35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Unsupervised Learning applied in MER and ECG Signals through Gaussians Mixtures with the Expectation-Maximization Algorithm and Variational Bayesian Inference

Hernán Darío Vargas Cardona, Álvaro Ángel Orozco and Mauricio A. Álvarez

Abstract— **Automatic identification of biosignals is one of the more studied fields in biomedical engineering. In this paper, we present an approach for the unsupervised recognition of biomedical signals: Microelectrode Recordings (MER) and Electrocardiography signals (ECG). The unsupervised learning is based in classic and bayesian estimation theory. We employ gaussian mixtures models with two estimation methods. The first is derived from the frequentist estimation theory, known as Expectation-Maximization (EM) algorithm. The second is obtained from bayesian probabilistic estimation and it is called variational inference. In this framework, both methods are used for parameters estimation of Gaussian mixtures. The mixtures models are used for unsupervised pattern classification, through the responsibility matrix. The algorithms are applied in two real databases acquired in Parkinson's disease surgeries and electrocardiograms. The results show an accuracy over 85% in MER and 90% in ECG for identification of two classes. These results are statistically equal or even better than parametric (Naive Bayes) and nonparametric classifiers (K-nearest neighbor).**

I. Introduction

The clinical need of the biosignal analysis, arises from the fact that diseases and dysfunctions in biological processes may be detected with digital signal processing and pattern recognition. For example, classification of electrocardiograms (ECG) into different disease categories is a difficult pattern recognition task. At this scenario, the processing of electrocardiogram signals is an effective method for diagnosing cardiac arrhythmias. Classification algorithms applied in ECG's can provide high accuracy results [1]. A typical heart beat is identified from the ECG and the component waves of the QRS, T, and P waves are characterized using measurements such as magnitude, duration and area. Excellent results are achieved extracting shape features of the ECG that discriminate between the different diagnostic categories.

Also, the interpretation of physiological signals known as Microelectrode Recording signals (MER-signals) is crucial for Deep Brain stimulation (DBS) of the Subthalamic Nucleus (STN) in Parkinson's Disease patients [2]. MER signals have a non-stationary behavior due to the contribution of several biochemical factors [3]. For this reason is necessary the development of robust methodologies for processing and classification of these signals. This procedure serves as medical support for the correct location of a target brain area and the respective implantation of microelectrodes.

H.D Vargas Cardona, A.A. Orozco and M.A. Álvarez are with the Department of Electrical Engineering, Universidad Tecnológica de Pereira, Pereira-Colombia, {hernan.vargas, malvarez, aaog}@utp.edu.co

Supervised classification algorithms, usually operate with the information provided by a set of samples, a set of patterns, or training examples with a correct class label. This data set is called training set. When, there is not a previous knowledge about the labels, it is needed a data analysis known as unsupervised learning. Unsupervised learning studies how systems can learn to represent particular input patterns in a way that reflects the statistical structure of the overall collection of input patterns. In contrast with supervised learning there are no explicit target outputs or labels associated with each input [4]. The goal is to build classifiers without prior information. Unsupervised learning consists of dividing the set of samples in groups of similar objects called clusters. In this context the cluster represents a data class.

Clustering algorithms have been used in speech recognition, image segmentation and computer vision [5], data mining, information retrieval and text mining [6], heterogeneous data analysis [7], web applications [8], computational biology and DNA analysis [9].

In this paper we use Gaussian mixture models with the Expectation-Maximization (EM) algorithm proposed by [10], and variational inference algorithm. The responsibility matrices: $\gamma(\mathbf{z})$ in EM algorithm and $\mathbf{r}$ in variational inference determine the class of each sample. The algorithms were tested on two different real databases: 1. Brain signals from microelectrode recording (MER) and, 2. Signals of electrocardiography (ECG). The raw MER-signals were processed using adaptive wavelets with two decomposition levels. Discrete wavelet transform with five decomposition levels was used for ECG signals. Processed databases (MER and ECG) have eight and fifteen features per sample respectively. Both databases have two classes. The results show an accuracy over 85% in MER and 90% in ECG for identification of two classes. This results are statistically equal or even better than some supervised classifiers.

