Evolutionary Algorithms and Rough Sets-based Hybrid Approach to Classificatory Decomposition of Cortical Evoked Potentials

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

[1] Computational Intelligence Laboratory,
Department of Electrical and Computer Engineering, University of Louisville,
Louisville, KY 40292, USA
{tomasz.smolinski, grzegorz.boratyn, jacek.zurada}@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 decomposition and classification of rat's cortical evoked potentials (EPs). The decomposition is based on learning of a sparse set of basis functions using Evolutionary Algorithms (EAs). The basis functions are generated in a potentially overcomplete dictionary of the EP components according to a probabilistic model of the data. Compared to the traditional, statistical signal decomposition techniques, this allows for a number of basis functions greater than the dimensionality of the input signals, which can be of a great advantage. However, there arises an issue of selecting the most significant components from the possibly overcomplete collection. This is especially important in classification problems performed on the decomposed representation of the data, where only those components that provide a substantial discernibility between EPs of different groups are relevant. In this paper, we propose an approach based on the Rough Set theory's (RS) feature selection mechanisms to deal with this problem. We design an EA and RS-based hybrid system capable of signal decomposition and, based on a reduced component set, signal classification.

1 Introduction

Signal decomposition plays a very important role in the analysis of Evoked Potentials (EPs) [1]. Among the most popular methods for EP decomposition, one will find Principal Component Analysis (PCA) [2], Independent Component Analysis (ICA) [3], [4] or wavelet-based analysis [5]. In general, a common way to represent real-valued EPs can be based upon a linear superposition of some basis functions (i.e. components). 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 characteristics of the basis functions they employ [5], [6].

An alternative and more general method of signal representation via transformation uses Sparse Coding with Overcomplete Bases (SCOB) [7], [8]. This methodology is based on the assumption that data can be represented by a set of statistically independent events (i.e. basis functions). An additional conjecture requires the probability distributions of those events to be sparse, meaning that the data can be usually described in terms of a relatively small number of those events. At the same time, an overcomplete representation allows for a greater number of basis functions than the dimensionality of the input signals, which provides much greater flexibility in terms of capturing structures hidden in data [9], [10], [11]. The SCOB methodology, due to the employment of an overcomplete representation, provides a powerful mechanism for a detailed data modeling. However, even if the sparseness of the basis functions is accounted for and preserved, the issue of selecting the most significant components from the possibly overcomplete collection is still crucial. This is especially important for some signal classification applications that can use SCOB as a data preprocessing/transformation tool. In such applications, only those components that provide the best discernibility between signals that belong to different groups or classes, are relevant.

While a similar idea of dimensionality reduction via a two-stage feature selection has already been proposed by Swiniarski in the hybridization of PCA and Rough Sets (RS) [12], it appears that the application of this approach to sparse coding with overcomplete bases is quite unique. Subsequently, we propose an algorithm for learning a potentially overcomplete basis of the EP components by viewing it as a probabilistic model of the observed data. From this model, we derive a simple and robust learning algorithm by maximizing the data likelihood over the basis.

2 Bayesian Motivated Model

The primary step in measuring the form of EPs is to decompose them into parts (i.e. components). Components can be expressed by some basis functions weighted by coefficients. Therefore, we assume that each data vector $\mathbf{x}$ can be described by a set of basis functions $\mathbf{M}$ and coefficients $\mathbf{a}$, plus some additive noise ε:

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

The unknown parameters to be estimated are $\mathbf{a}$ and $\mathbf{M}$. Developing efficient algorithms to solve this equation is an active research area. A given data point can have many possible representations. Nevertheless, this ambiguity can be removed by a proper choice of the prior probability of the basis coefficients, $P(\mathbf{a})$, which specifies the probability of the alternative representations. Standard approaches to signal representation do not specify the prior for the coefficients. A more general approach is to use the information theory and the probabilistic formulation of the problem [13], [14]. Rather than making prior assumption about the shape or form of the basis functions, those functions are adapted to the data using an algorithm that maximizes the log-probability of the data under the model.

