Genetic Algorithm and Tabu Search for Feature Selection

Sabra El Ferchichi [1], Kaouther Laabidi [1], Salah Zidi [2],

[1]LACS, ENIT Tunis, TUNISIE,

[2]LAGIS, USTL Lille, FRANCE,

Abstract: In this paper we propose a wrapper approach to select features involving the Support Vector Machines (SVM) combined with a metaheuristic optimization algorithm: Tabu Search and Genetic Algorithms. Feature selection is efficient in searching the most descriptive features which would contribute in increasing the performance of the inductive algorithm by reducing dimensionality and processing time. The process we propose is based on the use of the rate of misclassification as an evaluating criterion. First, we used the tabu algorithm to guide the search of the optimal set of features; then a genetic algorithm is implemented to reach the same goal. This procedure is applied on data from regulation of urban transport network systems. A comparison between the performances of each search engine (TS and GAs) used is then presented.

Keywords: Features Selection, Support Vector Machines, Tabu Search, Genetic algorithm and urban transport regulation.

Sabra El Ferchichi received the engineering degree in Electrical Engineering and the M.S degree in Control and Signal Processing from the National School Engineering of Tunis in 2007 and 2008. Where, she is currently pursuing her Ph.D. degree in Department of Electrical Engineering at National School Engineering of Tunis. Her research interests include Feature extraction and selection, learning machines, metaheuristic algorithms.

Kaouther Laabidi was born in Tunis, Tunisia. She received the Master degree from the "Ecole Suprieur de Sciences et de Technologie de Tunis", in 1995 and the Ph.D degree in Genie Electique from the "Ecole Nationale d'Ingnieurs de Tunis", in 2005. She is currently preparing the ability degree in the laboratory ACS "Analyse et Commande des Systmes".His research is related to the Ide ntification and control of complex systems.

Salah Zidi was born in Gafsa, Tunisia, in 1980. In 2003, he received the B.E. degree in electrical engineering from the National School of Engineers of Sfax, Tunisia. He received the D.E.A. degree in new technologies of dedicated computing systems from the National School of Engineers of Sfax, Tunisia, in 2004. He received the Ph.D. degrees in automatic control from the University of Lille, Lille, France, in 2007. He is currently a teacher and researcher at the University of Lille. Her research interests include transport, reconfiguration, scheduling, ant colony algorithms, and decision-support systems.

1. Introduction

Dealing with high dimensional data reveals of computational time's problem. Thus, feature dimensionality reduction is of a considerable importance to treat this problem and to improve the classifier's performance.

To address this problem, Feature Selection (FS) tries to find an optimal subset of features from the original set which contains the most pertinent ones taking into account regards to the classification task [1]. In this context, two main directions exist; they are based on the evaluation subset' criterion: the first is the filter approach; the second is the wrapper approach [2]. Filter method' principle is based on sorting variables according to their score, computed with the use of different criteria. The features with low scores are then eliminated. Wrapper methods rely on the classification' accuracy to determine the most optimal subset of features: the search for the optimal subset makes use of the classification algorithm itself to evaluate candidate subsets [3]. Other method implements the same concept as the wrapper is called embedded methods. It involves knowledge about the structure of the classifier on the selection process, there is no more separation between classification and selection task [4].

The wrapper approach requires a search space, operators, search engine, and an evaluation function [2]. The search space in our case is composed of a large set of candidate subsets of features, which makes the evaluation of all possible subsets quite impractical. As a result, the exhaustive search has been replaced by heuristic search strategies in an attempt to avoid this high-priced complexity. This solution is computationally feasible and represents a trade-off balance between solution quality and processing time as argued in face recognition [5] and image annotation field [6]. within this work, we propose a feature selection algorithm involving an SVM-driven tabu search and genetic algorithm. Support Vector Machines 'SVM' has been applied to a variety of areas and it has shown a great performance and success in areas such as Biomedical data [7, 9], speech recognition [8]. The classification accuracy is used as an evaluation function to access feature' relevance.

The paper is organised as follows. In section 2, we briefly introduce the Tabu Search and the Genetic Algorithm. Support Vector Machines and our wrapper approach to feature selection are presented in the section 3. Section 4 is reserved to simulation' results on data set concerning the regulation of urban transport networks, comparing SVM-driven Tabu Search and SVM-driven GAs. Finally, a conclusion to our work is developed.

