SPARSE CODING AND ROUGH SET THEORY-BASED HYBRID APPROACH TO THE CLASSIFICATORY DECOMPOSITION OF CORTICAL EVOKED POTENTIALS

Grzegorz M. Boratyn[1], Tomasz G. Smolinski[1], Mariofanna Milanova[2], and Andrzej Wrobel[3]

[1]Computational Intelligence Laboratory,
Department of Electrical and Computer Engineering, University of Louisville,
Louisville, KY 40292, USA
{grzegorz.boratyn, tomasz.smolinski}@louisville.edu

[2]Department of Computer Science, University of Arkansas at Little Rock,
Little Rock, AR 72204, USA
mgmilanova@ualr.edu

[3]Laboratory of Visual System,
Department of Neurophysiology, Nencki Institute of Experimental Biology,
3 Pasteur Street, 02-093 Warsaw, Poland
wrobel@nencki.gov.pl

ABSTRACT

This paper presents a novel approach to classification of decomposed cortical evoked potentials (EPs). The decomposition is based on learning of a sparse set of basis functions using an Artificial Neural Network (ANN). The basis functions are generated according to a probabilistic model of the data. In contrast to the traditional signal decomposition techniques (i.e. Principle Component Analysis or Independent Component Analysis), this allows for an overcomplete representation of the data (i.e. number of basis functions that is greater than the dimensionality of the input signals). Obviously, this can be of a great advantage. However, there arises an issue of selecting the most significant components from the whole collection. This is especially important in classification problems based upon the decomposed representation of the data, where only those components that provide a substantial discernibility between EPs of different groups are relevant. To deal with this problem, we propose an approach based on the Rough Set theory's (RS) feature selection mechanisms. We design a sparse coding- and RS-based hybrid system capable of signal decomposition and, based on a reduced component set, signal classification.

1. INTRODUCTION

Signal decomposition plays a crucial role in analysis of Evoked Potentials (EPs) or Event-Related Potentials (ERPs) [1], [2]. Among the most popular methods of EP decomposition one will find Principal Component Analysis (PCA) [3], Independent Component Analysis (ICA) [4], [5], [6] or wavelet-based analysis [7]. In general, a common way to represent real-valued EPs is via a linear superposition of some basis functions. For instance, a standard wavelet analysis produces coefficients for expressing a signal as a linear combination of "wavelet packets." Bases such as wavelets can provide a very useful representation of some signals, however they have serious limitations in terms of the number as well as the type of the basis functions they employ [7], [8].

An alternative and more general method of signal representation via transformation uses sparse coding [9], [10]. This methodology is based on the assumption that the data can be represented by a set of statistically independent events (i.e. basis functions). An additional conjecture is made that the appropriate form for the probability distribution of those events is that they are sparse (i.e. the data can be usually described in terms of a relatively small number of basis functions). At the same time, an overcomplete representation allows for a greater number of basis functions than the dimensionality of the input signals, which can provide much greater flexibility in terms of capturing structures hidden in data [11], [12], [13].

However, even if the sparseness of the basis functions is accounted for and preserved, the issue of selecting the most significant components from the basis set is still crucial. This is especially important for signal classification applications that use sparse coding as a data preprocessing (i.e. transformation) tool. In such applications, one is mostly interested in selecting those components that provide the best discernibility between signals that belong to different categories.

While a similar idea of data dimensionality reduction has already been utilized in a hybridization of PCA and Rough Sets (RS) [14], it appears that an application of this approach to the sparse coding is quite novel.

The paper is organized into the following sections: first, in Sect. 2, we discuss the utilized data model. Then, in Sect. 3, we discuss the application of the theory to the problem of selecting the most significant components in the sense of signal classification while in Sect. 4, the results of our numerical experiments based on the approach are presented. Sect. 5 provides a short summary.

2. DATA MODEL

The primary step of examining the form of EPs or ERPs is to decompose them into parts (i.e. components). Basis functions create a domain in which an EP measurement can be represented as a vector [1].

We assume that each data vector x is described with a set of basis functions **M** weighted by coefficients **a**, and additive noise *e*:

$$\mathbf{x} = \mathbf{M}\mathbf{a} + e \,. \tag{1}$$

Probabilistic formulation of the problem, presented in [15] and [16] is used here to estimate the values of **a** and **M**. A given data point can have many possible representations, nevertheless this ambiguity is removed by a proper choice for the prior probability of the basis coefficients [9], [10],

$$P(\mathbf{a} \mid \mathbf{M}) = \prod_i \exp\left(-S(a_i)\right), \tag{2}$$

which specifies the probability of the alternative representations. $S(a_i)$ is a sparseness term defined as $S(a_i) = b\log(1+(a_i / g)^2)$, where b and g are scaling factors.

