An Evolutionary Artificial Immune System for feature selection and parameters optimization of support vector machines for ERP assessment in a P300-based GKT

S. Shojaie[1], M. H. Moradi[1]
[1]Biomedical Engineering Faculty, Amirkabir University of Technology, Tehran, Iran
e-mail: s_shojaie@aut.ac .ir

Abstract- **Optimizing a classifier is a subject of great interest in the research area. A lot of methods inspired of biological metaphors are proposed for this task. This paper present a new algorithm based on the natural immune metaphors which select a proper subset of features and optimal parameters of a Support Vector Machines (SVM) classifier. The designed optimization method is validated for ERP assessment in a P300-based GKT (Guilty Knowledge Test). The result experiment shows the effectiveness of the method.**
***Keywords* - Artificial Immune System, Support vector machines, Feature selection, Classification**

I. Introduction

Many real-word problems involve the simultaneous feature selection and tuning parameters for Support Vector Machines (SVM) as one of the useful techniques for data classification. It is important that feature selection and parameters optimization are performed at the same time, because the kernel parameters are influenced by the feature subset choice and vice versa. The parameters that should be set for improvement the classification result, involve penalty parameter C and the kernel parameter γ for the radial basis function (RBF) kernel.

Genetic Algorithm (GA) as a global search method has successful application in solving many kinds of complex optimization problems [1]. Because of similarities between the basic computational theories of Evolutionary Algorithm (EA) and Artificial Immune System (AIS), mechanisms inspired by immune metaphors can be applied to the evolutionary optimization process. During the last decades, there has been growing interest in algorithms which use the natural principles such as evolution and Immunity. AIS had been widely used for combinational optimization task in the previous researches [2], [3]. In this work, we produce a novel Evolutionary Artificial Immune System (EAIS) that can be applied to optimize any classifier.

Biological immune system exhibits many information processing characteristics such as pattern matching, feature extraction, learning and memory, diversity and etc. AIS inspired by the Biological immune system, is a new computational intelligence method which has found application in a vast range of areas such as machine learning, data mining, pattern recognition, control optimization and etc.

Recording the brain potentials is one of the old non-invasive techniques for studying the brain functions [4]. This technique as a common used method, measures event-related changes in the Electroencephalography (EEG) known as Event-Related Potentials (ERP). ERPs are affected by the recognition of important events and have been extensively studied in the P300 waves. The P300-based GKT is a Guilty Knowledge Test which utilized P300 amplitude as an index of actual recognition of concealed information [5]. This test has been suggested as an alternative approach for conventional polygraphy. So, the designed GKT applied to several subjects and their respective brain signals were recorded. After preprocessing and removal the noises, for analysis of signals, a new approach consisting of different features and a SVM classifier was implemented.

The remainder of this paper is recognizes as follows: Section II gives background information of natural and artificial immune system. Also a brief introduction to the SVM is given in this section; Section III gives a brief description of our used data; Section IV describes our proposed method based immune metaphors for optimization task. Section V presents the experiment results and Section VI concludes this paper.

II. BACKGROUND

A. Natural and Artificial Immune Systems

The natural immune system is one of the most complex functional systems that protect the body from foreign disease causing pathogens called antigens. Defense mechanism of the body regulated by means of innate and adaptive immune responses. Between them, adaptive response is much more important because of its specific metaphors like self-regulation, recognition, memory acquisition, diversity, etc. the adaptive immune system consist of two types of lymphocyte known as T-cells and B-cells [6].

The principle of clonal selection as shown in Fig. 1, explains how the adaptive system recognizes and eliminates specific Ags by means of B-cells [7]. When a B-cell receptors named antibodies (Ab), recognize an Ag, it is selected to expanse. The number of clones generated, is proportional to the affinity of the selected cell and the Ag. So the highest affinity cells proliferate and the B-cell clones suffer a mutation process to produce B-cells with more affinity with the presented Ag. The mutation rate, opposite of clonal rate is inversely proportional to the affinity. Additionally, the best B-cells whose Abs present high affinity with the Ab, are selected as memory cells. Memory B-cells are kept for future responses to the same or similar antigenic patterns.