2. Heuristic Strategies for Feature Selection Algorithm

Search strategy is a key problem in feature subset selection. To come to a compromise between result optimality and computational time efficiency, search strategies such as heuristic or random search have been introduced to generate candidate feature subsets for evaluation step through a determined classifier [6]. We will give in the following paragraph an overview of the two heuristic strategies we have used in our work: Tabu Search (TS) and Genetic Algorithms (GAs).

A. Tabu search

It is a local metaheuristic method based on the exploration of the whole space of possible solution from an initial one. Within every iteration, the algorithm chooses the best solution among the neighbourhoods generated for the current solution. It uses a tabu list to memorize the last visited solutions in attempt to avoid local optima, it is forbidden to return to these tabu solutions [10]. Concepts like diversification and intensification have been introduced to ameliorate the procedure. On the one hand, diversification tries to explore the whole space of search and not trapped on specific area. On the other hand, intensification focuses on some areas which may contain the global solution [11].

B. Genetic algorithms

It is based on the mechanism of natural selection. They used operations found in natural genetics to guide itself through the paths in the search space. The procedure starts from a randomly chosen population of possible solutions. Every individual from the population is denoted as a chromosome. At each generation, a fitness function is used to evaluate the performance of each chromosome. on whose bases, we can determine if that chromosome would survive or not for the next generation. After selecting a set of best chromosomes, allowed reproducing, the algorithm makes use of the crossover and mutation operations to reconstruct a novel population which is more likely to reach the goal. This process continues until it converges on the optimal fitness requested or the required number of generations is reached. The optimal solution is then determined [10].

3. Wrapper Approach Involving Support Vector Machine for Feature Selection

The general concept is described by the following figure1:

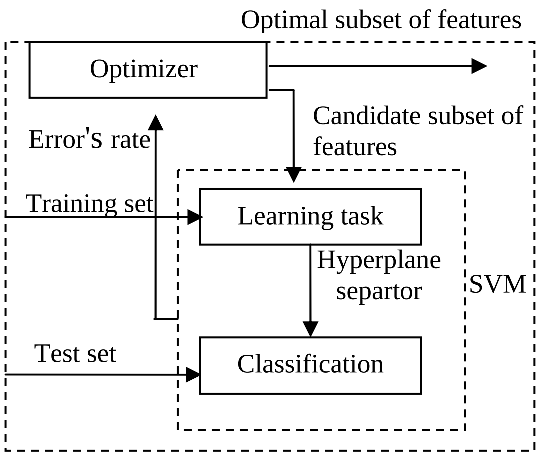

Figure 1. FS involving SVM-driven optimizer algorithm

We have implemented an algorithm of Tabu Search as an optimizer algorithm in a first time, and then we have used the Genetic Algorithms instead. We have used an SVM classifier in our approach. The criterion to evaluate classification' performance is the accuracy of classification or rate of misclassification. The optimal subset of feature is the one which is ranked on the bottom with the least rate of classification error. The data base is split in two subsets; one for the training task, the rest is used for the test task. Each feature is denoted by a variable η which can be either 0 or 1. If a feature is present in the candidate subset, it will be indicated by $\eta = 1$ otherwise, it will be assigned to 0.

A. SVM

SVM is a learning machine based on the structural risk minimization induction principle derived from the statistical theory developed by Vapnick in 1982 [12]. It was developed basically for the binary classification and then extended to the multiclass case [13]. Consider we have a training set composed of N vectors $X = \{x_i\}_{i \le N}$, each vector x_i has a label $y_i \in \{1, -1\}$.

The two classes are linearly separated if an hyperplane defined by $\langle w, x \rangle + b = 0$ can be found, where w is the weight and b is the offset. SVM attempts to find the most optimal hyperplane that separates perfectly the two classes (without training error) and maximizes the minimum distance from a training sample to the hyperplane by solving the following optimization system.

$$\begin{cases} \min\limits_{w,b} \frac{1}{2}\|w\|^2 \\ \forall i, y_i(\langle w, x_i \rangle + b) \ge 1 \end{cases} \quad (1)$$

For linearly non separable case, slack variable $\xi_i \ge 0$ is introduced to relax the previous constraints by accepting errors in separating data. We obtain the following formulation:

$$\begin{cases} \min\limits_{w,b,\xi_i} \frac{1}{2}\|w\|^2 + C\sum_i \xi_i \\ \forall i, y_i(\langle w, x_i \rangle + b) \ge 1 - \xi_i \\ \xi_i \ge 0 \end{cases} \quad (2)$$

