Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Identification of Spike Sources using Proximity Analysis through Hidden Markov Models

Alvaro Orozco, Mauricio Alvarez, Enrique Guijarro, Germán Castellanos

Abstract— **Hidden Markov Models have shown promising results for identification of spike sources in Parkinson's disease treatment, e.g., for deep brain stimulation. Usual classification criteria consist in maximum likelihood rule for the recognition of the correct class. In this paper, we present a different classification scheme based in proximity analysis. For this approach matrices of Markov process are transformed to another space where similarities and differences to other Markov processes are better revealed. The authors present the proximity analysis approach using hidden Markov models for the identification of spike sources (Thalamo and Subthalamo sources, Gpi and GPe sources). Results show that proximity analysis improves recognition performance for about 5% over traditional approach.**

I. INTRODUCTION

Parkinson's disease is one of the most studied movement disorders. It is believed that the misbehavior of some neurons in subthalamus nucleus (STN) or Globus Pallidus internus (GPi) leads in to disease. Many patients, suffering Parkinson's disease, have received surgery for implanting microelectrode stimulator in STN, GPi or Vim nucleus, a procedure known as Deep Brain Stimulation (DBS). Depending on the microelectrode stimulator localization, different symptoms can be disabled.

One of the most difficult task on these surgeries is to identify the brain region where the microelectrode is going to be placed. Many surgeons use microelectrode recordings (MER) (see figure 1) to obtain better precision and physiologic validation of the nominal target detection [1]. The surgeon listens the rhythmic static created by the action potentials of the neurons near the electrode and based on this judgement decides where the microelectrode actually is. Unfortunately, these signals may be ambiguous to even a skilled surgeon, requiring several attempts to locate the nuclei. Besides this procedure is inappropriate to train new practitioners and it often suffers form lack of repeatability [2]. Therefore, an open issue consists in developing automatized techniques to support surgeon decision.

This research was supported by grant PI04/0303 from the Instituto de Salud Carlos III (Fondo de Investigación Sanitaria) in the framework of the "Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica (I+D+I)" and by Colciencias under contract 1110-14-17904

A. Orozco and M. Alvarez are lecturers of Programa de Ingeniería Eléctrica, Universidad Tecnológica de Pereira, La Julita, Pereira, Colombia {aaog,malvarez}@utp.edu.co

E. Guijarro is with Centro de Investigación e Innovación en Bioingeniería, Universidad Politécnica de Valencia, Valencia, España eguijarro@eln.upv.es

G. Castellanos is with Universidad Nacional de Colombia, Departamento de Eléctrica, Electrónica y Computación, Campus La Nubia, Manizales, Colombia gcastell@telesat.com.co

In [3], the authors proposed hidden Markov models (HMM) for characterizing dynamic behavior of spike trains, with similar average classification results than other methods already suggested in the literature [4]. Adopting a supervised classification scheme, it was possible to obtain classification accuracies over 94% for databases including signals from different patients. Standard classification approach was applied, which consisted in training an HMM for each class and use them later as class-conditional densities in a Bayes classification paradigm. For example, assuming a priori equiprobable classes, an unknown sequence is classified into the class whose model shows the highest probability (*likelihood*) of having generated this sequence (this is the well-known *maximum-likelihood* (ML) classification rule).

In this paper, an alternative classification scheme is used by means of a proximity-based hidden Markov model approach. The general idea behind proximity-based analysis consists in that given a set of pairwise dissimilarity values, a new representation space can be build, in which each object is described by these values. The classification is then performed in this new representation space. In the context of hidden Markov models, proximity analysis was first exploited in [5]. In that work, authors show results of the proposed methodology in a 2D shape recognition task and a face recognition problem. In our work, we use the proximity approach for the classification of different spike sources represented as the parameters of hidden Markov models.

A main issue in this approach is the high dimensionality of vectors in proximity space. For this particular problem, standard dimensionality reduction techniques could be used. In this work, we employ principal component analysis (PCA).

The organization of the paper is as follows. Section II describes proximity analysis. In section III, a brief description of a HMM is made. In section IV, stochastic matrices are transformed to proximity space. In section V, results of the proposed methodology in two different databases are shown. Finally, in section VI discussion is stated.