Artificial Immune systems proposed in 1990s as a new computational research area [8]. AIS aim at solving complex engineering problems such as optimization. Like Artificial Neural Network (ANN), GA and another EA, AIS is capable of learning new information, save this information at its memory, recall the learned information and then perform its tasks such as pattern recognition and optimization. Clonal selection principle is one of the biological inspired techniques that employed for optimization tasks. Some authors believe that clonal selection model is similar to genetic algorithm without crossover. The main difference between AIS and GA is the population improvement stages. In a standard GA, the population is developed by reproduction, crossover and

 978-1-4244-2695-9/08/$25.00 ©2008 IEEE

mutation but doesn't attend to important properties such as affinity proportional mutation and proliferation.

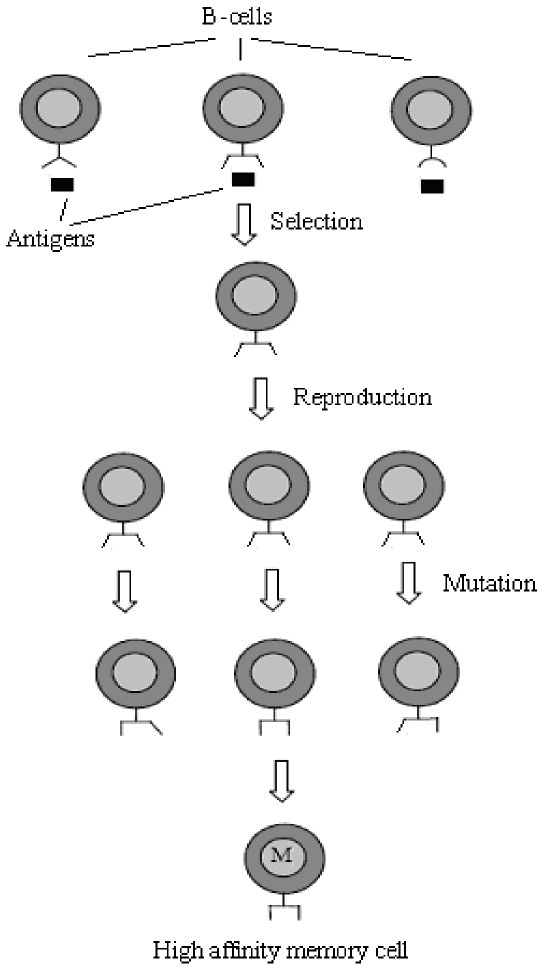

Fig. 1. Clonal selection principle.

B. Support Vector Machines

The SVM as one of the powerful techniques was first proposed by Vapnik (1995) and recently has been extensively used in pattern classification [9]. SVM maps the input data set into a higher dimensional feature space through several kernel functions [10]. SVM then attempts to find the optimal separation hyperplane with the maximum margin in the feature space. Creation of a hyperplane $w^T x + w_0 = 0$ (w is the vector of coefficients and w_0 is a bias term) so that the optimal separating hyperplane which has maximum margin linearly in the feature space is find and can be posed as quadratic optimization problem. After finding the optimal hyperplane, the decision function is built as following:

$$f(y) = \operatorname{sgn}\left[\sum_{i=1}^{N} \alpha_i y_i K_\theta(x_i, y) + b\right] \quad (1)$$

Where N is the sample size, $K_\theta(x_i, y)$ represents defined kernel function, θ a set of parameters and b is bias. α_i are non negative Lagrange multipliers that satisfy $\sum_{i=1}^{N} \alpha_i y_i = 0$ and are obtained by solving quadratic optimization problem:

$$w(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K_\theta(x_i, y) \quad (2)$$

SVM for obtaining the optimal solution uses different kernel function such as sigmoid, polynomial and radial basis function (RBF). In this study, RBF kernel function that is a effective option for classifying multi-dimensional data, is applied.

In the SVM classifier with RBF kernel, two parameter C and γ must be select appropriately [11]. Penalty parameter C influences on the rate of classification accuracy so that if it is set too small, the classification accuracy rate will be unsatisfied and if C is fitted excessively large, the rate of classification accuracy will be very high in the training phase but very low in the testing phase. So the choice of value for C is important factor to obtain desirable classification result. Variance parameter γ is the second parameter that is more effective on classification outcome. If it is selected too large, over-fitting will be resulted and vice versa.