The neural network model that learns basis functions and computes the coefficients values, proposed by Olshausen in [10], is used. The coefficients are inferred from **x** by maximizing the posterior probability $P(\mathbf{a}|\mathbf{x}, \mathbf{M})$, which can be expressed via Bayes' rule as:

$$P(\mathbf{a} \mid \mathbf{x}, \mathbf{M}) = P(\mathbf{x} \mid \mathbf{a}, \mathbf{M}) P(\mathbf{a} \mid \mathbf{M}) \,. \tag{3}$$

The first term of the right hand side of the proportion specifies the likelihood of the signal for a particular state of the model (given by **M** and **a**):

$$P(\mathbf{x} \mid \mathbf{a}, \mathbf{M}) \propto \exp\left(-\frac{l}{Z_{sN}} \left|\mathbf{x} - \mathbf{M}\mathbf{a}\right|^2 \right), \tag{4}$$

where $Z_{\sigma N}$ is normalizing constant, $l = 1/s^2$, and s is the standard deviation of the additive noise. The maximization of $P(\mathbf{a}|\mathbf{x}, \mathbf{M})$ is accomplished via gradient ascent on the log-probability which using (3) and (4) is given by:

$$\Delta\mathbf{a} = l_N \mathbf{M}^T \mathbf{e} - \frac{\partial S(\mathbf{a})}{\partial \mathbf{a}}, \tag{5}$$

where $\mathbf{e} = \mathbf{x} - \mathbf{M}\mathbf{a}$. The basis functions are found by minimizing model's estimate of the average code length of the model L:

$$L = -\left\langle \log P(\mathbf{x} \mid \mathbf{M}) \right\rangle, \tag{6}$$

where $\langle \cdot \rangle$ denotes average value , and

$$P(\mathbf{x} \mid \mathbf{M}) = \int P(\mathbf{x} \mid \mathbf{a}, \mathbf{M}) P(\mathbf{a} \mid \mathbf{M}) \, d\mathbf{a} \,. \tag{7}$$

The basis function are found via gradient descent on L:

$$\Delta\mathbf{M} \propto -\frac{\partial L}{\partial \mathbf{M}} \propto \left\langle \mathbf{e}\mathbf{a}^T \right\rangle. \tag{8}$$

After each learning step the values of basis functions need to be rescaled such that their L2 norms are changed by the factor $\frac{\left\langle a_i^2 \right\rangle}{s^2}$, to ensure appropriate variance of coefficients values (see [10]).

3. ROUGH SETS-BASED SELECTION OF CLASSIFICATION-RELEVANT BASIS FUNCTIONS

Sparse coding provides a very efficient and useful mechanism for data transformation. In traditional techniques, such as PCA, feature extraction is based upon minimization of the reconstruction error and the "most expressive" components are selected according to some statistical criteria [19], [20]. Sometimes, however, the reconstruction error is not so important while feature reduction task is crucial. This is especially true for any classification problem performed on the new representation of the data (i.e. coefficients for a given set of basis functions), for which one might be looking for the smallest possible set of components that explain all the variations between different classes of objects. In terms of evoked potentials, for instance, that would not only allow for a decomposition of the signals into some meaningful components, but also for determination of those among them that are the most significant for discernibility between different groups of signals. For this kind of problems, traditional approaches do not guarantee that selected components, as a feature vector in the new representation, will be competent for classification.

One way to deal with this problem, would be to apply the theory of rough sets [21], [22], [23]. In this case, especially useful will be the concept of reducts, inherently embedded in the theory. Intuitively, an application of the methodology of sparse coding will yield an adequate and detailed model of the input data, whilst the rough sets-based search for reducts will determine the most significant components in that model, in terms of data classification.

The only issue that must be addressed before applying rough sets to the search for a reduced set of components is the fact that the coefficients of the basis functions are real-valued. This problem, however, can be easily solved by utilizing some discretization techniques that will transform the real values of coefficients into intervals that will be assigned ordered, integer values (i.e. labels) [24], [25], [26].