II. Materials and Methods

A. Databases

The MER database of the Technological University of Pereira (DB-UTP) includes recordings of surgical procedures in patients with Parkinson's disease, whose ages are between 55 ± 6 (5 men, 2 woman). All the patients signed an informed consent form. Microelectrode recordings were obtained using the ISIS MER system (Inomed Medical GmbH). MER signals were labeled by neurophysiology and neurosurgery

978-1-4577-0216-7/13/$26.00 ©2013 IEEE

specialists from the Institute of Parkinson and Epilepsy of the Eje Cafetero, located in the city of Pereira, Colombia. In total, there are 700 neural recordings divided in two classes: 350 signals from Subthalamic Nucleus (STN), and 350 from other brain structures (Thalamus-Tal, Zone Incerta-ZI, Substantia Nigra reticulata-SNr). Each record lasted 1 second with sampling frequency of 25 kHz and 16-bit of resolution.

The ECG Database belongs to the research group Gamma ascribed to engineering faculty of the University of Quindío, located in Armenia, Colombia. This database consists of records processed with wavelet transform using 5 decomposition levels. For each record we extract 15 features. There are 1180 samples of normal beats and 1180 samples of the pathology: Left Bundle Branch Block (LBBB).

B. Feature Extraction

For MER signals we use adaptive wavelets with two decomposition levels. We apply the following metrics in approximation coefficients $\mathbf{C}_p$ $(p = 1,2)$:

$$M_p = max\left(\left|\mathbf{C}_p\right|\right), \textit{Maximum Value}$$

$$E_p = \frac{1}{2}\sum_N (\mathbf{C}_p)^2, \textit{Energy}$$

$$\mu_p = \frac{1}{N}\sum_N \left|\mathbf{C}_p\right|, \textit{Normalized Mean}$$

$$K_p = Kurtosis\left(\left|\mathbf{C}_p\right|\right), \textit{Kurtosis}$$

Finally, we obtain a feature vector $\mathbf{x} \in \mathbb{R}^{1X8}$ for each sample.

ECG signals were processed via discrete wavelet with 5 decomposition levels. We use statistical descriptors in approximation coefficients $\mathbf{C}_p$ $(p = 1,2,3,4,5)$. Then, we apply Principal Component Analysis (PCA) for obtain 15 features per sample.

C. EM Algorithm for Gaussians Mixtures

Given a Gaussian mixture model, the objective is to maximize the likelihood function respect to the parameters (means, covariances and mixing coefficients). The EM algorithm for Gaussian mixtures can be summarized in the following steps [11] :

1) Initialize the means μ_k, covariances Σ_k, mixing coefficients π_k, the convergence criterion, and evaluate the initial value of the log-likelihood (Eq. 1).

$$ln\, p(\mathbf{x} \mid \pi, \mu, \Sigma) = \sum_{n=1}^{N} ln\left\{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k)\right\} \quad (1)$$

2) E Step: Evaluate the responsibilities using the current parameters values.

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{n} \pi_j \mathcal{N}(x_n \mid \mu_j, \Sigma_j)}$$

3) M Step: Re-estimate the parameters using the responsibility matrix calculated in E step.

$$\mu_k^{new} = \frac{1}{N_k}\sum_{n=1}^{N} \gamma(z_{nk}) x_n,$$

$$\Sigma_k^{new} = \frac{1}{N_k}\sum_{n=1}^{N} \gamma(z_{nk})(x_n - \mu_k^{new})(x_n - \mu_k^{new})^T$$

$$\pi_k^{new} = \frac{N_k}{N}$$

Where

$$N_k = \sum_{n=1}^{N} \gamma(z_{nk})$$

4) Evaluate the log-likelihood (Eq. 1), and check for convergence of either the parameters or the log likelihood. If the convergence criterion is not satisfied return to E step.

