Hybridization of Independent Component Analysis, Rough Sets, and Multi-Objective Evolutionary Algorithms for Classificatory Decomposition of Cortical Evoked Potentials

Tomasz G. Smolinski[1], Grzegorz M. Boratyn[2], Mariofanna Milanova[3],
Roger Buchanan[4], and Astrid A. Prinz[1]

[1] Department of Biology, Emory University, Atlanta, GA 30322
{tomasz.smolinski, astrid.prinz}@emory.edu
[2] Kidney Disease Program, University of Louisville, Louisville, KY 40292
greg.boratyn@louisville.edu
[3] Department of Computer Science, University of Arkansas, Little Rock, AR 72204
mgmilanova@ualr.edu
[4] Department of Biology, Arkansas State University, Jonesboro, AR 72467
rbuck@astate.edu

Abstract. This article presents a continuation of our research aiming at improving the effectiveness of signal decomposition algorithms by providing them with "classification-awareness." We investigate hybridization of multi-objective evolutionary algorithms (MOEA) and rough sets (RS) to perform the task of decomposition in the light of the underlying classification problem itself. In this part of the study, we also investigate the idea of utilizing the Independent Component Analysis (ICA) to initialize the population in the MOEA.

1 Introduction

The signals recorded from the surface of the cerebral cortex are composites of the electrical activity of a large number of individual cells and often constitute a mixture of a group of signals produced by many different sources (*e.g.*, specific neuronal structures). In order to separate those superimposed signal patterns and analyze them independently, 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 this case) [1]. Application of stimuli that affect the observed signals often has an effect only on a subset of the sources. The information about which sources are affected by the stimuli can provide interesting insight into the problem of neural activity analysis, but cannot be measured directly. Based on the assumption that each of the sources produces a signal that is statistically independent on the others, the observed signals can be decomposed into constituents that model the sources. Those modeled sources are referred to as basis functions. Each of the observed signals is a linear combination of the basis functions. Due to the fact that some sources can have stronger influence in some locations than others,

J.C. Rajapakse, L. Wong, and R. Acharya (Eds.): PRIB 2006, LNBI 4146, pp. 174–183, 2006.

each source can be present in each observed signal with a different magnitude. 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 the contribution of a source in generation of a given signal.

Independent Component Analysis (ICA) can be useful in this kind of analysis, as it allows 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 verified by designing a classifier that is able to "predict" under which condition a given signal was registered, solely based on the discovered independent components. However, statistical criteria used in ICA, often turn out to be insufficient to build an accurate coefficients-based classifier.

Classificatory decomposition is a general term that describes our research study that attempts to improve the effectiveness of signal decomposition techniques by providing them with "classification-awareness." The description of previous stages of the study and some examples of applications can be found in [2,3,4]. Currently, we are investigating a hybridization of multi-objective evolutionary algorithms (MOEA) and rough sets (RS) to perform decomposition in the light of the classification problem itself. The idea is to look for basis functions whose coefficients allow for an accurate classification while preserving the reconstruction.

In this article, we propose a simple extension of the well-known multi-objective evolutionary algorithm VEGA, which we call end-VEGA (elitist-non-dominated-VEGA). The extension supplies the algorithm with the considerations related to elitism and non-dominance, lack of which is known to be VEGA's main drawback. We also investigate the idea of utilizing the ICA to initialize the population in the MOEA. The details of the modifications as well as a short theoretical background are given below.

2 Theoretical Background

2.1 Independent Component Analysis

Independent Component Analysis (ICA) is a signal processing technique originally developed to deal with the cocktail-party problem [5]. ICA is perhaps the most widely used method in Blind Source Separation (BSS) in various implementations and practical applications [6]. The basic idea in ICA is to represent a set of random variables using basis functions (or sources), which are as much statistically independent as possible. The Central Limit Theorem states that the distribution of a sum of independent random variables, under certain conditions, tends toward a Gaussian distribution. Thus a sum of two independent random variables usually has a distribution that is closer to Gaussian than any of the two original random variables. Therefore, the key concept in ICA is based on maximization of non-Gaussianity of the sources. There are various quantitative measures of non-Gaussianity, one of the most popular one among which is kurtosis (*i.e.*, the fourthorder cumulant). One of the most popular ICA algorithms based on finding the local maximum of the absolute value of kurtosis is FastICA [7].