To solve this quadratic optimisation, the dual Lagrange formulation can be used and the decision function is defined as follows:

$$f(x) = \operatorname{sign}\left(\sum_{i \in SV} \alpha_i y_i \langle x_i, x \rangle + b\right) \quad (3)$$

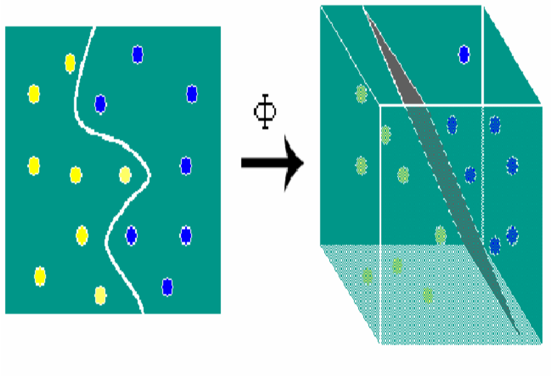

Figure 2. Redescription space

For nonlinearly cases, the idea is to get to another dimension bigger than the original one where a linear hyperplane could be found. Figure 2 illustrates this non linear transformation noted Φ, generally unknown, kernel function is thus introduced to turn around. The initial system (1) is then transformed into this form:

$$\begin{cases} \max\limits_{\alpha_i} \sum_i^l \alpha_i - \frac{1}{2}\sum_{i,j}^l \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ C \ge \alpha_i \ge 0 \\ \sum_i^l \alpha_i y_i = 0 \end{cases} \quad (4)$$

B. SVM-driven Tabu search

In this section, we propose a feature selection algorithm based on SVM-driven Tabu search. It uses the Support Vector Machines as an inductive algorithm. Performance of each subset of features in classification task namely misclassification' rate is considered as an evaluating criterion. the Tabu algorithm has been used as an engine to guide the search.

We start by generating a set of

neighbourhoods; each one will be evaluated by calculating the misclassification rate with an SVM classifier. The best neighbourhood is then compared to the optimal solution we had obtained. If it is the best, it will be chosen and stocked in a tabu list. It is forbidden to return to this solution. The algorithm continues to run until a stop criterion is met. Finally, it will give the optimal solution and the corresponding subset of features. This procedure is illustrated by figure 3a.

C. SVM-driven GAs

In this section, we propose a feature selection algorithm based on SVM-driven Genetic algorithms. It also uses the Support Vector Machines as an inductive algorithm. Performance of each subset of features in classification task is considered as an evaluating criterion. The Genetic algorithm has been used as an engine to guide the search.

We start by generating a population of random subset of features; each subset is referred to as a chromosome and will be evaluated by calculating the misclassification' rate. According to the evaluation function, the individuals are ranked. We used an N/2 elitism strategy to select the best set of chromosomes. After selection step, the crossover operator is used depending on the probability. It is made of two different parts, chosen randomly each time, of the chromosome. The mutation operator is applied to diversify the solutions in attempt to accelerate the process, and it is controlled by a probability factor. The algorithm continues to run until it reaches the required number of generations. Finally, it will give the optimal solution and the corresponding subset of features. This procedure is illustrated by figure 3b.

4. Simulation' Results

We have applied the feature selection algorithm we proposed on data collected from regulation of urban transport networks systems. In fact, transport system could be affected by random incidents which may cause disturbance in the planning table made to organize the exploitation of transportation networks.

To re-establish the theoretical scheduling, a regulation system is implicated. A vector composed of 24 attributes is used to describe precisely this perturbation. A first bloc of classification of different perturbations is developed within the support vector machines. Our work is to search among the space of these attributes, those which are the most reliable ones. We applied FS using TS and FS using GAs in the case of binary classification.

Results and comparison

The initial solution is composed of all 24 parameters. The corresponding error rate is 3%. When first Tabu search is used as a search engine, within our selective approach, we have reached an error rate of 1.5%. The algorithm reaches this result through 3000 iterations. It starts from the initial solution and at each iteration, we generate 7 neighbourhoods; only two bits from the 24 attributes change randomly their status. To diversify our search, we have used a random jump on different positions. The Tabu list is fixed randomly to 44.