D. Variational Mixture of Gaussians

Full development of variational inference for Gaussian mixtures can be found in [12]. The variational posterior distribution is given by,

$$q(\pi, \mu, \Lambda) = q(\pi)\prod_{k=1}^{K} q(\mu_k, \Lambda_k)$$

Where, $q(\pi) = Dir(\pi \mid \alpha_k)$ is a Dirichlet distribution and $q(\mu_k, \Lambda_k)$ is a Gaussian-Wishart distribution and is given by,

$$q(\mu_k, \Lambda_k) = \mathcal{N}(\mu_k \mid m_k, (\beta_k \Lambda_k)^{-1})\mathcal{W}(\Lambda_k \mid W_k, v_k)$$

Similarly to the EM algorithm, the variational inference algorithm is performed in the following steps:

1) Initialize all parameters and hyperparameters of the distribution functions and the prior: $\alpha_0, \beta_0, m_0, v_0, W_0$. Also initialize the responsibility matrix r_{nk}.
2) Calculate the following statistical values,

$$N_k = \sum_{n=1}^{N} r_{nk}$$

$$\bar{x}_k = \frac{1}{N_k}\sum_{n=1}^{N} r_{nk}\mathbf{x}_n$$

$$\bar{S}_k = \frac{1}{N_k}\sum_{n=1}^{N} r_{nk}(\mathbf{x}_n - \bar{x}_k)(\mathbf{x}_n - \bar{x}_k)^T$$

3) Find the hyperparameters $\alpha_k, \beta_k, m_k, W_k, v_k$, corresponding to posterior distributions of mixing coefficients π_k and the parameters μ_k, Λ_k.

$$\alpha_k = \alpha_0 + N_k$$

$$\beta_k = \beta_0 + N_k$$

$$m_k = \frac{1}{\beta_k}(\beta_0 m_0 + N_k \bar{x}_k)$$

$$W_k^{-1} = W_0^{-1} + N_k S_k + \frac{\beta_0 N_k}{\beta_0 + N_k}(\bar{x}_k - m_0)(\bar{x}_k - m_0)^T$$

$$v_k = v_0 + N_k$$

4) Recalculate the responsibility matrix r_{nk} with the updated parameters in step 3,

$$r_{nk} \propto \tilde{\pi}_k \tilde{\Lambda}_k^{1/2} exp\left\{-\frac{D}{2\beta_k} - \frac{v_k}{2}(\mathbf{x}_n - m_k)(\mathbf{x}_n - m_k)^T\right\}$$

Where,

$$ln\,\tilde{\Lambda}_k = \sum_{i=1}^{D} \psi\left(\frac{v_k + 1 - i}{2}\right) + D\,ln\,2 + ln\,|W_k|$$

$$ln\,\tilde{\pi}_k = \psi(\alpha_k) - \psi(\hat{\alpha})$$

$\psi(x)$ is the digamma function and $\hat{\alpha} = \sum_k \alpha_k$.

5) Check for convergence using the variational lower bound [12], otherwise proceed to step 2.

III. Experimental Results

The clustering algorithms based in Gaussian mixtures with Expectation-Maximization (EM) and variational inference (VI) were tested on two databases (See subsection II-A). For comparison, we test different parametric and non-parametric classifiers. Within the parametric family, we use the Naive Bayes classifier with a shared covariance matrix among classes, known as the linear discriminant classifier (LDC) and the Naive Bayes classifier with a different covariance matrix per class, known as the quadratic discriminant classifier (QDC). Within the non-parametric family, we use the K-nearest neighbors (KNN) algorithm with K = 1 and K = 3 (KNN1 and KNN3, respectively). We evaluate the classifiers with the mean accuracy and the Area Under Curve (AUC) of the Receiver Operating Characteristic (ROC) [13]. We analyze the statistical significance of our results with a Kruskal-Wallis test (null hypothesis for equal medians) over 50 repetitions of each classifier. If the null hypothesis is not rejected, the difference between the algorithms is not statistically significant [14].

Table I shows mean accuracy results. Table II shows average AUC results.

TABLE I

Mean accuracy and standard deviations for different classifiers applied to databases

	EM	VI	LDC	QDC	KNN1	KNN3
MER-DB	86.74±4.86	85.21±3.45	78.43±2.32	69.62±2.56	88.50±2.15	86.50±2.26
ECG-DB	91.12±5.43	90.09±3.18	83.65±2.67	82.77±2.48	89.34±1.26	86.50±2.26

TABLE II

Area Under Curve for all classifiers

	EM	VI	LDC	QDC	KNN1	KNN3
MER-DB	0.977±0.014	0.968±0.008	0.599±0.014	0.634±0.015	0.948±0.008	0.886±0.019
ECG-DB	0.988±0.021	0.976±0.003	0.621±0.011	0.687±0.016	0.943±0.010	0.912±0.011

Figure 1 shows the clustered data from ECG database with EM algorithm. Figure 2 shows the clustered data from MER database with variational inference.