II. PROXIMITY ANALYSIS

The first step in a patter recognition problem is the search for an appropriate representation of objects. Then, a decision rule can be constructed, which discriminates between different categories and which is able to generalize well. Conventionally, objects are represented by characteristic features. In some cases, however, a feasible feature-based description of objects might be difficult to obtain or inefficient for learning purposes. One alternative representation is based on

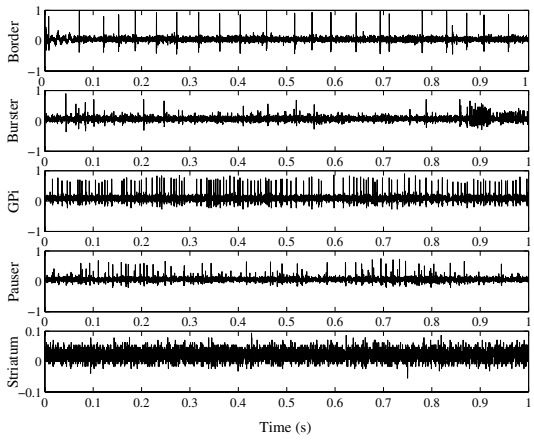

Fig. 1. Signals from different brain regions. Border, Burster, Pauser and Striatum are regions around GPi zone.

similarity or dissimilarity relations between objects. When properly defined, it might be advantageous for solving class identification problems [6].

The general idea behind proximity-based analysis consists in that given a set of pairwise dissimilarity values, a new representation space can be build, in which each object is described by these values. The classification is then performed in this new representation space. Experimental research has shown that proximity-based philosophy improves performance in complex classification tasks [6]. Resemblances and diferencies between proximity analysis and kernel methods have been stated in [7].

III. Hidden Markov Models

A hidden Markov model is basically a Markov chain where the output observation is a random variable generated according to an output probabilistic function associated with each state. It is defined by a set of parameters $\lambda = (\mathbf{A}, \mathbf{B}, \boldsymbol{\pi})$, where $\boldsymbol{\pi}$ is the initial state distribution, $\mathbf{A}$ is the transition probability matrix and $\mathbf{B}$ is the output probability matrix. Output probabilities can also be modeled through continuous probability density functions [8].

Usually, as a classification rule, a maximum-likelihood (ML) criteria is carried out, where an unknown sequence $\mathbf{O}$ is assigned to the class showing the highest likelihood $P(\mathbf{O}|\lambda)$. This requires training L HMMs for a problem of L classes. In the proximity-based approach, instead of training one HMM for each class, we could train one model for each training sequence and assign an unknown sequence $\mathbf{O}$, represented for a particular λ set, to the class of the model showing the highest likelihood.

Next section describes formally the process of transforming sequences of spike trains to the proximity space.

IV. Transformation of HMM matrices

A. Distance Definition

The most important point for practical application of the proximity analysis in an HMM context is the definition of an appropriate distance between a pair of hidden Markov models.

In [5], a simple distance measure is employed: $P_{ij} = \hat{P}(\mathbf{O}_i|\lambda_j)$, the probability that the given model λ_j generates the observation sequence $\mathbf{O}_i$, normalized w.r.t. T_i, the length of the observation sequence $\mathbf{O}_i$. This number is calculated using the *forward-backward* procedure [8]. The above expression used as a distance is not symmetric, a useful property that a metric should have. In order to circumvent this matter and due to good results in previous experiments [3], we use the definition of distance proposed in [8] between a pair of sequences represented as parameters of an HMM. The *similarity measure* (SM) between two distinct observation sequences $\mathbf{O}_1$ and $\mathbf{O}_2$ of discrete symbols is defined as:

$$\sigma(\mathbf{O}_1, \mathbf{O}_2) = \sqrt{\frac{P_{21}P_{12}}{P_{11}P_{22}}}$$

where, $P_{ij} = \hat{P}(\mathbf{O}_i|\lambda_j) = P(\mathbf{O}_i|\lambda_j)^{1/T_i}$. In some cases, it may be more convenient to represent the similarity between two trajectories through a distance measure $d(\mathbf{O}_1, \mathbf{O}_2)$, rather than a similarity measure. Given the similarity measure $\sigma(\mathbf{O}_1, \mathbf{O}_2)$, such a distance is expressed as [8]

$$d(\mathbf{O}_1, \mathbf{O}_2) = -\log \sigma(\mathbf{O}_1, \mathbf{O}_2) \tag{1}$$