4. EXPERIMENTS AND RESULTS

4.1. Data

In the experiments conducted at the Laboratory of Visual System, Nencki Institute of Experimental Biology, Warsaw, Poland, a piezoelectric stimulator was attached to a vibrissa of a rat [27], [28]. An electrical impulse of 5 V amplitude and 1 ms duration was applied to the stimulator causing the vibrissa deflection. Evoked Potentials were then registered – each of them related to a single stimulus.

Evoked potentials have been used for many years as a measurement of dynamic events occurring in nervous systems that accompany and are related to some defined sequences of behavior [1]. Based on same previous work, a hypothesis about a relation between two components of the registered evoked potentials and particular brain structures (i.e. supra- and infra-granular pyramidal cells) was stated. In order to verify the hypothesis, two additional types of stimuli were applied: 1) a cooling event applied to the surface of the cortex (allowing to temporarily "switch off" some structures of the brain), and 2) a supplementary aversive stimulus – electrical shock applied to the rat's ear (in order to cope with the phenomenon of habituation). Main goal of these experiments was to investigate those stimuli in the sense of their impact on the brain activity represented by the registered EPs.

A single, four-level electrode positioned in the cortex of a rat collected the data. The electrode registered brain activity in a form of evoked potentials on four depths (i.e. channels) simultaneously as described in [27]. Each evoked potential was sampled with frequency of 2kHz and is described in the database by 100 values. The complete database consists of four separate data sets for each of the four channels with 882 records in each data set.

Because of the fact that the third channel's electrode (0.4 mm) was located in the closest position to the granular cells (laying in the middle between supra- and infra-granular, pyramidal cells – see [27], [28], [29]) and yielded the most "representative" perspective at the activity of the cortex, quite often this level was acknowledged the most meaningful and interesting one and was given particular attention.

4.2. Analysis

A sequence of experiments was performed in order to verify and analyze the performance of the proposed approach. The overall effectiveness of the algorithm, in the light of previous findings, was considered. The most important issue was to investigate if the sparse coding-based approach was capable of determining similar components to the ones obtained in previous work by PCA and ICA (see [29] and [8]). Secondly, it was crucial to explore the ability of the system to automatically select those of the components that really mattered in terms of the discrimination between the registered EPs. Those components, were assumed to explain most of the differences between EPs in the database (ultimately between the normal and cooled potentials).

The complete set of 882 evoked potentials registered on the 3[d] channel was used as the input to the neural network. Based on the conclusions derived from some preliminary work on the same data (i.e. having too many basis functions, some of them appeared to be completely insignificant – see [8]) the goal of the algorithm was to determine a set of 10 basis functions. The graphical representation of the computed basis functions is shown in Fig. 1.

It is important to point out that the "polarization" of the basis functions is not really relevant, since the coefficients can also take negative values.

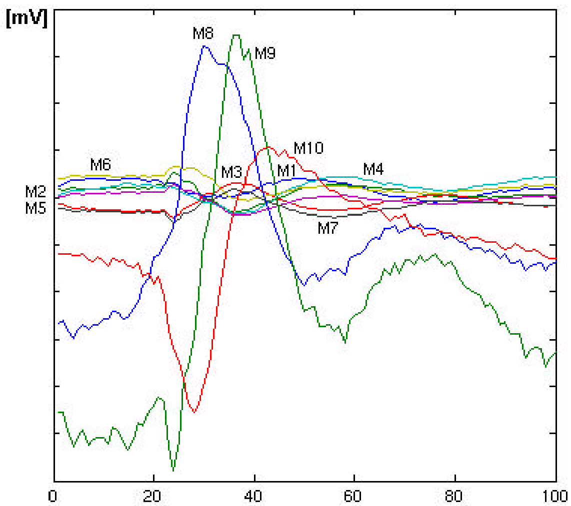

Figure 1: 10 basis functions computed from the complete data set (M*x* denotes the *x*-th basis function).

