Multi-Objective Evolutionary Algorithms and Rough Sets for Decomposition and Analysis of Cortical Evoked Potentials

Tomasz G. Smolinski, *Member, IEEE*, Mariofanna Milanova, *Member, IEEE*,
Grzegorz M. Boratyn, *Member, IEEE*, Roger Buchanan, and Astrid A. Prinz

Abstract—Signal decomposition techniques prove to be useful in the analysis of neural activity, as they allow for identification of supposedly distinct neuronal structures (*i.e.*, sources of activity). Applied to measurements of brain activity in a controlled setting as well as under exposure to an external stimulus, they allow for analysis of the impact of the stimulus on those structures. The link between the stimulus and a given source can be confirmed by a classifier that is able to “predict” if a given signal was registered under one or the other condition, solely based on the components. Very often, however, statistical criteria used in traditional decomposition techniques turn out to be insufficient to build an accurate classifier. Therefore, we propose to utilize a novel hybrid technique based on multi-objective evolutionary algorithms (MOEA) and rough sets (RS) that will perform decomposition in the light of the classification problem itself.

Index Terms—Signal decomposition, classification, multi-objective evolutionary algorithms, rough sets.

I. INTRODUCTION

SIGNALS recorded from the surface of the cerebral cortex are composites of the electrical activity of a large number of individual cells. Because the “average” activity of a large number of cells is being recorded, one would expect that several different processes, each with a characteristic activity pattern, would be occurring simultaneously. In order to separate those superimposed signal patterns, we propose to utilize an experimental technique based on measuring neural activity in a controlled setting (normal) as well as under exposure to some external stimulus (nicotine, in our case). Comparison of the measured signals in those two situations can provide interesting insight into the problem of neural activity analysis.

The signals that are observed and measured with physical devices are often not produced by just one source, but constitute a mixture of a number of signals produced by many different sources (*e.g.*, specific neuronal structures). Application of stimuli that affect the observed signals often has an effect only on a subset of the sources. Which sources are affected by the stimuli is interesting, but cannot be measured directly. Based on the assumption that each of the sources produces a

signal with different statistical properties, the observed signals can be decomposed into constituents that model the source signals. Each of the observed signals is a linear mixture of those discovered components. However, each component can have a different magnitude in each observed signal (due to the fact that some sources can have stronger influence in some locations than others). The source magnitudes are modeled as coefficients in the aforementioned linear combination. The change in the coefficients, as a result of applied stimuli, corresponds to the change in generation of the signal by the corresponding sources (also called *basis functions*).

Signal decomposition techniques can be useful in this kind of analysis, as they allow for determination of an impact of the external stimuli on some specific neuronal structures, supposedly represented by the discovered components. The link between the stimulus and a given source can be confirmed by a classifier that is able to “predict” under which condition a given signal was registered, solely based on the components. However, statistical criteria used in traditional decomposition techniques, like Principal Component Analysis (PCA) or Independent Component Analysis (ICA), often turn out to be insufficient to build an accurate classifier. Therefore, we propose a novel hybrid technique based on multi-objective evolutionary algorithms (MOEA) and rough sets (RS) that will perform decomposition in the light of the classification problem itself: we will be looking for basis functions whose coefficients allow for an accurate classification.

II. CLASSIFICATORY DECOMPOSITION

The main concept of classificatory decomposition was motivated by the hybridization of evolutionary algorithms (EAs) with sparse coding with overcomplete bases (SCOB) introduced in [5]. Using this approach, the basis functions as well as the coefficients are being evolved by optimization of a fitness function that minimizes the reconstruction error and at the same time maximizes the sparseness of the basis function coding. This methodology produces a set of basis functions and a set of sparse (*i.e.*, “as few as possible”) coefficients. This may significantly reduce dimensionality of a given problem but, as any other traditional decomposition technique, does not assure the classificatory usefulness of the resultant model.

In the first attempt to incorporate classification accuracy into signal decomposition, a rough sets-based search for reducts was applied in post-processing of the discovered components

T. Smolinski and A. Prinz are with Emory University, Atlanta, GA 30322, USA (e-mail: {tomasz.smolinski, astrid.prinz}@emory.edu)

M. Milanova is with University of Arkansas, Little Rock, AR 72204, USA (e-mail: mgmilanova@ualr.edu)

G. Boratyn is with University of Louisville, Louisville, KY 40292, USA (e-mail: greg.boratyn@louisville.edu)

R. Buchanan is with Arkansas State University, State University, AR 72467, USA (e-mail: rbuck@astate.edu)

[8]. This 2-stage approach allowed for selection of the most important components in the light of the classification task dealt with (a 2-category classification problem of evoked potentials), but did not escape the limitations of the decomposition process itself.