2.2 Multi-Objective Evolutionary Algorithms

Many decision making or design problems involve optimization of multiple, rather than single, objectives simultaneously. In the case of a single objective, the goal is to obtain the best global minimum or maximum (depending on the nature of the given optimization problem), while with multi-objective optimization, there usually does not exist a single solution that is optimal with respect to all objectives. Therefore, the goal of multi-objective optimization is to find a *set* of solutions such that no other solution in the search space is superior to them when *all* objectives are considered. This set is known as **Pareto-optimal** or **non-dominated set** [8].

Since evolutionary algorithms (EA) work with a population of individuals, a number of Pareto-optimal solutions can be found in a single run. Therefore, an application of EAs to multi-objective optimization seems natural. The first practical MOEA implementation was the Vector Evaluated Genetic Algorithm (VEGA) proposed in [9]. Although it opened a new avenue in multi-objective optimization research, the algorithm seemed to have some serious limitations, at least partially due to the lack of considerations of dominance and elitism [8]. To deal with the first of the above considerations, a non-dominated sorting procedure was suggested in [10] and various implementations based on that idea of rewarding non-dominated solutions followed [11]. Elitism, in other words the notion that "elite" individuals cannot be expelled from the active gene-pool by worse individuals, has recently been indicated to be a very important factor in MOEAs that can significantly improve their performance [12]. Both these aspects, while preserving the simplicity of implementation of the original VEGA, were taken into consideration in the design of the end-VEGA algorithm proposed here.

2.3 Rough Sets

The theory of rough sets (RS) deals with the classificatory analysis of data tables [13]. The main idea behind it is the so-called **indiscernibility relation** that describes objects indistinguishable from one another. The indiscernibility relation induces a split of the universe, by dividing it into disjoint equivalence classes, denoted as $[\mathrm{x}]_B$ (for some object x described by a set of attributes B). These classes can be used to build new partitions of the universe. Partitions 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 $\underline{B}X = \{\mathrm{x} \mid [\mathrm{x}]_B \subseteq X\}$ and $\overline{B}X = \{\mathrm{x} \mid [\mathrm{x}]_B \cap X \neq \emptyset\}$. Only the objects in $\underline{B}X$ can be with certainty classified as members of X, based on the knowledge in B.

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

$$\gamma_B(X) = \frac{card\underline{B}X \cup \underline{B}\neg X}{cardU}, \quad (1)$$

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

Another very important aspect of rough analysis is data reduction by means of keeping 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**. Finding a global minimal reduct (*i.e.*, reduct with a minimal cardinality among all reducts) is an NP-hard problem. However, there are many heuristics (including utilization of genetic algorithms [14]) designed to deal with this problem.

3 ICA, RS, and MOEA-Based Classificatory Decomposition

The main concept of classificatory decomposition was motivated by the hybridization of EAs with sparse coding with overcomplete bases (SCOB) introduced in [15]. 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 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 [4].

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), an application of MOEAs seems natural. In the experiments described here, we investigate a simple extension of VEGA, which supplies it with elitism and non-dominance, lack of which is known to be its main drawback. We call this extended algorithm end-VEGA (elitist-non-dominated-VEGA).

3.1 End-VEGA

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. Then, the crossover between the subpopulations is performed as with traditional EAs, with an introduction of random mutations.