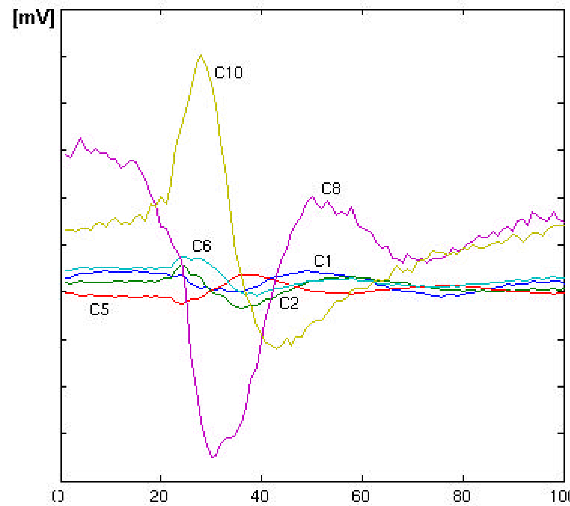

Figure 2: Sample reduced averaged components in the 3rd channel (C*x* denotes averaged the *x*-th component).
Discretization method: Equal Frequency Bin
Reduction method: Johnson Algorithm (reduct: {C1, C2, C5, C6, C8, C10})

Based on this new representation of the input data (i.e. basis functions + coefficients), a rough sets' search for reducts was applied in order to determine the set of classification-relevant components. First, however, the coefficients were discretized using three different discretization techniques – *Equal Width Bin*, *Equal Frequency Bin*, and *Holte's One Rule Discretizer* (for more information on these techniques see [24], [25], [26]). After the discretization, the Johnson's algorithm [30], [31] for searching for reducts was applied.

Various configurations of the discretization and/or reduction algorithms' were investigated and one the most interesting results are shown in Fig. 2.

Since the classificatory attribute (i.e. cooling event) was only approximately defined in our database, it was impossible to directly determine the classification accuracy based on the discretized and reduced data. However, the most important part of this project was to verify the coherency of the results obtained with our approach with the results produced by other methods and, based on this, improve and extend the process of EP analysis by providing an automatic methodology for signal decomposition and selection of significant components.

This goal was successfully achieved since the characteristics of some basis functions determined by the neural network, were very similar to the first two components received via both, PCA and ICA (see [29], [8]), and those two basis functions, were always selected by the reduction algorithms. Additionally, as it can be clearly seen in Fig. 2, after the signal decomposition, the system pointed out several components that provide an ability to discern between the EPs in the database (guaranteed by the reduction algorithm – indiscernibility relation holds).

In all the experiments described in this section, the neural network implemented by Olshausen [32] was used. Additionally, the Rosetta system [30] along with some authors' implementations of rough sets were employed for the RS-based value discretization and feature selection/reduction.

5. CONCLUSIONS

On the basis of the experiments and the analysis described above we can conclude that the proposed sparse coding- and RS-based hybrid system provides a useful and effective tool in terms of classification of evoked potentials. Our results, obtained via the methodology of sparse coding, were coherent with previous work in terms of the signal's main components. This suggests that this approach delivers useful capabilities in terms of signal decomposition and classification. On the other hand, the system provides a significant extension to the traditional approaches thanks to the mechanisms of an automatic determination of relevant components, in terms of signal classification.

6. REFERENCES

[1] W. J. Freeman, "Measurement of Cortical Evoked Potentials by Decomposition of their Wave Forms," *J. of Cybernetics and Information Science*, 2-4, pp. 22-56, 1979

[2] I. Gath and G.F. Inbar (eds.), *Advances in Processing and Pattern Analysis of Biological Signals*, Kluwer Academic/Plenum Publishers, New York, NY, 1996

[3] R. M. Chapman and J. W. McCrary, “EP Component Identification and Measurement by Principal Components Analysis,” *Brain and Cognition*, 27, pp. 288-301, 1995

[4] S. Makeig, A. J. Bell, T.-P. Jung, and T. J. Sejnowski, “Independent Component Analysis of Electroencephalographic Data,” *Advances in Neural Information Processing Systems*, 8, pp. 145-151, 1996

[5] S. Makeig, T.-P. Jung, A. J. Bell, D. Ghahremani, and T. J. Sejnowski, “Blind Separation of Auditory Event-related Brain Responses into Independent Components,” *Proc. Natl. Acad. Sci. USA*, 94, pp. 10979-10984, 1997

[6] S. Makeig, et al., *ICA Toolbox Tutorial*, Available: http://www.cnl.salk.edu/~scott/tutorial/

[7] J. Raz, L. Dickerson, and B. Turetsky, “A Wavelet Packet Model of Evoked Potentials,” *Brain and Language*, 66, pp. 61-88, 1999

[8] M. Milanova, T. G. Smolinski, G. M. Boratyn, J. M. Zurada, and A. Wrobel, “Sparse Correlation Kernel Analysis and Evolutionary Algorithm-based Modeling of the Sensory Activity within the Rat’s Barrel Cortex,” *Proc. of the International Workshop on Pattern Recognition with Support Vector Machines* (SVM2002), pp. 198-212, 2002