In the 1-stage approach proposed here, the sparseness term is replaced by a rough sets-derived data reduction-driven classification accuracy measure. This should assure that the result will be both “valid” (*i.e.*, via the reconstruction constraint) and useful for the classification task. Furthermore, since the classification-related constituent also searches for a reduct, the classification is done with as few as possible basis functions. Finally, the single-objective EA utilized in the aforementioned technique is replaced by a multi-objective approach, in which the EA deals with the reconstruction error and classification accuracy, both at the same time.

A. Utilized Aspects of the Theory of Rough Sets

The theory of rough sets (RS) deals with the classificatory analysis of data tables [6]. The main idea behind it is the so-called indiscernibility relation that describes objects indistinguishable from one another.

Let $IS = (U, A)$ be an information system, where U is a non-empty finite set of objects (universe) and A is a non-empty finite set of attributes. With any $B \subseteq A$ there is associated an equivalence relation $IND_{IS}(B)$:

$$IND_{IS}(B) = \{(x, x') \in U^2 \mid \forall a \in B, a(x) = a(x')\}. \quad (1)$$

This **indiscernibility relation** induces a partitioning of the universe, by dividing it into disjoint equivalence classes, denoted as $[x]_B$. These partitions can be used to build new subsets of the universe. Subsets that are most often of interest are those that contain objects that belong to the same decision class. It may happen, however, that a concept cannot be defined in a crisp manner. The main goal of rough set analysis is to synthesize approximations of concepts from acquired data. The concepts are represented by lower and upper approximations. Although it may be impossible to precisely define some concept X , we can approximate it using the information contained in B by constructing the **B -lower** and **B -upper approximations** of X , denoted by $\underline{B}X$ and $\overline{B}X$ respectively, where, $\underline{B}X = \{x \mid [x]_B \subseteq X\}$ and $\overline{B}X = \{x \mid [x]_B \cap X \neq \emptyset\}$. Only the objects in $\underline{B}X$ can be with certainty classified as members of X , based on the knowledge conveyed by B .

A rough set can be characterized numerically by the so-called **quality of classification**:

$$\gamma_B(X) = \frac{card \underline{B}X \cup \overline{B}X}{card U}, \quad (2)$$

where $\underline{B}X$ is the lower approximation of X , $\overline{B}X$ is the lower approximation of the set of objects that do not belong to X , and U is the set of all objects.

All these considerations represent one way of reducing data by identifying equivalence classes, *i.e.*, objects that are indistinguishable using the available attributes. That is a much more efficient representation, since only one element of the

equivalence class is needed to characterize the entire class. The other consideration in terms of data reduction is to keep only those attributes that preserve the indiscernibility relation and, consequently, the set approximation. The rejected attributes are redundant since their removal cannot worsen the classification.

There are usually several such subsets of attributes and those that are minimal are called **reducts**. The following presents a formal definition of a reduct: a “minimal” $R \subseteq A$, such that:

$$\forall k, n \forall a_i \in R : ((a_i(o_k) = a_i(o_n)) \Rightarrow (d(o_k) = d(o_n))), \quad (3)$$

where $k, n = 1..N$ (N is the number of objects in the decision table), $o_{k|n}$ is the $k^{th}|n^{th}$ object, $d(o_{k|n})$ is the value of the decision attribute for the $k^{th}|n^{th}$ object, and $i = 1..M$ (M is the number of conditional attributes). The reduct is “minimal” if: $\neg \exists P \subset R : (3)$ is satisfied for P .

Computing equivalence classes is straightforward. Finding a global minimal reduct, on the other hand (*i.e.*, reduct with a minimal cardinality among all reducts), is NP-hard. There are many heuristics (including utilization of genetic algorithms) designed to deal with this problem. For more details and examples of reduct-finding algorithms see *e.g.*, [9].

B. Utilized Aspects of the Multi-Objective Optimization Using Evolutionary Algorithms

Since the approach proposed here is based upon finding a solution satisfying two potentially conflicting goals (*i.e.*, component-based reconstruction accuracy vs. classification accuracy), EAs for multi-objective optimization were used throughout the performed experiments. For a detailed introduction to multi-objective optimization using EAs see [1].

Based on the results of the study performed by Zitzler, Deb and Thiele on evaluation and comparison of multi-objective evolutionary algorithms [10] and simplicity of its implementation, the well-known Vector Evaluated Genetic Algorithm (VEGA) was chosen for the initial experiments in this research study. The main idea in VEGA is to randomly divide the population, in each generation, into equal subpopulations. Each subpopulation is assigned fitness based on a different objective function (the appropriate fitness functions will be presented below). Then, the crossover between the subpopulations is performed as with traditional EAs, with an introduction of random mutations.

