} be a random matrix as described in Equation 1. Then, with probability at least $1 - S^{-\beta}$, $\forall u, v \in \mathcal{S}$

$$(1 - \epsilon) \|u - v\|^2 \leq \|\mathbf{P}u - \mathbf{P}v\|^2 \leq (1 + \epsilon) \|u - v\|^2 \quad (3)$$

It is important to note that Theorem 2.1 is independent of the original dimension d .

3. NEURO-FUZZY CLASSIFICATION

Fuzzy classification is the task of partitioning a feature space into fuzzy classes. A method of employing adaptive networks to solve a fuzzy classification problem is worked out in [5]. An adaptive network is a multi-layer feed-forward network in which each node performs a particular function based on incoming signals and a set of parameter pertaining to this

node. Of particular interest is the adaptive neuro-fuzzy classifier with linguistic hedges [10]. It is based on fuzzy rules. A fuzzy classification rule that has two inputs $\{x_1, x_2\}$ and one output y is defined with linguistic hedges as:

If x_1 is A_1 with p_1 hedge and x_2 is A_2 with p_2 hedge THEN y is C_1 class, where A_1 and A_2 denote linguistic terms that are defined on X_1 and X_2 feature space; p_1 and p_2 denote linguistic hedges, respectively; C_1 denotes the class label of output y .

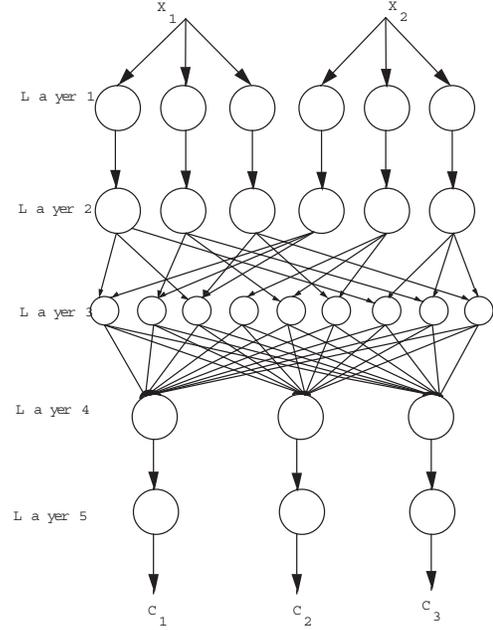


Fig. 1. Architecture of the neuro-fuzzy classifier with linguistic hedges. The feature space with two inputs $\{x_1, x_2\}$ is partitioned into three classes $\{C_1, C_2, C_3\}$

The architecture of the used adaptive neuro-fuzzy classifier is illustrated in Figure 1. Here, the nodes in the same layer have the same type of node functions. The layers are defined as follows:

- **Layer 1: Membership layer.** A Gaussian membership function is used in this paper. It is defined as follows

$$\mu_{ij}(x_{sj}) = \exp\left(\frac{-(x_{sj} - c_{ij})^2}{2\sigma_{ij}^2}\right), \quad (4)$$

where $\mu_{ij}(x_{sj})$ is the membership grade of the i -th rule and the j -th feature; x_{sj} denotes the s -th sample and the j -th feature of input matrix $\mathbf{X}\{\mathbf{X} \in \mathbb{R}^{N \times D}\}$; c_{ij} and σ_{ij} are the center and the width of Gaussian function, respectively.

- **Layer 2: Power layer.** The secondary meanings of fuzzy sets are calculated with their linguistic hedges $\alpha_{ijs} = [\mu_{ij}(x_{sj})]^{p_{ij}}$, where α_{ijs} denotes the modified membership grades of $\mu_{ij}(x_{sj})$; p_{ij} denotes the

linguistic hedges value of the i -th rule and the j -th feature.