The coefficients $\mathbf{a}$ from (1) can be inferred from $\mathbf{x}$ by maximizing the conditional probability of $\mathbf{a}$, given $\mathbf{x}$ and $\mathbf{M}$, which can be expressed via Bayes' rule as:

$$\mathbf{a} = \arg\max_{\mathbf{a}} P(\mathbf{a}|\mathbf{x}, \mathbf{M}) \propto \arg\max_{\mathbf{a}} P(\mathbf{x}|\mathbf{a}, \mathbf{M}) P(\mathbf{a}). \tag{2}$$

The first term of the right hand side of the proportion specifies the likelihood of the signal under the model for a given state of the coefficients:

$$P(\mathbf{x}|\mathbf{a}, \mathbf{M}) \propto \exp\left(-\frac{\lambda}{Z_{\sigma N}} |\mathbf{x} - \mathbf{Ma}|^2\right), \tag{3}$$

where $Z_{\sigma N}$ is normalizing constant, $\lambda = 1/\sigma^2$, and σ is the standard deviation of the additive noise. The second term specifies the prior probability distribution over the basis coefficients, where:

$$P(\mathbf{a}) = \prod_j \exp\left(-S(a_j)\right), \tag{4}$$

where a_j is the coefficient of the j-th basis function and $S(a_j)$ is a sparseness term given by $\beta \log(1 + (a_j/\gamma)^2)$, where β and γ are scaling factors. This sparse coding constraint encourages the model to use relatively few basis functions to represent the input signal. This leads to approximate redundancy reduction [15].

Thus, the maximization of the log-probability in (2) becomes:

$$\mathbf{a} = \arg\min_{\mathbf{a}} \left(\frac{\lambda_N}{2} |\mathbf{x} - \mathbf{Ma}|^2 + \sum_j S(a_j)\right). \tag{5}$$

3 Evolutionary algorithm for proposed sparse coding

From the model presented in Sect. 2, we derive a simple and robust learning algorithm by maximizing the data likelihood over the basis functions.

Some research has been previously done in applying Genetic Algorithms (GAs) to the blind source separation (BSS) and ICA [16]. In our work, an Evolutionary Algorithm (EA) is used to solve the problem of finding the best representation of a given signal in terms of basis functions and coefficients. The EA searches for an optimum by iteratively changing a population of temporary solutions encoded into chromosomes [17]. Each chromosome represents the matrix of basis functions $\mathbf{M}$ and the matrix of coefficients $\mathbf{a}$. Fitness function, minimized in our case, is based on (5) and consists of two parts: 1) the error of the reconstructed signals and 2) the sparse cost of the values of the coefficients:

$$f = \sum_i \left(\sum_t \left| x_i(t) - \sum_j a_{ij} M_j(t) \right| + \sum_j S(a_{ij}) \right) , \tag{6}$$

where $x_i(t)$ is the value of the i-th input signal at time t, $M_j(t)$ is the value of the j-th basis function at time t, and a_{ij} is the value of the coefficient for the j-th basis function for the i-th input signal.

4 Rough sets-based selection of classification-relevant components from a potentially overcomplete set of basis functions

The SCOB methodology provides a very efficient mechanism for data transformation. The fact that the collection of basis functions is potentially overcomplete allows for a very detailed and accurate modeling. On the other hand, this can cause a given problem to become more difficult to analyze, due to the increase of the conceptual dimensionality of the task. 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 [18]. Sometimes, however, the reconstruction error is not important, while the 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), in which one is looking for the smallest possible set of components that explain all the variations between different classes (i.e. groups) of objects. In terms of evoked potentials, for instance, traditional approaches do not guarantee that selected components, as a feature vector in the new representation, will be competent for classification.

One possibility for dealing with this problem, is to apply the theory of rough sets [19], [20]. In this case, especially useful will be the concept of reducts, inherently embedded in the theory. Intuitively, an application of the SCOB methodology will yield an adequate and detailed model of the input data, whilst the RS-based search for reducts will determine the most significant components in that model, in terms of data classification.

Obviously, since the RS theory operates on integer-valued data by principle, the real values of coefficients representing the signals need to be first discretized (i.e. divided into intervals that will be assigned ordered, integer values) [21], [22].

5 Experiments and results

5.1 Data

In the neuro-physiological experiments underlying our project, a piezoelectric stimulator was attached to a vibrissa of a rat. 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. 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, a series of additional stimuli was applied to the surface of the cortex – cooling events allowing to temporarily "switch off" some structures of the brain. The 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 (for a detailed description of the study, see [23], [24]).

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. Each EP was then sampled and is represented 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 was acknowledged as the most "representative" perspective at the activity of the cortex, it was usually chosen as the input to our experiments.

5.2 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 system was capable of determining components similar to the ones obtained in previous work by PCA [24] and ICA [6]. Furthermore, it was crucial to explore the ability of the system to automatically select 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, especially between *normal* and *cooled* potentials.

In all the experiments described in this section, an evolutionary algorithm implemented by the authors of this article was used for the signal decomposition (see also [6]). Additionally, the Rosetta system [25] along with some authors' implementations of rough sets were employed for the RS-based value discretization and feature selection/reduction.