C. Chromosome Coding

Each chromosome is a complete solution to a given classificatory decomposition problem and provides a description of both the set of basis functions and the coefficients for all the signals in the training data set. For example, if there are N signals with n samples each, and the task is to find M basis functions, the chromosome will be coded in the following way:

Fig. 1. Chromosome coding in classificatory decomposition

Each of the M basis functions has the length of the original input signal (*i.e.*, n), and there are N vectors of coefficients (*i.e.*, each vector corresponds to one signal in the training set) of dimensionality equal to the number of basis functions (*i.e.*, each coefficient corresponds to one basis function).

D. Fitness Evaluation

1) *Reconstruction error*: The problem of minimization of the reconstruction error is a relatively simple task. Once a particular distance measure has been decided upon, virtually any optimization algorithm can be used to minimize the distance between the original signal and the reconstructed one. The measure employed in this project is the well known 2-norm [3], referred to in signal processing as the signal energy-based measure. In order to deal with raw signals which can be large (thus causing the energy-based distance measure to be large as well), a simple normalization of the energy-based measure by the energy of the original signal is proposed:

$$D_{NORM} = \frac{\sum_{t=1}^n (\mathbf{x}_t - (\mathbf{M}\mathbf{a})_t)^2}{\sum_{t=1}^n (\mathbf{x}_t)^2}, \quad (4)$$

where $\mathbf{x}$ represents the original signal, $\mathbf{M}$ is the matrix of basis functions, $\mathbf{a}$ is a set of coefficients, and $t = 1..n$ where n is the number of samples in the signal.

Subsequently, the reconstruction error fitness function f_{REC} for a representative p takes the following form:

$$f_{REC}(p) = \frac{\sum_{i=1}^N D_{NORM}^i}{N}, \quad (5)$$

where D_{NORM}^i is the normalized reconstruction error for the i^{th} signal and N is the total number of the input signals.

2) *Classification accuracy and reduction in the number of coefficients and basis functions*: The problem of maximizing the classificatory competency of the decomposition scheme, and at the same time reducing the number of computed basis functions, can be dealt with by the application of rough sets. In this project, the rough sets-based *quality of classification*, as introduced in (2), is used for the purpose of estimating the classificatory aptitude.

The quality of classification is estimated directly on the candidate reduct, which can be computed by any of the existing algorithms/heuristics. Note that the main objective that deals with the classificatory capability of decomposition can actually be considered a bi-objective optimization problem itself. On one hand, we are looking for the best possible classification accuracy, but on the other, we want to use as few basis functions as possible. However, based on previous applications of EAs in the search for reducts, as described in [9], it is easy to see that this problem can be dealt with by minimizing a single-objective fitness function that is simply a summation of the classification error and the relative length of the reduct, as shown in (6).

$$f_{CLASS}(p) = (1 - \gamma_R) + \frac{L(R)}{M}, \quad (6)$$

where p is a given representative (*i.e.*, chromosome), $L(R)$ is the length of the potential reduct R (*i.e.*, the number of

attributes used in the representative), normalized by the total number of conditional attributes M , and γ_R is the *quality of classification* coefficient for the reduct R .

An interesting question here is what to do with the coefficients (and the corresponding basis functions) that are not a part of the reduct. Since we are looking for the best possible classification accuracy, while using as few basis functions as possible, some mechanism capable of emphasizing the “important” coefficients/basis functions would be advisable. A solution to this problem is possible due to the application of the “hard” fitness computation idea, which allows the fitness function itself to introduce changes directly to the genetic material of the evaluated chromosome [7]. In this paper we propose to utilize a coefficients/basis functions annihilation approach, which simply zeroes-out the “not important” genetic material. The idea here is that if we remove the basis functions that are not vital in the classification process, the EA will improve the remaining basis functions in order to compensate for an increase in the reconstruction error.

III. EXPERIMENTAL DATA

The dataset used in this study was derived from neurophysiological experiments performed at Arkansas State University. In the experiments, recordings were collected from the vertex (cerebral cortex) of two rats. One of the animals had been exposed to the cigarette smoke in utero (*i.e.*, mother of the animal was exposed to cigarette smoke during pregnancy), while the other had not. The recordings were in the form of evoked potentials (EPs) of a duration of 1 second triggered by an auditory stimulus (for more details on the experiments see [4]). The research problem here is to investigate and understand how treatments (like nicotine) could alter responses to discrete stimuli. 10 signals were registered for the unexposed animal and 9 for the exposed one. The EPs were sampled at the rate of 7 kHz. The signals (smoothed with a mean filter) for the unexposed and exposed rats are shown in Fig. 2.