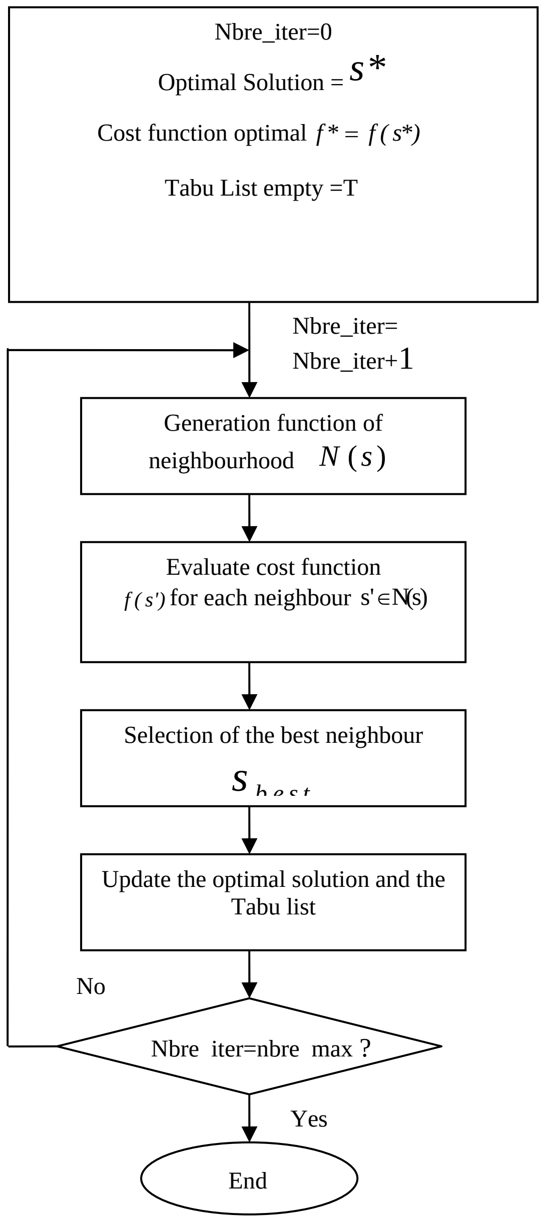

Figure 3.a. Feature selection algorithm using TS

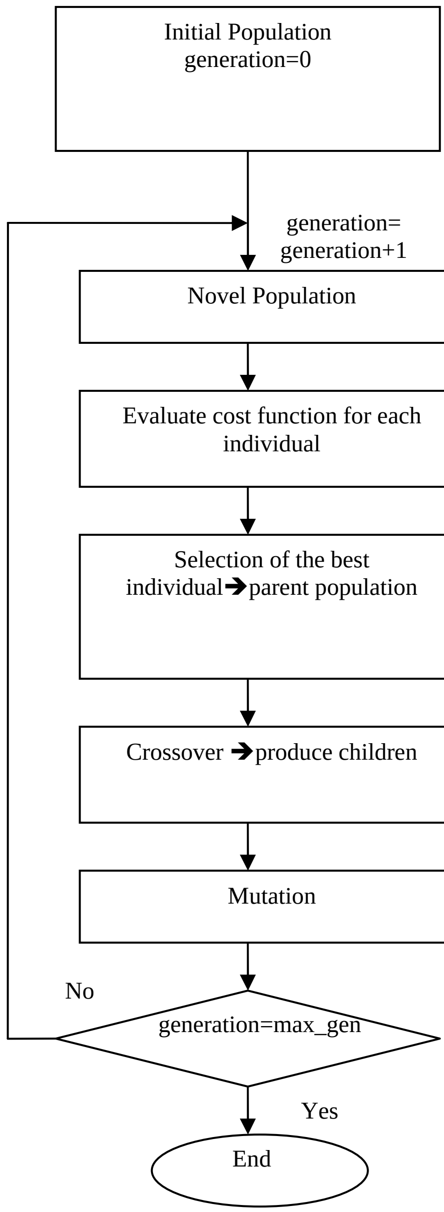

Figure 3.b. Feature selection algorithm using GAs

Then, Genetic Algorithm is used as a search engine. The initial population is set arbitrarily; it is fixed at 30 individuals. The algorithm reaches the error rate of 1.5% through 200 generations. Within every generation, the algorithm evaluates the population and selects the best group to constitute the parent set which will be used in reproduction task.