[9] B. Olshausen and D. J. Field, “Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1?,” *Vision Research*, 37(23), pp. 3311-3325, 1997

[10] B. Olshausen, “Sparse Codes and Spikes,” In: R. P. N. Rao, B. A. Olshausen, and M. S. Lewicki (eds.), *Probabilistic Models of Perception and Brain Function*, MIT Press, Cambridge, MA, pp. 245-260, 2001

[11] S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic Decomposition by Basis Pursuit,” *Technical report, Dept. Stat.*, Stanford University, 1996

[12] S. G. Mallat and Z. Zhang, “Matching Pursuits with Time-Frequency Dictionaries,” *IEEE Trans. on Signal Processing*, 41(12), pp. 3397-3415, 1993

[13] M. Lewicki and T. Sejnowski, “Learning Overcomplete Representations,” *Neural Computation*, 12, pp. 337-365, 2000

[14] R. Swiniarski, “Rough Sets and Principal Component Analysis and Their Applications in Data Model Building and Classification,” In: S. K. Pal and A. Skowron (eds.), *Rough Fuzzy Hybridization: A New Trend in Decision-Making*, Springer-Verlag, Singapore, pp. 275-300, 1999

[15] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995

[16] M. S. Lewicki and B. A. Olshausen, “Probabilistic Framework for Adaptation and Comparison of Image Codes,” *J. Opt. Soc. of Am.*, 16, pp. 1587-1600, 1999

[17] D. J. Field, “What is the Goal of Sensory Coding?,” *Neural Computation*, 6, pp. 559-601, 1994

[18] M. Yoshioka and S. Omatu, “Independent Component Analysis Using Time Delayed Sampling,” Presented at the *IEEE International Joint Conference on Neural Networks*, Como, Italy, July 24-27, 2000

[19] I. T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, New York, 1986

[20] E. Oja, “A Simplified Neural Model as a Principal Components Analyzer,” *J. of Mathematical Biology*, 15, pp. 267-273, 1982

[21] Z. Pawlak, “Rough Sets,” *International J. of Computer and Information Sciences*, 11, pp. 341-356, 1982

[22] W. Marek and Z. Pawlak, “Rough Sets and Information Systems,” *Fundamenta Matematicae*, 17, pp. 105-115, 1984

[23] Z. Pawlak, “Knowledge, Reasoning, and Classification – A Rough Set Perspective,” *Bulletin of the European Association for Theoretical Computer Science* (EATCS), 38, pp. 199-210, 1989

[24] S. H. Nguyen and H. S. Nguyen, “Discretization methods with back-tracking,” *Proc. of the Fifth European Congress on Intelligent Techniques and Soft Computing* (EUFIT'97), pp. 201-205, 1997

[25] S. H. Nguyen and H. S. Nguyen, “Discretization Methods in Data Mining,” In: L. Polkowski and A. Skowron (eds.), *Rough Sets in Knowledge Discovery*, Physica-Verlag, Heidelberg, pp. 451-482, 1998

[26] S. H. Nguyen, “Discretization Problems for Rough Set Methods,” *Proc. of the First International Conference on Rough Sets and Current Trend in Computing* (RSCTC'98), pp. 545-552, 1998

[27] E. Kublik and P. Musial, “Badanie ukladów czuciowych metoda potencjalów wywolanych,” *Kosmos*, 46, pp. 327-336, 1997 (*in Polish*)

[28] A. Wrobel, E. Kublik, and P. Musial, “Gating of the Sensory Activity within Barrel Cortex of the Awake Rat,” *Exp. Brain Res.*, 123, pp. 117-123, 1998

[29] E. Kublik, P. Musial, and A. Wrobel, “Identification of Principal Components in Cortical Evoked Potentials by Brief Surface Cooling,” *Clinical Neurophysiology*, 112, pp. 1720-1725, 2001

[30] *ROSETTA: A Rough Set Toolkit for Analysis of Data*, Available: http://www.idi.ntnu.no/~aleks/rosetta/

[31] A. Øhrn, *ROSETTA Technical Reference Manual* (update: May 25, 2001), Available: http://www.idi.ntnu.no/~aleks/rosetta/help/manual.pdf

[32] B. Olshausen, *SPARSENET: Sparse coding simulation software*, Available: http://redwood.ucdavis.edu/bruno/sparsenet.html