III. USED DATA SOURCE

The applied method for generation of lie-detection dataset was as follows [12]:

A. subject

Sixty-two subjects who were generally undergraduate or postgraduate students participated in the experiment.

B. Data acquisition

The EEG was recorded using Ag/Ag-Cl electrodes placed at the Fz (Frontal), Cz (Central) and Pz (Parietal) sites (10–20 international system). All sites were referenced to linked mastoids and only the results from Pz will be reported here. Electrooculogram (EOG) was recorded from sub and supraorbital electrodes above and below the right eye and subjects were grounded at the forehead. Brain electrical activities were amplified and digitized at a rate of 256 samples per second. Achieved data were subsequently analyzed offline using MATLAB software. Prior to data analysis, all data were digitally filtered in the [0.3– 30] Hz range, which is the common frequency range in P300-based GKT studies.

C. Procedure

After a training phase about the protocol, a box – containing a jewel – was given to the subject. The examiner left the room and permitted the subject to perform his role in the scenario. In this step the subject could choose and implement one of two possible roles (Guilty/Innocent). The guilty subject was expected to open the box, examine the jewel precisely and imagine that he/she has stolen the jewel. The subject was asked to memorize the jewel details, so he/she can act as the actual robber. For reassurance, the subject was asked to write the detail of the object on a piece of paper and deliver it after the test. The innocent subject, however, had nothing to do with the box and thus had no information about the object.

All subjects were initially trained and then performed a mock crime scenario. After a training phase about the protocol, a box containing a jewel was given to the subject. The examiner left the room let subject play his/her role in the scenario. The subject could choose one of the guilty or innocent roles. The guilty subject opened the box, saw the jewel precisely and imagined that he/she has stolen the jewel. The subject asked to memorize the jewel details and for reassurance write that on a piece of paper for delivering after the test. The innocent subject had never opened the box and thus didn't have any information about object. Then the examiner returned to the room and executed the P300-based GKT for the knowledge of the scenario. After the attachment of electrodes and starting the recording, stimuli consisting of single pictures of five pieces of jewelry were presented randomly and successively on a monitor to the subject. Each item remained 1.1s on the screen, of which 1s was used for processing. The inter-stimulus interval was 2s. During the test, pictures of five items were randomly presented one at a time (each one with 30 iterations). These objects contained one target, one probe and three irrelevants. The probe was the object in the box. The target was another object, which has been previously presented to the subject in the training phase at the start of the protocol.

The subject gave one push button in each hand, right hand click as "YES" and left one click as "NO". Then, subject was asked to reply to the items by pressing one of two buttons; "YES" for familiar items and "NO" for unknowns. All subjects were proposed to press "YES" for targets and "NO" for probes and irrelevants. Thus, innocents and guilties, both replied honestly to targets and irrelevants; but fore probes, the innocents replied honestly while guilty answered falsely.

Each subject participated in two experiments in which the boxes and all displayed jewels were different. The jewel was a gold coin in the first experiment and a hairpin in the second one. Subjects chose a guilty role in one experiment and innocent role in the other. Signals from these two experiments were analyzed independently. So, for 62 subjects, 124 tests were performed. 32 subjects chose the innocent role and 30 subjects chose guilty role in the first experiment. A few test results were removed due to misdoing of protocol or inappropriately recorded signals. Finally, 59 guilty cases and 51 innocent were chosen to be used.

D. Data analysis

In this paper, we used classification based method for analyzing of the recorded signals. This is done by extraction of several suitable and common used features from the raw data which will be averaged for each subject.

Four different feature set was proposed: morphologic features (18 features), frequency features (3 features), wavelet features (32 features) and wavelet energy and wavelet entropy features (16 features). So, totally 69 features were extracted. For more detailed information refer to [13].