A probability of crossover is fixed at 0.6, two positions are chosen and the crossover is accomplished.

Then, the mutation operator modifies each time a random chosen bit from a fixed individual close to the solution obtained by the tabu search. It's controlled by a probability factor. A fraction of the best individuals are excluded from this step, aiming at conserving the most fortunate individuals of the original population and accelerating the search procedure.

Figure 4.a and 4.b show the evolution of the Tabu Search algorithm through 3000 iterations in relation to the error' rate and the dimension. The Same way, Figure 5.a and 5.b present the results of Genetic Algorithms implementation relative to the error' rate and the dimension.

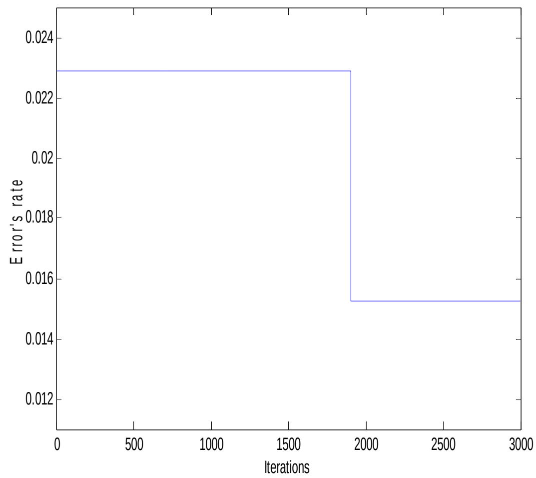

Figure 4.a. Error' rate through Tabu Search

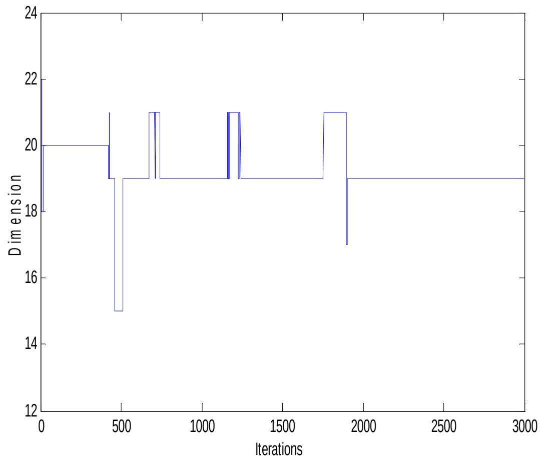

Figure 4.b. Dimension through Tabu search

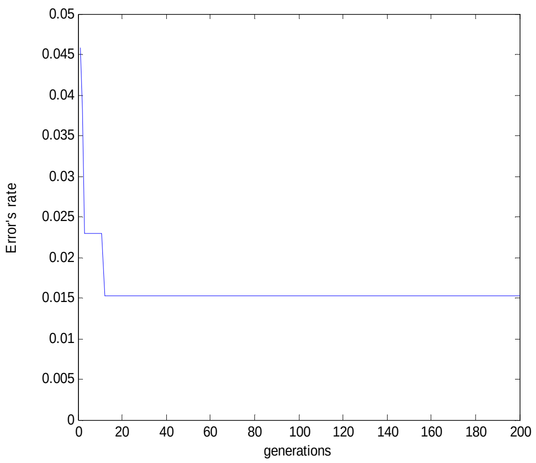

Figure 5.a. Error' rate through GAs

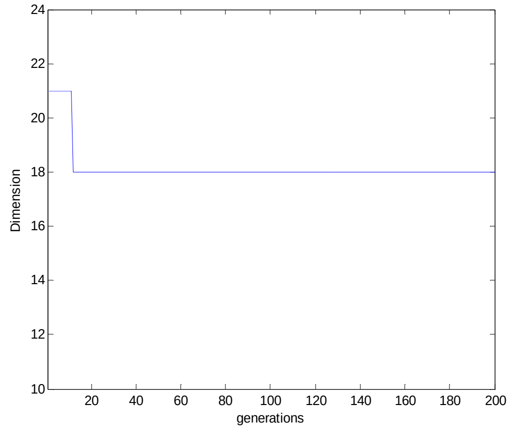

Figure 5.b. Dimension through Gas

Comparing the two type of engine search, we notice that genetic algorithm reaches the optimal solution within only 200 generations. On the other side, the tabu search needs at least 2000 iterations to get to the optimal solution. As a result GAs has converged faster than Tabu Search.