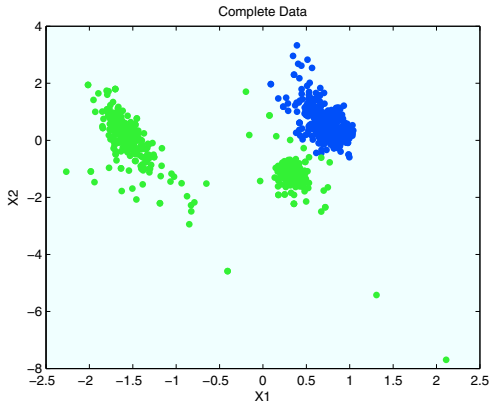

Fig. 1. Clustered data from ECG database for EM algorithm. X_1 and X_2 are features extracted from approximation coefficients of the first decomposition level using discrete Wavelet transform. Blue points are normal heartbeats and green points are the LBBB (Left Bundle Branch Block).

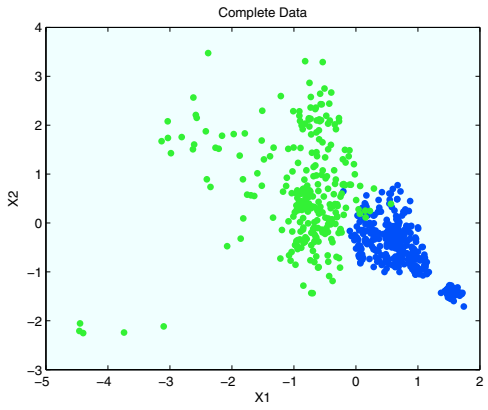

Fig. 2. Clustered data from MER database for variational inference. X_1 and X_2 are features extracted from approximation coefficients of the first decomposition level using adaptive filter banks. Blue points are samples from Subthalamic Nucleus and green points are samples from Thalamus, Zone Incerta and Substantia Nigra reticulata.

IV. Discussion

The analysis and results discussion are organized in the following items:

1) The accuracy results in table I, show a similar accuracy for some classifiers. The null hypothesis of equal means between EM, VI, KNN1 and KNN3 is not rejected. According to this analysis, the difference in accuracy performances between EM, VI, KNN3 and KNN1 is not statistically significant for both databases.

However, VI and EM exhibit better performance than LCD and QDC.

2) The table II shows the AUC results for all classifiers. Mean AUC performances for unsupervised classifiers with EM and VI are superior to mean AUC of parametric (LDC, QDC) and non-parametric classifiers (KNN1,KNN3). This result is highly relevant, because the AUC is a measurement of true positive ratio in a classifier. Making a multiple comparison test, we find that EM and VI do not reject the null hypothesis of equal means, so they are not statistically different. But, The null hypothesis of equal means between the unsupervised classifiers group (EM, VI) and the standard classifiers group (LDC, QDC, KNN1, KNN3) is rejected.
3) Figures 1 and 2 show the clustered data for two features. The separation of features X_1 y X_2 in both databases is not linear and we can see an overlap in many records. The average Rand index of EM is 0.82, while the rand index for VI is 0.76. This result shows a better performance in EM than VI. Another problem is the existence of outliers that disperse the variance of each Gaussian distribution in the mixture.
4) An important observation is the poor stability of EM algorithm, since it is sensitive to initial parameters values. This is because the EM convergence is made on local minimums. Tables I and II, present a higher standard deviation in EM than VI. In this work the initial values of all parameters is random, although it is possible to establish methodologies that reduce this issue.
5) Both algorithms (EM and VI) presented convergence problems due to infinitesimally small values in parameters. In some instances less than 10^{-250}. This quantization difficulty makes the algorithms do not converge, also it generates singular matrices. To address this, we normalize the feature matrix $\mathbf{X}$. The normalization of $\mathbf{X}$ affects directly the positive results in learning algorithms either supervised or unsupervised. One possible alternative is to use regularization methods to prevent excessively small values in parameters, which can be further work to refine the proposed methods.