IV. EVOLUTIONARY ARTIFICIAL IMMUNE SYSTEM ALGORITHM

Evolutionary Artificial Immune System (EAIS) is a general adaptive optimization search methodology based on clonal selection model of immune system. Clonal selection, affinity maturation and memory cell introduction are the used immune metaphors in this paper. Our proposed algorithm works with a set of candidate solution called antibodies and after iterative computations, introduces the optimal solution. Because of using SVM classifier with RBF kernel, the features and two parameters (C, γ) must be optimized by our proposed EAIS. So each antibody involves three parts: first part for features, second for C and third for γ. We used the binary coding system for representation of antibodies. The binary bit string Ab representation of our design, showed in Fig.2. In this figure, $ab_1^f \sim ab_{N_f}^f$ considered for feature mask, $ab_1^c \sim ab_{N_c}^c$ and $ab_1^\gamma \sim ab_{N_\gamma}^\gamma$ represent the value of parameters C and γ respectively.

ab_1^f	$\cdots$	$ab_{N_f}^f$	ab_1^c	$\cdots$	$ab_{N_c}^c$	ab_1^γ	$\cdots$	$ab_{N_\gamma}^\gamma$

Fig. 2. The designed antibody that comprises three parts (feature mask, *C* and γ).

The proposed method (as shown in Fig.4) has 8 main steps explained as follows [14]:

(1) Initializing: randomly create an initial subpopulation of antibodies (*ab*). Each *ab* in *Ab* population is characterized by a binary string.
(2) Evaluation: evaluate each *ab* in *Ab* population according to fitness function adopted by user.
(3) Cloning: for each $ab \in Ab$, create a set of clones (*clone_Ab*). Number of clones depends on the fitness value of that *ab*. The maximum number of clones belongs to the best *ab*.
(4) Mutation: mutate each clone and add mutated population to Ab. Mutation is performed by changing the position of "1"s in the binary string describing the *ab*. Number of changed bits depend on the fitness value of $ab \in clone\_Ab$. The best *ab* changes less and vice versa.
(5) Diversity: add a number of new randomly generated *ab*s to *Ab* population.
(6) Evaluation: evaluate the behaviour of $ab \in Ab$,
(7) Suppression: hold the best member of *Ab* by suppression among subpopulation. So keep the size of *Ab* constant by removing worst *ab*.
(8) Repeat: iterate from (2) until the termination condition is met. Stopping criteria chosen by user. For example, the process is repeated for a given number of iteration or it stops if no improvement is observed in fitness.

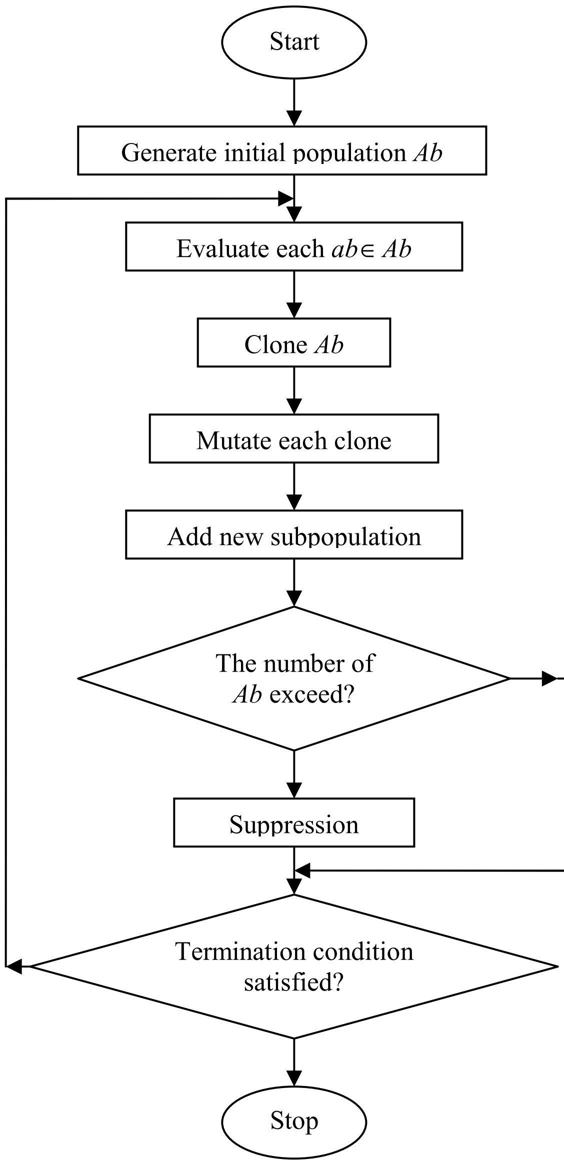

Fig. 2. Flowchart of the proposed EAIS-based optimization method.