The complete set of 882 evoked potentials, registered on the 3^{rd} channel, was used as the input to the evolutionary algorithm. 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 [6]) the goal of the algorithm was to determine a set of 10 basis functions (note: not really overcomplete in this case). A graphical representation of the discovered 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.

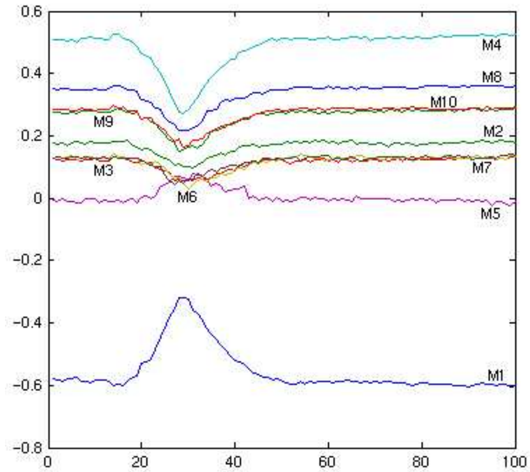

Fig. 1. 10 basis functions computed from the complete data set (Mx denotes the x-th basis function).

Based on this new representation of the input data (i.e. coefficients for the basis functions), RS-driven search for reducts was applied – after the prior discretization, the Johnson's reduction algorithm [25] was launched.

Various configurations of the discretization and/or reduction algorithms were investigated. The most interesting results are shown in Fig. 2, where the averaged selected components of the signals registered on the 3^{rd} channel (i.e. basis functions weighted by the coefficients of the signals) are shown.

Since the decision 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 two basis functions determined by the evolutionary algorithm, were extremely similar to the first two components received with both, PCA and ICA (see [24], [6]), and those two basis functions were always selected by the reduction algorithms. Additionally, as it can be clearly seen in Fig. 2, the system, after the signal decomposition, pointed out several other important components that provide an ability to discern between the EPs in the database (guaranteed by the reduction algorithm – indiscernibility relation holds). Additionally, the algorithm determined some clear differences between two main classes of the analyzed evoked potentials – *normal* vs. *cooled*, which was the main goal of the neuro-physiological experiments underlying our project.

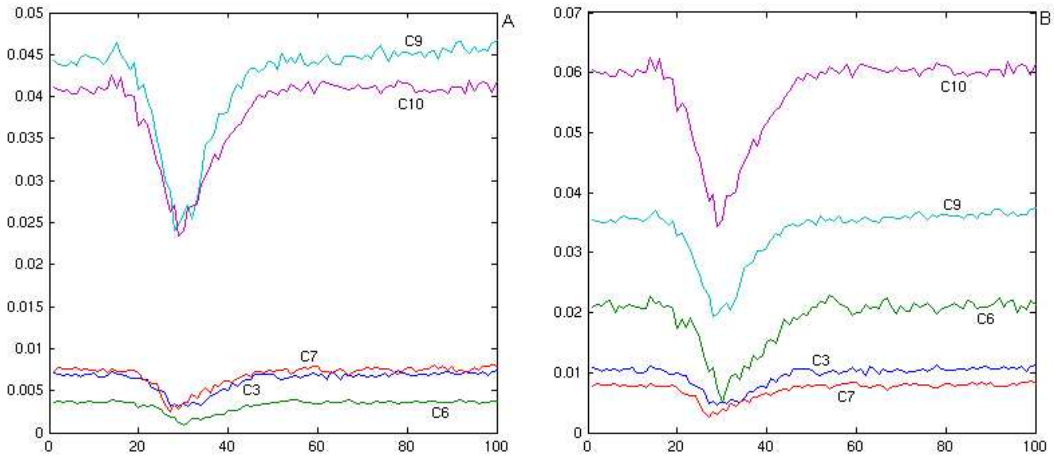

Fig. 2. Comparison of the reduced averaged components between *normal* (A) and *cooled* (B) 3^{rd} channel (Cx denotes the averaged x-th component). Discretization method: *Equal Frequency Bin*. Reduction method: *Johnson's Algorithm* (reduct: *[C3, C6, C7, C9, C10]*).

6 Conclusions

On the basis of the experiments described above, we can conclude that the proposed EA and RS-based hybrid system provides a useful and effective tool in terms of EP decomposition and classification. Our results, obtained via the SCOB methodology, were coherent with previous work in terms of the signal's main components, which suggests that this approach delivers comparable capabilities in terms of signal decomposition. On the other hand, the system provides a significant extension to the traditional approaches thanks to the potentially overcomplete representation of the input data as well as the mechanisms for an automatic determination of relevant components, in terms of signal classification.