We can notice through the figure 4.b that in Tabu Search the dimension is not necessarily decreasing as the misclassification's rate decreases. For the same value of misclassification' rate, for example 0.0229, Tabu Search is able to present a diversity of solutions (different feature subsets) with different dimensions (20, 21, 19) which is not the case with the genetic algorithm (only 18). GAs tries to achieve better solutions by using knowledge from previous generations and getting finally to a unique final solution (dimension is 22 for the local optimum 0.0229).

5. Conclusion

Ameliorating classification accuracy is the main concern of learning machines. The feature selection algorithm aims at resolving this issue by searching for an optimal subset of features from the original one. Selecting the most pertinent feature would reduce the processing time and reduce the dimensionality of data.

We have proposed in this work a wrapper approach based on Support Vector machines as an inductive algorithm. Since the search space is very large, non conventional approaches would present an acceptable solution to avoid this prohibitive complexity. We have implemented in a first time SVM driven by tabu search, then SVM driven by genetic algorithm.

As our principle concern is the classifier's accuracy, reducing dimensionality of data is one way to get to our target. In that progress, as Genetic Algorithms performs previous solution in an iterative way, it gets to the optimal solution much faster than the tabu search, albeit that TS gain on dimensionality (it could attend the dimension 17). Tabu Search has also the priority to get to a diversity of solutions for the same rate of misclassification.

REFERENCES

1. BADEA, LIVIU, **Simultaneous Feature selection and Clustering for Gene Expression Data using Nonnegative Matrix Factorization with offset**, Studies in Informatics and Control, Vol. 16, No. 4, 2007.
2. KOHAVI, RON and GEORGE. H. JOHN, **Wrapper for feature subset selection,** ELSIEVER Artificial Intelligence, Vol. 97, No. 2-1, 1997, pp.273-324.
3. GUYON, ISABELLE and ANDRÉ ELISSEEFF, **An Introduction to Variable and Feature selection,** Journal of Machine Learning Research, Vol.3, Special Issue, March 2003, pp. 1157-1182.
4. LAL, T. N., O. CHAPELLE, J. WESTON and A. ELISSEEFF, **Embedded methods, Feature Extraction: Foundations and Applications**, 2006, Berlin, Germany, pp.137-165.
5. HAMIDREZA RASHIDY KANAN, KARIM FAEZ, **An improved feature selection method based on Ant Colony Optimization (ACO) evaluated on face recognition system**, ELSIEVER Applied Mathematics and Computation, Vol. 205, No. 2, 2008, pp.716-725.
6. JIANJIANG LU, TIANZHONG ZHAO, YAFEI ZHANG, **Feature selection based-on genetic algorithm for image annotation**, ELSIEVER Knowledge-Based Systems, Vol.21, No. 8, 2008, pp. 887-891.
7. LUYIN ZHAO, LILLA BOROCZKY, K.P. LEE, **False positive reduction for lung nodule CAD using support vector machines and genetic algorithms**, ELSIEVER International Congress Series, Vol.1281, May 2005, pp. 1109-1114.
8. SALOMON, JESPER, **Support Vector Machines for Phomene classification**, Master of Science School of Artificial Intelligence Division of Informatics University of Edinburgh, 2001.
9. ZHOU, NINA and LIPO WANG, **A Novel Vector Machine with class-dependent Features for Biomedical Data**, IEEE International Conference on Systems, Man, and Cybernetics, Vol.2, October 8-11, 2006,Taipei,Taiwan, pp.1666-1670.
10. DREO, J. A. PETROWSKI, P. SIARRY, E. TAILLARD, **Métaheuristiques pour l'optimisation difficile**, Editions EYROLLES, 2003.
11. AYAS, JEAN et MARC ANDRE VIAU, la recherche Tabou, cours Polytechnique Montréal, Novembre 2004.
12. VAPNIK, V.N., **An Overview of Statistical Learning Theory**, IEEE transactions on Neural Networks, Vol.10, No.5, September 1999, pp. 988-999.
13. HSU, CHIH-WEI and CHIH JEN LIN, **A Comparison of Methods for Multiclass Support Vector machines,** IEEE Transactions on Neural Networks, Vol.13, No.2, March 2002, pp. 415-425.