V. EXPERIMENTAL RESULTS

To evaluate our proposed approach, we used lie detection dataset consists of 110 samples, 69 features and two classes collected by Abootalebi at the Tehran Amirkabir university [13].

A. Performance evaluation methods

For test results to be more valid, leave-one-out (LOO) cross validation is used in this research. In this method, the classification algorithm trained and tested N times (N is the number of samples). In each case, one of the sample represented average signals of a subject is taken as test data and remaining N-1 samples are added to form training data.

In this study, the classification accuracy of the proposed system measured as following:

$$accuracy = \frac{\sum_{i=1}^{N} assess(i)}{N} \tag{3}$$

$$assess(i) = \begin{cases} 1 & if\ classify(i)\ is\ true \\ 0 & otherwise \end{cases}$$

Where N is the number of subjects and *classify(i)* returns the classification of *i*th subject by SVM classifier.

B. Results and discussion

The details of parameters setting in our experiment, is shown in Table1. We designed the maximum generation 650 for termination criteria and classification accuracy as the fitness function.

As can be seen from table1, EAIS algorithm significantly improves the classification accuracy after optimization of SVM classifier. It obtained 85.45% classification accuracy using only 19 features. The increase of 21.81% in classification accuracy is good evidence for effectiveness of proposed approach.

TABLE I
USED PARAMETERS IN EAIS METHOD

used parameters	value
N_f (number of Ab bits for feature)	69
N_c (number of Ab bits for parameter C)	15
N_γ (number of Ab bits for parameter γ)	15

TABLE 2
OUR METHOD'S OPTIMIZATION RESULT FOR CLASSIFICATION OF LIE-DETECTION DATA SET AND COMPARISION WITH UNOPTIMISED METHOD

number of features	parameters of SVM		accuracy
	C	γ	
69	1	1	63.64
19	1987	0.001	85.45

For comparison of implemented technique with GA-based approach, we applied both methods with the same condition to solve the corresponding optimization problem. These methods are very similar in many aspects and the main difference between them is in the population evolution stage. Despite of GA, our proposed algorithm accounts for important properties such as affinity proportional proliferation and mutation in improvement population stage. Fig.3 shows that the rate of classification accuracy improvement in EAIS is higher than GA such as EAIS reaches 85.45% accuracy after 294 generation but GA obtained this result after 541 generation. On the other hand, the evolution of the population in EAIS method occurs faster than GA. Finally, the number of selected features and the best pairs of C, γ for each method are shown in Table3.

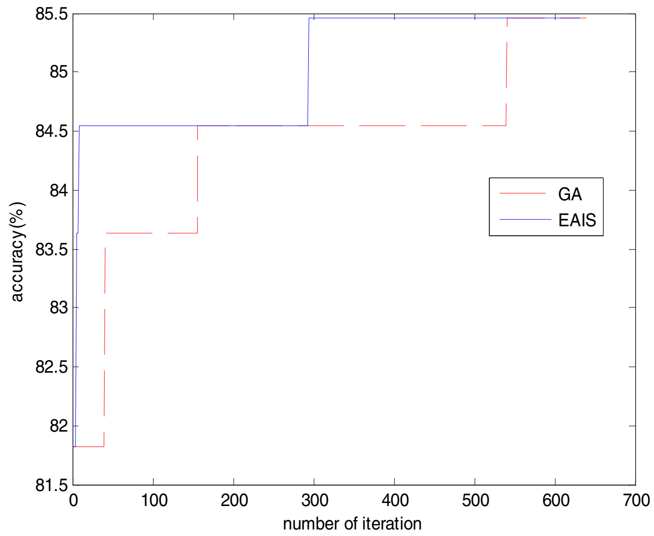

Fig. 3. Illustration of classification accuracy versus the iteration number.