As indicated earlier, VEGA has several quite significant limitations related to the lack of dominance and elitism. To address the former, we propose a simple approach based on multiplying the fitness of a given individual by the number of solutions that this individual is dominated by (plus 1 to ensure that the fitness function of a non-dominated solution is not multiplied by 0). Since the fitness function is being minimized in this project, the dominated solutions will be adequately penalized. To deal with the latter, we utilize the idea of an external sequential archive [12] to keep track of the best-so-far (*i.e.*, non-dominated) solutions and to make sure that their genetic material is in the active gene-pool.

3.2 Chromosome Coding

Each chromosome is a complete solution to a given classificatory decomposition task 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, for N signals with n samples each, and the task of finding M basis functions, the chromosome will be coded in the following way:

Basis functions part				Coefficients part			
$m_{11},\dots,m_{1n}$	$m_{21},\dots,m_{2n}$	$\dots$	$m_{M1},\dots,m_{Mn}$	$a_{11},a_{12},\dots,a_{1M}$	$a_{21},a_{22},\dots,a_{2M}$	$\dots$	$a_{N1},a_{N2},\dots,a_{NM}$

Fig. 1. Chromosome coding

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).

3.3 ICA-Based Population Initialization

The idea behind using ICA to initialize the starting population in end-VEGA is very simple: rather than beginning the search at a random location in the search space, the chromosomes are supplied with a starting point that is already quite satisfactory in one of the objectives (*i.e.*, reconstruction). Since based on previous results described in [4], the reconstruction accuracy objective seems to be much more difficult to optimize (especially for a small number of high-dimensional signals), this approach is plausible. Depending on the parameters of both ICA and end-VEGA, various variants of the initialization can be pursued.

If the number of independent components (IC) is higher than the number of the maximum allowed classificatory components (CC), then the subset of the ICs used to initialize the population is chosen randomly. In the opposite case, the "missing" initial CCs can assume random values, just as in the case without ICA.

3.4 Fitness Evaluation

Reconstruction error. The problem of minimization of the reconstruction error is intuitively simple. 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, referred to in signal processing as the signal energy-based measure [16]. 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 [3]:

$$D_{NORM} = \frac{\sum_{t=1}^{\mathrm{n}} (\mathbf{x}_t - (\mathbf{Ma})_t)^2}{\sum_{t=1}^{\mathrm{n}} (\mathbf{x}_t)^2}, \tag{2}$$

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

$$f_{REC}(p) = \frac{\sum_{i=1}^{\mathrm{N}} D_{NORM}^{i}}{N}, \tag{3}$$

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

Classification accuracy and reduction in the number of coefficients and basis functions. The problem of maximizing the classificatory competence 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 (1), 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 (in this project, algorithms from the Rough Set Library were utilized [17]). 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 [14], we decided to deal with it 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 (4).

$$f_{CLASS}(p) = \big(1 - \gamma_R\big) + \frac{L(R)}{\mathrm{M}}, \tag{4}$$

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 [3]. 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.

4 Experimental Data

The dataset used in this study was derived from neurophysiological experiments performed at Arkansas State University [1]. In the experiments, recordings in the form of evoked potentials (EP) of a duration of 1 second triggered by an auditory stimulus were collected from the 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 research problem here is to investigate 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 original signals for the unexposed and exposed rats are shown in Fig. 2.

5 Analysis

In the first step of the analysis described in this paper, the FastICA algorithm [18] was utilized to compute the ICs to be used in the initial population in the MOEA. The algorithm yielded 19 ICs along with the corresponding coefficients. As typical with ICA, the reconstruction was nearly perfect, but the coefficients were hardly useful for differentiation between the two EP classes (unexposed vs. exposed).