B. HMM matrices in proximity space

Consider a classification problem with L classes. For each class $\ell \in \{1, 2, \ldots, L\}$, there is a set of N_ℓ training sequences $\mathcal{S}_\ell = \{\mathbf{O}_1^\ell, \ldots, \mathbf{O}_{N_\ell}^\ell\}$. Thus, $N = \sum_\ell N_\ell$ is the total size of the training set $\mathcal{S} = \bigcup_{\ell=1}^{L} \mathcal{S}_\ell$.

Let $\mathcal{R} = \{\mathbf{P}_1, \ldots, \mathbf{P}_R\}$ be a set of R representative objects; these objects may belong to the set of training sequences ($\mathcal{R} \subseteq \mathcal{S}$). Let $\mathcal{D}_\mathcal{R}(\mathbf{O})$ be a function that returns the vector of proximities between an arbitrary sequence $\mathbf{O}$ and all sequences in $\mathcal{R}$, which is (using equation (1))

$$\mathcal{D}_\mathcal{R}(\mathbf{O}) = \begin{bmatrix} d(\mathbf{O}, \mathbf{P}_1) \\ \vdots \\ d(\mathbf{O}, \mathbf{P}_R) \end{bmatrix} \in \mathbb{R}^R$$

Once "the proximity space" $\mathbb{R}^R$ (where the proximity vectors exist) is defined, any standard classifier can, in principle, be used.

Regarding the choice of $\mathcal{R}$, different approaches can be adopted; the basic one is to choose $\mathcal{R} = \mathcal{S}$, the whole training set. With this choice the dimensionality of the proximity space is equal to N, the cardinality of the training set. In this case, a classifier must be designed on an N-dimensional space using only N training sequences, suggesting that some dimensionality reduction technique should be adopted.

C. Dimensionality Reduction

For vector dimensionality reduction in the proximity space, we use a standard multivariate technique named Principal Component Analysis (PCA). PCA searches for a new coordinate system under which the retained variance under linear projection is maximal [9].

V. Experimental Setup

A. Databases

Two different databases, as shown in table I, were used for experiments. For the subthalamic nucleus trajectorie, database (DB1) includes regions such as thalamus and subthalamic nucleus. To get GPi zone, database (DB2) includes boundary regions such as Border, Burster, Pauser and Striatum regions (see figure 1), jointly named as GPe (Globus Pallidus externus). All signals are 10 seconds length.

TABLE I
DATABASES USED FOR EXPERIMENTS

Database	Region	Samples
DB1	Thalamus	50
	Subthalamic Nucleus	50
DB2	Border	8
	Burster	8
	GPi	10
	Pauser	5
	Striatum	3

B. Feature extraction

Each signal is segmented in frames of 250 milliseconds (ms) length with an overlapping of 50 ms, following results from [2]. For each frame, a Haar wavelet transform is applied (due to its better time resolution property) and for each scale, the two wavelet coefficientes with the largest magnitude are retained and used as part of the observation vector. This is done in this way due to wavelet's energy compaction property [10].

C. Structure of hidden Markov models

In order to decrease computational cost, discrete hidden Markov models are used to describe different target regions. Ergodic topology models [11] were trained with 5 and 10 states and 64 and 128 words for the codebook [11], [12]. The parameters λ, describing the HMM model of each region class, are estimated using the standard Expectation-Maximization algorithm [12].

D. Validation

Two different validation approaches are examined: classical approach based on ML criteria, where likelihood ratio is calculated and the model having the highest value is selected as the correct one and the proximity analysis approach using five different classifiers, namely, Regularized Linear Normal density-based Classifier (RLNC), Regularized Quadratic Normal density-based Classifier (RQNC) [6], Parzen Classifier, 1-Nearest Neighbor Classifier (1-NN) and 3-Nearest Neighbor Classifier (3-NN) [13]. For the proximity analysis, definition of distance in equation (1) is employed. Due to reduced database sizes, *leave-one-out* validation method is used and experiments are repeated 5 times.