TABLE3

COMPARISON OF EAIS AND GA APPROACH IN OPTIMIZATION TASK

method	number of selected features	best parameters	
		C	γ
EAIS	19	1987	0.001
GA	11	17204	0.01

VI. CONCLUSION

In this work, we proposed a strategy based on natural immune mechanism to select the best feature subset and optimize the parameters of SVM classifier. We conducted experiments to evaluate our proposed method for ERP assessment in a P300-based GKT. Results showed that EAIS is a successful optimizer method that can develop the classification performance with fewer features.

Generally, compared the EAIS with GA approach proved the effectiveness of our method. EAIS because of best capabilities in population evolution has a higher performance of convergence.

The proposed approach can also be applied to improve the classification accuracy of any classifier that need to be optimized.

REFERENCES

[1] C. L. Huang and C. J. Wang, "A GA-based feature selection and parameter optimization for support vector machines," *expert system with applications. Elsevier. Taipei*, vol. 31, pp. 231-240, 2006.

[2] L. Jiao, Senior member, IEEE and L. wang, "A novel genetic algorithm based on immunity," *Transaction on systems, Man, and cybernetics. IEEE*, Vol. 30, pp. 5, September 2000.

[3] K. C. Tan, C. K. Goh and A. A. Mamun, "An evolutionary artificial immune system for mukti-objective optimization," *European Journal of Operational Research. Elsevier*, Vol. 187, pp. 371-392, 2008.

[4] V. Abootalebi, M. H. Moradi and M. A. Khalilzadeh, "Detection of the cognitive components of brain potentials using wavelet coefficients," *Iranian Journal of biomedical engineering. Tehran. Iran*, Vol. 1, pp. 25-46, 2004.

[5] J. P. Rosenfeld, "Event-Related Potentials in the Detection of Deception, Malingering, and False Memories," in *Handbook of Polygraph Testing,* M. Kleiner, Ed. New York: Academic Press, 2002, ch. 10, pp. 265-286.

[6] F. Gonzalez. "A study of Artificial Immune Systems Applied to Anomaly Detection," May 2003.

[7] L. N. de Castro and J. Timmis, "Artificial Immune Systems: A Novel Paradigm to Pattern Recognition," Computing Laboratory University of Kent.

[8] S. N. Omkar, R. Khandelwal, S. Yathindra, G. N. Naik and S. Gopalakrishnan, "Artificial immune system for multi-objective design optimization of composite structures," *Engineering Application of Artificial Intelligence. Elsevier. India*, January 2008.

[9] C. Cortes and V. Vapnik, "Support-Vector networks," *Machine Learning*, vol. 20, pp. 273-297, 1995.

[10] E. Comak and A. Arslan, "A new training method for support vector machines: Clustering k-NN support vector machines," *Expert systems with Applications. Elsevie*, Vol. 35, pp. 564-568, 2008.

[11] S. W. Lin, K. C. Ying, S.C. Chen and Z. J. Lee, "Particle swarm optimization for parameter determination and feature selection of support vector machines," *Expert systems with Applications. Elsevie. Taiwan*, Vol. 35, 1817-1824, 2008.

[12] V. Abootalebi, M. H. Moradi and M. A. Khalilzadeh, "A comparison of methods for ERP assessment in a P300-based GKT," *International Journal of Psychophysiology. Elsevier. Tehran. Iran*, Vol. 62, pp. 309-320, May 2006.

[13] V. Abootalebi, M. H. Moradi and M. A. Khalilzadeh, "Analysis of the cognitive components of brain potentials and its application on lie-detection," *Iranian PHD Thesis. biomedical engineering. Tehran. Iran*, 2006.

[14] M. Bereta and T. Burczynski, , "Comparing binary and real-valued coding in hybrid immune algorithm for feature selection and classification of ECG signals," *Engineering Application of Artificial Intelligence. Elsevie*, Vol. 20, pp. 571-585, November 2006.