

© 2013, TextRoad Publication

# Memetic Feature Selection Algorithm Based on Efficient Filter Local Search

# Mohadeseh Montazeri<sup>1,2</sup>, Hamid Reza Naji<sup>3</sup>, Mitra Montazeri<sup>4,5,1</sup>

<sup>1</sup>Research Center for Modeling in Health, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran. <sup>2</sup> Technical and Vocational University, Kerman, Iran,

<sup>3</sup> College of Electrical and Computer Engineering, Graduate University of Advanced Technology, Kerman, Iran.

<sup>4</sup> Medical Informatics Research Center , Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran. <sup>5</sup> Computer Engineering Department, Shahid Bahonar University, Kerman, Iran.

> Received: June 5 2013 Accepted: July 3 2013

# ABSTRACT

Feature selection is a problem of finding efficient features among all features in which the final feature set can improve accuracy and reduce complexity. In this paper, we have proposed a method based on memetic algorithm to find an efficient feature subset for classification purposes. It incorporates a filter method in wrapper method to improve classification performance and accelerates the search in identifying core feature subsets. Especially, this methoddeletes or adds a feature from a feature subset based on the multivariate feature information. Empirical study on commonly data sets of the University of California, Irvine shows that the proposed method outperforms existing methods. Furthermore, we have investigated several major issues of memetic algorithm to identify a good balance between local search and genetic search so as to maximize search quality in the hybrid filter and wrapper memetic algorithm.

KEYWORDS: Feature Selection, Memetic Algorithms, Meta-Heuristic Algorithms, Local search.

# 1. INTRODUCTION

Most of the real world classification problems require supervised [1] or semi-supervised [2] learning which use class information to establish a model to predict unseen instance. In these models, the underlying class probabilities and class conditional probabilities are unknown [3], and some instances are associated with a class label. In real-world situations, relevant features often did not have an unknown priori. Therefore, many candidate features are introduced to better represent the domain. Although in theory, more feature increases discriminating power, but in practice, more feature not only reduces the learning speed of the learning process but also makes the learning process of the classifier over-fit to the training data.

Feature Selection is the problem of selecting informative feature among all features in which a selected feature subset has lower cardinality and have higher accuracy. Feature selection is an attractive research field which is improved recently in statistical pattern recognition [4], machine learning [5, 6], data mining [7, 8] and statistics [9, 10]. It has been proved that feature selection causes:

- Performance improvement, to gain higher accuracy;
- Data understanding, to visualize the data simply;
- data reduction, to speed up learning ;

In general, a feature selection algorithm consists of four basic steps: subset generation, subset evaluation, stopping criterion, and result validation [11]. Subset generation is a search procedure. Basically, it generates subsets of features for evaluation. Cardinality of features in the original data set is set to N, then the total number of candidate subset is  $2^N$  which is an exhaustive search [32]. Each subset which is generated in the generation procedure should be evaluated. It is done by a certain evaluation criterion then with respect to this criterion it is compared with the best one. If this subset is better, then it replaces with the previous best one. Without a proper stopping criterion the feature selection process is exhaustive search. The best feature subset needs to be validated and compare to the original set.

Various aspects of feature selection have been studied by researchers. There are two key aspects: feature evaluation and search strategies. Feature evaluation is how to measure the goodness of a feature subset[12, 13]. There are filter models[14-16] and wrapper models[17-19] with different emphases on dimensionality reduction or accuracy enhancement.

The filters methods are based on the intrinsic properties of data. These methods seek the relevant features and eliminate the irrelevant ones. This method finds efficient features in one of two ways: univariate method and multivariate method. In univariate method, it computes some simple score S(i) that measures how informative each feature  $x_i$  is about the class labels y. using this method have three problems. First problem, features that are not individually relevant should become deleted but they may become relevant in the context of others [20]; second problem, always the relevant feature is not useful one because of possible redundancies [21]. Third problem, when the features were ranked according to their scores S(i), if the number of effective feature is not determine, decide how many features is difficult and time consuming. Therefore, the second method in filter approach is attended. In this method, it takes into account feature dependencies. This method potentially achieves better results because they consider feature dependence [21] but it is obvious, they need search strategies in feature space to find the best feature subset.

\*Corresponding Author: Mitra Montazeri, Medical Informatics Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran. Email: mmontazeri@kmu.ac.ir, montazeri\_mitra@yahoo.com.

#### Montazeri et al., 2013

All filter based methods are fast and allow them to be used with any learner. An important disadvantage of filter methods is that they ignore the interaction with the classifier therefore they have low accuracy. Beside this method, wrapper approach is placed. Wrapper approach is a procedure that "wraps" around a learning algorithm, and accuracy of this learning algorithm is to be used to evaluate the goodness of different feature subsets. Due to the learning algorithm is called repeatedly, its execution is slow. But it has high accuracy. Recently, due to combine the advantages of both models, hybrid models has been attractive.

Filters and wrappers differ mostly by the evaluation criterion. Both methods can make use of search strategies to explore the space of all possible feature combinations that is usually too large to be explored exhaustively. Therefore, another important aspect i.e. Search strategies [12, 22]were studied by researchers.

Heuristic search employs heuristics in conducting search. Due to polynomial complexity, it can be implemented very faster than exact searches. These methods do not guarantee the optimality or feasibility. Therefore many researches have been done to increase and guarantee the optimality.

Among heuristic algorithms, meta-heuristic algorithm which is kind of heuristic algorithm, by tightening a focus on good solutions and improving upon them (Exploration), and to encourage the exploration of the solution space by broadening the focus of the search into new areas (diversification)can search solution space effectively[23].Exploration and Exploitation are two competing goals govern the design of global search methods. Exploration is important to ensure global reliability, i.e., every part of the domain is searched enough to provide a reliable estimate of the global optimum; exploitation is also important since it concentrates the search effort around the best solutions found so far by searching their neighborhoods to produce better solutions[34].