E. Results over Database DB1

Table II shows average classification results obtained using different approaches for database DB1 with 64 codewords for the codebook. Table III shows average classification results obtained using different approaches for database DB1 with 128 codewords for the codebook. From tables II and III it can be seen that, independently of the number of states and the number of codewords employed for the codebook, non-parametric classifiers such us Parzen classifier and 1-NN classifier, allows to obtain better classification results for classes Thalamus and Subthalamic Nucleus.

TABLE II
CLASSIFICATION RESULTS FOR DATABASE DB1 AND 64 CODEWORDS

Method	5 States		10 States	
	Mean (%)	Std (%)	Mean (%)	Std (%)
ML	94.0	1.80	93.0	3.08
RLNC	99.4	0.50	99.9	0.01
RQNC	99.8	0.50	99.0	0.00
Parzen	99.9	0.01	99.9	0.01
1-NN	99.9	0.01	99.9	0.01
3-NN	97.4	0.50	97.6	0.58

TABLE III
CLASSIFICATION RESULTS FOR DATABASE DB1 AND 128 CODEWORDS

Method	5 States		10 States	
	Mean (%)	Std (%)	Mean (%)	Std (%)
ML	97.2	1.3	97.4	2.19
RLNC	99	0.00	99	0.00
RQNC	99	0.00	99.4	0.96
Parzen	99.9	0.01	99.9	0.01
1-NN	99.9	0.01	99.9	0.01
3-NN	97	0.00	98	0.00

For DB1, vectors in the proximity space have a dimension of 100 (equal to the number of observations in the database). Table IV summarizes the best classification results obtained using PCA for dimensionality reduction in the proximity space. For this case HMMs with 5 states and 64 codewords were used.

Table IV shows that almost all classifiers obtain high accuracies (99.9%) with reduced dimension, 15 components, for the principal component analysis. Experiments also show that, below 15 components, Parzen classifier and 1-NN classifier obtain, with only 5 components, a 99% classification accuracy and with 1 component more than 90%.

TABLE IV
BEST CLASSIFICATION RESULTS FOR DATABASE DB1 USING PCA IN THE PROXIMITY SPACE

Classifier	Reduced Dimension	Mean (%)	Std (%)
RLNC	15	99.9	0.01
RQNC	15	99.9	0.01
Parzen	15	99.9	0.01
1-NN	15	99.9	0.01
3-NN	80	97.8	0.50

F. Results over Database DB2

Table V shows average classification results using different approaches over database DB2 with 64 words in the codebook. Table VI shows classification results over the same database, this time with 128 codewords.

Similar to results obtained for database DB1, tables V and VI indicate that the highest average accuracies are obtained using the non-parametric classifiers Parzen and 1-NN. Standard deviations in both cases are less than for other classifiers.

TABLE V
CLASSIFICATION RESULTS FOR DATABASE DB2 AND 64 CODEWORDS

Method	5 States		10 States	
	Mean (%)	Std (%)	Mean (%)	Std (%)
ML	92.12	3.60	89.41	5.34
RLNC	92.94	3.35	93.53	3.22
RQNC	87.06	2.63	83.53	3.35
Parzen	97.06	2.08	97.06	2.08
1-NN	96.47	3.22	95.88	3.35
3-NN	85.29	2.94	82.94	2.46

TABLE VI
CLASSIFICATION RESULTS FOR DATABASE DB2 AND 128 CODEWORDS

Method	5 States		10 States	
	Mean (%)	Std (%)	Mean (%)	Std (%)
ML	93.00	2.63	92.94	4.46
RLNC	94.71	1.32	91.76	3.83
RQNC	87.65	1.32	83.53	4.46
Parzen	97.06	2.08	94.71	3.83
1-NN	96.47	2.46	95.29	3.35
3-NN	95.29	2.63	92.35	3.90

TABLE VII
BEST CLASSIFICATION RESULTS FOR DATABASE DB2 USING PCA IN THE PROXIMITY SPACE

Classifier	Reduced Dimension	Mean (%)	Std (%)
RLNC	12	97.06	2.08
RQNC	9	95.88	1.61
Parzen	6	97.65	2.46
1-NN	3	98.24	2.63
3-NN	3	97.06	2.08

Transforming vectors in the proximity space using PCA allows to attain 98.24% classification accuracy by means of the 1-NN classifier with just 3 components. This can be seen in table VII. For this experiment HMMs with 5 states and 64 codewords were examined.