References

1. Freeman, W. J.: Measurement of Cortical Evoked Potentials by Decomposition of their Wave Forms. J. of Cybernetics and Information Science, **2-4** (1979) 22–56
2. Chapman, R.M. and McCrary, J.W.: EP Component Identification and Measurement by Principal Components Analysis. Brain and Cognition. **27** (1995) 288–301
3. Makeig, S., Bell, A. J., Jung, T.-P., and Sejnowski, T. J.: Independent Component Analysis of Electroencephalographic Data. Advances in Neural Information Processing Systems **8** (1996) 145–151
4. Makeig, S., Jung, T.-P., Bell, A. J., Ghahremani, D., and Sejnowski, T. J.: Blind Separation of Auditory Event-related Brain Responses into Independent Components. Proc. Natl. Acad. Sci. USA **94** (1997) 10979–10984

5. Raz, J., Dickerson, L., and Turetsky, B.: A Wavelet Packet Model of Evoked Potentials. Brain and Language **66** (1999) 61–88
6. Milanova, M., Smolinski, T. G., Boratyn, G. M., Zurada, J. M., and Wrobel, A.: 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), Niagara Falls, Canada, August 10 (2002)
7. Olshausen, B. and Field, D. J.: Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1? Vision Research **37(23)** (1997) 3311–3325
8. Olshausen, B.: Sparse Codes and Spikes. In: Rao, R. P. N., Olshausen, B. A., Lewicki, M. S. (eds.): Probabilistic Models of Perception and Brain Function. MIT Press, Cambridge, MA (2001) 245–260
9. Chen, S., Donoho, D. L., and Saunders, M. A.: Atomic Decomposition by Basis Pursuit. Technical report, Dept. Stat., Stanford University (1996)
10. Mallat, S. G. and Zhang, Z.: Matching Pursuits with Time-Frequency Dictionaries. IEEE Trans. on Signal Processing **41(12)** (1993) 3397–3415
11. Lewicki, M. and Sejnowski, T.: Learning Overcomplete Representations. Neural Computation **12** (2000) 337–365
12. Swiniarski, R.: Rough Sets and Principal Component Analysis and Their Applications in Data Model Building and Classification. In: Pal, S. K. and Skowron, A. (eds.): Rough Fuzzy Hybridization: A New Trend in Decision-Making. Springer-Verlag, Singapore (1999) 275–300
13. Vapnik, V.: The Nature of Statistical Learning Theory. Springer-Verlag, New York (1995)
14. Lewicki, M. S. and Olshausen, B. A.: Probabilistic Framework for Adaptation and Comparison of Image Codes. J. Opt. Soc. of Am., **16** (1999) 1587–1600
15. Field, D. J.: What is the Goal of Sensory Coding? Neural Computation **6** (1994) 559–601
16. Yoshioka, M. and Omatu, S.: Independent Component Analysis Using Time Delayed Sampling. IEEE International Joint Conference on Neural Networks, Como, Italy, July 24-27 (2000)
17. Michalewicz, Z.: Genetic Algorithms + Data Structures = Evolution Programs. Springer-Verlag, New York (1996)
18. Jolliffe, I. T.: Principal Component Analysis. Springer-Verlag, New York (1986)
19. Pawlak, Z.: Rough Sets. International J. of Computer and Information Sciences **11** (1982) 341–356
20. Marek, W. and Pawlak, Z.: Rough Sets and Information Systems. Fundamenta Matematicae **17** (1984) 105–115
21. Nguyen, S. H. and Nguyen, H. S.: Discretization Methods in Data Mining. In: Polkowski, L. and Skowron A. (eds.): Rough Sets in Knowledge Discovery. Physica-Verlag, Heidelberg (1998) 451–482
22. Nguyen, S. H.: Discretization Problems for Rough Set Methods. Proc. of the First International Conference on Rough Sets and Current Trend in Computing (RSCTC'98), June 1998, Warsaw, Poland (1998) 545–552.
23. Wrobel, A., Kublik, E., and Musial, P.: Gating of the Sensory Activity within Barrel Cortex of the Awake Rat. Exp. Brain Res. **123** (1998) 117–123
24. Kublik, E., Musial, P., and Wrobel, A.: Identification of Principal Components in Cortical Evoked Potentials by Brief Surface Cooling. Clinical Neurophysiology **112** (2001) 1720–1725
25. ROSETTA: A Rough Set Toolkit for Analysis of Data.
Available: http://www.idi.ntnu.no/∼aleks/rosetta/