- Layer 3: **Fuzzification layer.** The β_{is} firing strength of the i -th rule is $\beta_{is} = \prod_{j=1}^D \alpha_{ijs}$, where D denotes the number of features.
- Layer 4: **Defuzzification layer.** Here weighted outputs are calculated, and every rule can affect each class according to their weights. However, if a rule controls a specific class region, the weight between this rule output and the specific class is to be bigger than the other class weights. The weighted output for the s -th sample that belongs to the k -th class is computed: $o_{sk} = \sum_{i=1}^U \beta_{is} w_{ik}$, where U is the number of rules, w_{ik} represents the degree of belonging to the k -th class that is controlled with the i -th rule.
- Layer 5: **Normalization layer.** The outputs of the network should be normalized because sometimes the summation of weights may be bigger than 1. The normalized degree of the s -th sample that belongs to the k -th class is computed as follows: $h_{sk} = \frac{o_{sk}}{\sum_{l=1}^K o_{sl}} = \frac{o_{sk}}{\delta_s}$, with $\delta_s = \sum_{l=1}^K o_{sl}$, where K is the number of classes. Then, the class label of the s -th sample is determined by $C_s = \max_{k=1,2,\dots,k} \{h_{sk}\}$

In this paper, the scale conjugate gradient (SCG) method is used to adapt the network parameters. This is a second order supervised training and derivative-based method. In particular, it determines the second order derivatives of parameters from their first-order derivatives. The SCG has a super linear convergence rate which is two times faster than that of the back-propagation algorithm.

4. EXPERIMENTAL RESULTS

4.1. ECG database

For experimenting the proposed paradigm of RP-based neuro-fuzzy classification, we have used the Physikalisch-Technische Bundesanstalt (PTB) Diagnostic ECG database available on <http://physionet.org/>. Each record includes 15 simultaneously measured signals: the conventional 12 leads (i, ii, iii, avr, avl, avf, v1, v2, v3, v4, v5, v6) together with the 3 Frank lead ECGs (vx, vy, vz). Each signal is digitized at 1000 samples per second, with 16-bit resolution over a range of $\pm 16.384mV$. There are several diagnostic classes represented in the database.

For a purpose of experiment, we create a restricted version of this database where only 60 subjects have been considered: 20 during normal sinus rhythm, 20 with myocardial infarction and 20 presenting cardiomyopathy. Our database is divided into the training and test sets, each with 30 subjects of the three classes (i.e. 10 normal, 10 myocardial infarction,

and 10 cardiomyopathy). We consider that the significant information is concentrated only in the PQRST zone of the 12 leads. First, an ECG pre-processing step has consisted of extracting the useful information from a record, namely extracting a characteristic beat. The selected beat contains $n = 700$ samples.

4.2. Feature vector

We have used the RP concept for generating the feature vector for each of the records in database. The random matrices were generated based on Section 2 and using a Gaussian distribution. An example of a normal beat and its randomly projected version is shown in Figure 2. The beat is with reduced dimensionality from 700 to 50 samples.

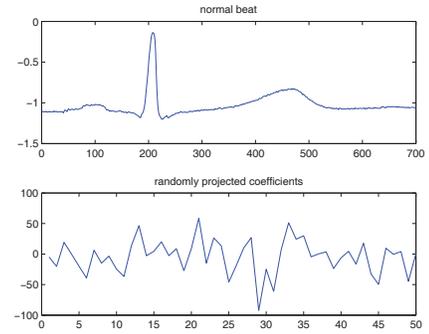


Fig. 2. 1-lead ECG beat during normal sinus rhythm (above) and its random projection using 50 coefficients (below)

To integrate the information from the available channels, the RMS of the randomly projected coefficients in the $M = 12$ leads is computed:

$$R_{rms_n} = \sqrt{\frac{1}{M} \sum_{i=1}^M R_n^2}. \quad (5)$$

4.3. Random projection-based ECG classification

After extracting the RP, as described in Section 2, on the entire database, the neuro-fuzzy classifier from Section 3 is further applied. Initially, each feature dimension is partitioned into three homogeneously distributed overlapping regions. The constructed network is trained with the training data set for 100 epochs. Different experiments on the randomly projected database are performed, changing the number of coefficients in the range from 5 to 50.

4.4. Results analysis

The performance of the classifier was evaluated by computing the Jaccard index: $J = \frac{TP}{m} \times 100, \%$ with $m = TP +$

Table 1. Recognition rate of the adaptive neuro-fuzzy classifier as a function of retained coefficients

number of coefficients	recognition rate training set (%)	recognition rate test set (%)
5	95	45
10	95	85
15	95	95
20	100	90
25	100	100
50	100	100

$FN + FP$. The recognition rate in the sense of Jaccard index is given in Table 1 for different number of retained RP coefficients.