V. Conclusions and Future Work

A considerable advantage of unsupervised learning is that it does not need a-priori information of class labels for pattern recognition. In this paper, we applied an approach composed of Gaussians mixtures and parameter estimation algorithms (Expectation Maximization and Variational Bayesian Inference) to perform clustering or unsupervised learning on two real databases. The basic idea is to use the advantage of Gaussians mixtures and the clustering possibility through responsibility matrix from both algorithms. For comparison, we employed standard classifiers like the parametric LDC,QDC and non-parametric KNN. Also, We performed a Kruskal-Wallis test to compare average performances among the classifiers. The difference in accuracy performances between EM, VI, KNN3 and KNN1 is not statistically significant for both databases. However, the AUC performances in EM and VI are superior than the others classifiers.

The most important difficulty to develop this methods is the instability of EM algorithm. It is sensitive to initial parameters values. This is because the EM convergence is made on local minimums. Also, there were quantization problems of very small values. For this reason, we normalized the processed data $\mathbf{X}$. The normalization affected the positive results in all learning algorithms.

In next works, we propose these methods for multiclass classification problems.

Acknowledgments

This research is developed under the projects:*"Desarrollo de un sistema automático de mapeo cerebral y monitoreo intraoperatorio cortical y profundo: aplicación neurocirugía"* and *"Desarrollo de un sistema efectivo y apropiado de estimación de volumen de tejido activo para el mejoramiento de los resultados terapéuticos en pacientes con enfermedad de parkinson intervenidos quirúrgicamente"*, both financed by Colciencias with codes 111045426008 and 111056934461 respectively. We also thank the contributions of the research group Gamma.

References

[1] J.A. Nasiri, M. Delescluse, M. Naghibzadeh, H.S. Yazdiand and B. Naghibzadeh, *ECG Arrhythmia Classification with Support Vector Machines and Genetic Algorithm*. In Third UK Sim European Symposium on Computer Modeling and Simulation.IEEE Computer Society, pp 187-192, 2009.

[2] C. Hamani, E.O. Richter, Y. Andrade, W. Hutchison, J.A. Saint-Cyr and A.M. Lozano, *Correspondence of microelectrode mapping with magnetic resonance imaging for subthalamic nucleus procedures*. Surg Neurol 63: 249–253, 2005.

[3] C. Pouzat, M. Delescluse, P. Viot, and J. Diebolt, *Improved spikesorting by modeling firing statistics and burst-dependent spike amplitude attenuation: a markov chain monte carlo approach*. J Neurophysiol 91, 2910-2928, 2004.

[4] H.B. Barlow, *Unsupervised learning*. Neural Computation, 1, 295-311, 1989.

[5] A. K. Jain, R. Duin, and J. Mao, *Statistical Pattern Recognition: A Review*. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol 22, No. 1, pp. 4 – 37, 2000.

[6] C. Zhai and J. A. Lafferty, *study of smoothing methods for language models applied to information retrieval*. ACM Transactions on Information Systems (TOIS), Vol 22, No. 2, pp. 179 – 214 , 2004.

[7] I. Cadez, P. Smyth and H. Mannila, *Probabilistic modelling of transactional data with applications to Profiling, Visualization and Prediction*. In Proceedings of the 7th ACM SIGKDD, 37 – 46, San Francisco, CA, 2001.

[8] J. Heer and E. Chi , *Identification of Web user traffic composition using multimodal clustering and information scent*. 1st SIAM ICDM, Workshop on Web Mining, 51 – 58, Chicago, IL, 2001.

[9] Y. Xu, V. Olman and Do. Xu, *Clustering gene expression data using a graph-theoretic approach: an application of minimum spanning trees*. Bioinformatics Vol. 18, No. 4, pp. 536 – 545, 2002.

[10] A. Dempster, N. Laird and D. Rubin, *Maximum likelihood from incomplete data via the EM algorithm*. Journal of the Royal Statistical Society, Series B, 1977.

[11] C. Bishop, *Pattern Recognition and Machine Learning, Springer*, Cap 9 pp. 423-440, 2006.

[12] C. Bishop, *Pattern Recognition and Machine Learning, Springer*, Springer, Cap 10 pp. 474-482, 2006.

[13] T. Fawcett, *An introduction to ROC analysis*, Pattern Recognition Letters, pp. 861–874, 2006.

[14] J. Pizarro, E. Guerrero and P. L. Galindo, *Multiple comparison procedures applied to model selection*, Neurocomputing, 48: 155-173, 2002.