VI. DISCUSSION

Average classification results using proximity analysis through hidden Markov models are greater than those obtained using traditional ML rule for identification of spike sources.

In the proximity space, non-parametric classifiers allow higher classification accuracies. In particular, for database DB1, 1-NN classifier and Parzen classifier get the highest accuracy with the lowest reduction factor $(15/100)$. For database DB2, one nearest neighbor classifier has better performance with a suitable reduction factor $(3/34)$.

Although high classification accuracies were obtained using non-parametric classifiers, some disadvantages of the method must be solved (specially when the number of samples is big enough): large storage requirement, large computational effort and, in particular for the nearest neighbor classifier, sensitivity to noisy examples. In order to overcome these issues, prototype selection techniques [14] should be a direction for future work.

VII. ACKNOWLEDGMENTS

The authors gratefully acknowledge the contributions of Colciencias, the Measurement and Instrumentation Laboratory of Universidad Tecnológica de Pereira, Colombia and the Center for Research and Innovation on Bioengineering of Universidad Politécnica de Valencia, Spain.

REFERENCES

[1] J.H. Falkenberg , J. McNames , K.J. Burchiel Statistical methods of analysis and visualization of extracellular microelectrode recordings, *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology*, vol. 3, 2003, pp 2515-2518

[2] R.A. Santiago , J. MacNames , K. Burchiel and George G. Lendaris Developments in understanding neuronal spike trains and functional specializations in brain regions, *Neural Networks*, vol. 16, 2003, pp 601–607

[3] A. A. Orozco, C. Guarnizo, M. Álvarez G. Castellanos and E. Guijarro, Spike Source Identification using Artificial Intelligence Techniques *Proceedings of The 3RD IEE International Seminar on Medical Applications of Signal Processing*, vol. 1, 2005, pp 105-109

[4] Q. Fu, M. Clements and K. Mewes. Neural cell type recognition between globus pallidus externus and globus pallidus internus by gaussian mixture modeling *Proceedings of the ICASSP*, vol. 1, 2005, pp 53-56

[5] M. Bicego, V. Murino and M. Figuereido Similarity-based classification of sequences using hidden markov models, *Pattern Recognition*, vol. 37, 2004, pp 2281-2291

[6] E. Pekalska and R. Duin. Dissimilarity representations allow for building good classifiers. *Pattern Recognition Letters*, vol. 23, 2002, pp 943-956

[7] E. Pekalska, R. Duin and P. Paclík. A Generalized kernel approach to Dissimilarity-based Classification. *Journal of Machine Learning Research*, vol. 2, 2001, pp 175-211

[8] M.C. Nechyba and Y. Xu Stochastic Similarity for Validating Human Control Strategy Models. *IEEE Transactions On Robotics And Automation*, vol. 14, 1998, pp 437-451

[9] R. A. Johnson and D.W. Wichern. *Applied Multivariate Statistical Analysis*. Prentice Hall, Upper Saddle River, New Jersey 2002.

[10] K.C. Ho and S.D. Blunt. Adaptive sparse system identification using wavelets. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*. vol. 49, 2002, pp. 656-667.

[11] L. Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of The IEEE*, vol. 77, 1989, 257-286

[12] X. Huang, A. Acero and H.W. Hon *Spoken Language Processing*, Prentice Hall, Upper Saddle River, New Jersey; 2001

[13] R.O. Duda, P.E. Hart and D.G. Stork. *Pattern Classification*, John Wiley & Sons, 2001.

[14] E. Pekalska, R. Duin and P. Paclík. Prototype selection for dissimilarity based classifiers. *Pattern Recognition*, vol. 39, 2006, pp 189-208