Fig. 2. Input EPs for the unexposed (red solid line) and exposed (green dashed line) animal.

As can be seen in this figure, differentiation between signals from exposed and unexposed animals is not a trivial task.

Fig. 3. An exemplary set of discovered basis functions.

IV. ANALYSIS

In order to investigate the feasibility of the proposed approach, a number of MOEAs was launched simultaneously on a computer cluster of 40 processors. The number of maximum possible generations was set to 100 while the size of the population was varied between 30 and 50. Mutation probability was initialized with a small random value and was being adapted along the evolution process (*i.e.*, increased if no progress in the fitness functions was observed and reset when there was an improvement). Crossover probability was randomly determined in each generation (between 0% and 100%). Single-point crossover was utilized.

In most cases, the classification accuracy of the generated model was high (over 80%). The maximum number of basis functions to be discovered was set to 5, 10, or 20 and the determined number of the basis functions required to preserve that precision (driven by the search for reduct) oscillated around 4 (other basis functions and the corresponding coefficients were removed by the annihilation algorithm). Unfortunately, the average reconstruction error tended to be rather high (even over 0.5). An exemplary set of four discovered basis functions is shown in Fig. 3 and the corresponding coefficients in Fig. 4. It is important to point out that in order to apply rough sets on the set of the coefficients, discretization had to be performed first. As a result, the values of the coefficients are limited.

Even though an accurate analysis of the discovered model is obscured by its high reconstruction error, an interesting observation can be derived from the presented coefficients, namely the fact that the basis function M4 is only present in the recordings from the unexposed animal. It is easy to see that this is the basis function with the highest amplitude of the negative waveform in the beginning of the signals. This may indicate that the source of activity corresponding to that waveform is impaired by nicotine exposure. Further analysis of such phenomena can provide an interesting insight into the mechanisms behind the influence of nicotine on cortical neural activity.

Fig. 4. Corresponding coefficients. Horizontal axis: indexes of the registered EPs. Vertical axis: discretized values of the coefficients for the basis functions.

V. CONCLUSION

This article presented a general framework for the methodology of classificatory decomposition of signals based on hybridization of multi-objective evolutionary algorithms and rough sets. The methodology combines the agility of rough sets in their accuracy and efficiency for solving data classification problems with the signal decomposition techniques' ability to explain the hidden behavior of the modeled system. In the described analysis, the apparent disproportion in difficulty between the two objectives in the utilized MOEA (*i.e.*, finding a reliable classifier for a 2-class problem seems much easier than reconstructing 19 signals with 7,000 samples each), caused the reconstruction error to remain high. This issue will be addressed in future research by utilizing other MOEAs.

ACKNOWLEDGMENT

The Rough Set Library [2] was used in parts of this project.

REFERENCES

- [1] K. Deb, *Multi-objective optimization using evolutionary algorithms*. Wiley, 2001.
- [2] M. Gawryś and J. Sienkiewicz, "RSL - The Rough Set Library version 2.0," Warsaw University of Technology, Poland, Tech. Rep., May 1994.
- [3] E. Kreyszig, *Introductory functional analysis with applications*. Wiley, 1978.
- [4] N. Mamiya *et al.*, "Nicotine suppresses the p13 auditory evoked potential by acting on the pedunculopontine nucleus in the rat," *Exp Brain Res*, vol. 164, pp. 109–119, 2005.
- [5] M. Milanova *et al.*, "Correlation kernel analysis and evolutionary algorithm-based modeling of the sensory activity within the rat's barrel cortex," *Lecture Notes in Computer Science*, pp. 198–212, 2002.
- [6] Z. Pawlak, *Rough sets - Theoretical aspects of reasoning about data*. Kluwer, 1991.
- [7] T. G. Smolinski, "Classificatory decomposition for time series classification and clustering," Ph.D. dissertation, Univ. of Louisville, KY, 2004.
- [8] T. G. Smolinski *et al.*, "Evolutionary algorithms and rough sets-based hybrid approach to classificatory decomposition of cortical evoked potentials," *Lecture Notes in Artificial Intelligence*, pp. 621–628, 2002.
- [9] J. Wróblewski, "Finding minimal reducts using genetic algorithms," in *Proc. 2nd Annual Joint Conference on Information Sciences*, Wrightsville Beach, NC, 1995, pp. 186–189.
- [10] E. Zitzler, K. Deb, and L. Thiele, "Comparison of multiobjective evolutionary algorithms: Empirical results," *Evolutionary Computation J.*, vol. 8, no. 2, pp. 125–148, 2000.