These first results are very promising even though they concern a restricted data set. We must stress here that the data set has not been analyzed with statistical methods, a better examination of the membership function could further improve the results while leading to smaller number of retained RP coefficients. Examination of the average recognition rate over several executions of the proposed paradigm would be needed when a bigger dataset is considered.

4.5. Computational complexity of the feature extraction technique

One of the main advantages of the proposed RP-based feature extraction technique is its low computational complexity, which makes it able to operate in real-time on a wireless wearable sensor platform. These platforms have very limited processing, memory storage and energy resources. As an example, the commercial Shimmer™ platform [11] is equipped with an ultra-low-power 16-bit microcontroller that runs at a maximum clock frequency of 8MHz and includes 10KB of RAM. If we want to implement PCA, and taking into account a dimensionality reduction of the ECG signals from 700 to 25 to obtain 100% recognition rate (Figure 1), we need a matrix of 17500 elements of 2 bytes (35KB of memory just to store the coefficients matrix). However, with RP we only have 3 possible values, therefore we can represent each element of the matrix with only 2 bits, which reduces the required memory for the random matrix to only 4.4KB. The probability of having 0's in the random matrix is also very high (2/3), thus, 2/3 of the multiplications can be replaced by a simple comparison.

5. CONCLUSIONS AND FUTURE WORK

In this paper we have presented an ECG classification model based on RP feature extraction and adaptive neuro-fuzzy classification. The experimental results are run on an ECG database containing three different classes (normal sinus rhythm, myocardial infarction and cardiomyopathy). This

method shows promising results leading to 100% recognition rate for only 25 retained RP coefficients. Furthermore, the RP-based feature extraction technique has a low computational complexity, which makes it able to operate in real-time on a wireless wearable sensor platform. As a future work, an implementation of the proposed model for ECG classification on one of these platforms is previewed.

6. REFERENCES

- [1] P. de Chazal and Reilly R. B., "A comparizon of the ECG classification performance of different feature sets," in *Computers in Cardiology*, 2000, pp. 327–330.
- [2] M. L. Soria and Martinez J. P., "An ECG classification model based on multilead wavelet transform features," in *Computers in Cardiology*, 2007, pp. 105–108.
- [3] V.E Neagoe, I-F. Iatan, and Grunwald S., "A neuro-fuzzy approach to classification of ECG signals for ischemic heart disease diagnostic," in *AMIA Symposium Proceedings*, 2003, pp. 494–498.
- [4] R. Benali and Chikh M.A., "Neuro-fuzzy classifier for cardiac arrythmias recognition," in *J. Theoretical and Applied Information Technology*, 2009, pp. 577–583.
- [5] Ch.T. Sun and Jang J.S., "A neuro-fuzzy classifier and its applications," in *Contemporary Mathematics*, 1993, vol. 26, pp. 189–206.
- [6] I. Guler and E. D. Ubeyli, "Application of adaptive neuro fuzzy inference system for detection of electrocardiographic changes in patients with partial epilepsy using feature extraction," in *Expert Systems with Application*, 2004, vol. 27, pp. 323–330.
- [7] B. M. Shen et al, "Multi-lead ECG classification based on independent component analysis and support vector machine," in *Int. Conf. on Biomedical Engineering and Informatics (BMEI)*, 2010, vol. 37, pp. 960–964.
- [8] W.B. Johnson and J. Lindenstrauss, "Extensions of Lipschitz mapping into Hilbert space," in *Contemporary Mathematics*, 1984, vol. 26, pp. 189–206.
- [9] D. Achlioptas, "Database-friendly random projections: Johson-Lindenstrauss with binary coins," in *Journal of Computer and System Sciences*, 2003, vol. 66, pp. 671–687.
- [10] B. Cetisli, "Development of adaptive neuro-fuzzy classifier using linguistic hedges: Part 1," in *Expert Systems with Applications*, 2010, vol. 37, pp. 6093–6101.
- [11] A. Burns et al, "SHIMMER a wireless sensor platform for noninvasive biomedical research," *IEEE Sensors*, vol. 10, no. 9, pp. 1527–1534, Sept. 2010.