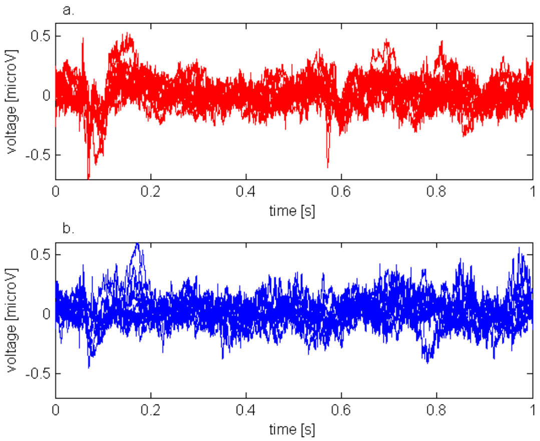

Fig. 2. Input EPs for the unexposed (a) and exposed (b) animal

In order to investigate the feasibility of the proposed approach, a number of MOEAs was launched simultaneously. The number of maximum possible generations was set to 200 (there was no significant improvement of convergence observed with a larger number of generations) while the size of the population was set to 30, 50, and 100. Mutation probability was initialized with a small random value and was being adapted along the evolution process (*i.e.*, increased by a fixed step in each generation 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.

Several variants of the ICA used to initialize the population in end-VEGA were tried. Both initialization of the full as well as a part of the population were simulated. In the first case, the changes in the basis functions can only be introduced by mutation, while in the second, some randomness is present from the beginning.

The maximum allowable number of basis functions was set to 5, 10, or 19. In the first two cases, a random subset of 5 or 10 ICs (out of all 19) was chosen for each chromosome, and in the second, a permutation of all 19 ICs was used.

In most cases, the classification accuracy was reasonably high (over 80%) and the problems appeared to be mostly related to 2 unexposed EPs being classified as exposed. The determined number of the basis functions required to preserve that precision (driven by the search for a reduct) oscillated around 4, 7, and 12, for the maximum allowable number of 5, 10, and 19 of the basis functions respectively.

The average reconstruction error was significantly improved compared to the previous study [4], especially in the case of the full set of the ICs being used to initialize the MOEA (note that this set was however reduced to about 12 ICs, thus indicating the ICs important for classification and at the same time “improving” them to account for the increase in the reconstruction error caused by removing the other 7 components). The modifications in end-VEGA, although

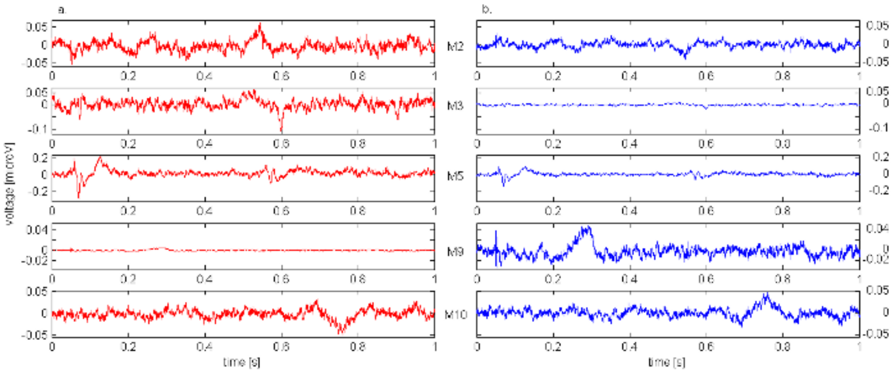

Fig. 3. Selected averaged components for the unexposed (a) and exposed (b) animal

improved the reconstruction slightly and sped up the overall convergence of the algorithm, worked much better in tandem with ICA. An exemplary set of components, averaged for the unexposed and exposed animal separately, for arbitrarily selected 5 basis functions, is shown in Fig. 3. The figure represents an average contribution of the basis functions in generation of the EPs in the unexposed and the exposed animal respectively.

Even a quick visual analysis of the figure reveals significant differences in how the sources are represented in the unexposed and the exposed rat. The dissimilarities can be simply expressed by amplitude variations (M3, M5, M9), or can be as major as the sign reversal (M2, M10). Further analysis of such phenomena can provide interesting insight into the mechanisms behind the influence of nicotine on the cortical neural activity.

6 Conclusions

This article presented a general framework for the methodology of classificatory decomposition of signals based on hybridization of independent component analysis, multi-objective evolutionary algorithms, and rough sets. The preliminary results described here are very promising and further investigation of other MOEAs and/or RS-based classification accuracy measures should be pursued. The incorporation of ICA-derived basis functions and coefficients as the starting point in the MOEA significantly improved the reconstruction error and more closely related the concept of classificatory decomposition to the traditional signal decomposition techniques.

Acknowledgments. RSL - The Rough Set Library [17] and the open-source FastICA MATLAB package [18] were used in parts of this project.

Research partially sponsored by Burroughs-Wellcome Fund CASI Award, National Institutes of Health (NIH) NCRR grant P20 RR-16460, and grant no. 301-435-0888 from the National Center for Research Resources (NCRR), a component of the NIH.

References

1. Mamiya, N., Buchanan, R., Wallace, T., Skinner, D., Garcia, E.: Nicotine suppresses the P13 auditory evoked potential by actingon the pedunculopontine nucleusin the rat. Exp Brain Res **164** (2005) 109–119
2. Smolinski, T.G., Boratyn, G.M., Milanova, M., Zurada, J.M., Wrobel, A.: Evolutionary algorithms and rough sets-based hybrid approach to classificatory decomposition of cortical evoked potentials. Lecture Notes in Artificial Intelligence **2475** (2002) 621–628
3. Smolinski, T.G.: Classificatory Decomposition for Time Series Classification and Clustering. PhD thesis, Univ. of Louisville, Louisville (2004)
4. Smolinski, T.G., Milanova, M., Boratyn, G.M., Buchanan, R., Prinz, A.: Multi-objective evolutionary algorithms and rough sets for decomposition and analysis of cortical evoked potentials. In: Proc. IEEE International Conference on Granular Computing, Atlanta, GA (2006) 635–638
5. Amari, S., Cichocki, A., Yang, H.H.: A new learning algorithm for blind signal separation. In Touretzky, D.S., Mozer, M.C., Hasselmo, M.E., eds.: Advances in Neural Information Processing Systems. Volume 8. The MIT Press (1996) 757–763
6. Hyvarinen, A., Oja, E.: Independent component analysis: Algorithms and applications. Neural Networks **13** (2000) 411–430
7. Hyvarinen, A.: Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. on Neural Networks **10** (1999) 626–634
8. Deb, K.: Multi-objective optimization using evolutionary algorithms. Wiley (2001)
9. Schaffer, J.D.: Some Experiments in machine learning using vector evaluated genetic algorithms. PhD thesis, Vanderbilt University (1984)
10. Goldberg, D.E.: Genetic algorithms in search, optimization, and machine learning. Addison-Wesley (1989)
11. Srinivas, N., Deb, K.: Multiobjective optimization using nondominated sorting in genetic algorithms. Evolutionary Computation **2** (1994) 221–248
12. Laumanns, M., Zitzler, E., Thiele, L.: A unified model for multi-objective evolutionary algorithms with elitism. In: Proceedings of the 2000 Congress on Evolutionary Computation CEC00, IEEE Press (2000) 46–53
13. Pawlak, Z.: Rough sets - Theoretical aspects of reasoning about data. Kluwer (1991)
14. Wróblewski, J.: Finding minimal reducts using genetic algorithms. In: Proc. 2nd Annual Joint Conference on Information Sciences, Wrightsville Beach, NC (1995) 186–189
15. Milanova, M., Smolinski, T.G., Boratyn, G.M., Zurada, J.M., Wrobel, A.: Correlation kernel analysis and evolutionary algorithm-based modeling of the sensory activity within the rat's barrel cortex. Lecture Notes in Computer Science **2388** (2002) 198–212
16. Kreyszig, E.: Introductory functional analysis with applications. Wiley, New York, NY (1978)
17. Gawryś, M., Sienkiewicz, J.: RSL - The Rough Set Library version 2.0. Technical report, Warsaw University of Technology, Poland (1994)
18. Hurri, J.: The FastICA package for MATLAB. (http://www.cis.hut.fi/projects/ica/fastica/)