Meta-heuristc is capable of global exploration and locating new regions of the solution space to identify potential candidates, but there is no further focus on the exploitation aspect when a potential region is identified[24]. Thus, Mimetic Algorithms(MA) which incorporate local improvement search into meta-heuristics, were proposed [25].

Mimetic algorithms are population-based meta-heuristic search approaches that have been receiving increasing attention in the recent years. Generally, Mimetic algorithms may be regarded as a marriage between a population-based global search and local improvement procedures. Experimental studies have been shown that a hybrid of a meta-heuristic and a local search is capable of more efficient search capabilities[26].

In this paper, we propose a feature selection algorithm based on mimetic algorithm(MFS). The goal of MFS is to improve classification performance and accelerate the search to identify important feature subsets. In particular, our local search is based on filter method in order to tune the population of Genetic Algorithm (GA) solutions by adding or deleting features based on multivariate feature information. Hence, our focus is on filter methods that are able to assess the goodness of one feature in the context of others. We denote popular filter method, Pearson Correlation Coefficient, as our filter method in this correspondence. Furthermore, we investigate the balance between exploitation and exploration. Empirical study of MF Son most commonly used data sets from the University of California, Irvine (UCI) repository[27] indicates that it outperforms recent existing methods.

The rest of this article is organized as follows: Section 2 describes preliminaries of some definitions about Pearson Correlation Coefficient and feature selection. In Section 3, we explain MFS. The experimental results and conclusion are presented in Section 4 and 5, respectively.

#### 2 Preliminaries

In this section, several basic concepts about Pearson Correlation Coefficient are given, and then the formalism of feature selection is presented.

#### 2.1 Pearson Correlation Coefficient

Correlation coefficients are the simplest approach to feature relevance measurements. In contrast with information theoretic and decision tree approaches they avoid problems with probability density estimation and discretization of continuous features and therefore are treated first.

The linear correlation coefficient of Pearson is very popular in statistics and represents the normalized measure of the strength of linear relationship between variables[28]. For random variable X with values x and random variable Y with values y, while a vector of d data points  $x_i, y_i, i = 1, ..., d$ , it is defined as:

$$r_{x,y} = \frac{\sum_{i} (x_{i} - \bar{x}_{i})(y_{i} - \bar{y}_{i})}{\sqrt{\sum_{i} (x_{i} - \bar{x}_{i})^{2} \sum_{j} (y_{i} - \bar{y}_{i})^{2}}}$$
(1)

Where, as usual,  $\overline{x}$  is the mean of the  $x_i$ 's,  $\overline{y}$  is the mean of the  $y_i$ 's.  $r_{xy}$  is equal to  $\pm 1$  if X and Y are linearly dependent and zero if they are completely uncorrelated (Random variables may be correlated positively or negatively).

If a group of k features variables has already been existed, correlation coefficients estimate correlation between this group and the class variable, including inter-correlations between the features. With the correlation between features and classes, relevance of a group of features grows, and decreases with growing inter-correlation. These ideas have been discussed in theory of psychological measurements [29] and in the literature on decision making and aggregating opinions [30]. In 1964, Ghiselli[29] proposed equation (2):

$$J_k = \frac{k\bar{r}_{cf}}{\sqrt{k+k(k-1)\bar{r}_{ff}}}$$
(2)

Where  $\overline{r_{cf}}$  is the average correlation coefficient between these k features and the output variables and the average between different features as  $\overline{r_{ff}}$ . This formula is obtained from Pearson's correlation coefficient with all variables standardized. It has been used in the Correlation-based Feature Selection(CFS) algorithm[31].

## 2.2 Feature Selection Problem

We can formulate the problem of selecting the subset of features with superior classificatory performance as follows: Let F be the original set of features with cardinality N. Let m represent the number of features in the selected subset, X,  $X \subseteq F$ . Let the feature selection criterion function for the set X be represented by J(X) which is supposed a high value of J to indicate a better feature subset. In formal, the problem of feature selection is to find a subset  $X \subseteq F$  which has two important properties:

- m < N
- J(x) > J(N)

It has been demonstrated that searching for the minimum feature subset is NP-hard[32].

## 3 MFS

Fig.1shows the procedure of MFS. In the first step, the GA population is initialized randomly which each chromosome is encoding a candidate feature subset. Then, on the elite chromosomes, a local search or meme is applied. The mechanism of local improvement can be reaching a local optimum or improving the solution. Genetic operators are then used to generate the next population. This process repeats until the stopping conditions are satisfied.

MFS incorporates a filter method in wrapper method to improve classification performance and accelerates the search in identifying core feature subsets. Our GA is based on wrapper method and our local search is based on filter method which it tunes the population of GA solutions by adding or deleting features based on multivariate feature information. Our focus is on filter methods that are able to assess the goodness of one feature in the context of others and finally, each improved salutation validate by classification evaluation in GA algorithm which is a wrapper method.

In fact, from the perspective of search strategies, we have investigated several major issues of MA to identify a good balance between local search and genetic search so as to maximize search quality in the hybrid filter and wrapper MA. We investigate the balance between exploitation and exploration. From the perspective of feature evaluation, we have both benefit of wrapper and filter. For example, in wrapper if we have redundant feature our process is very time consuming. Because of that first, by filter method redundant feature deleted then they send to wrapper method in order to evaluate feature subset accurately.

Proc	edure of MFS					
1	1 Begin					
2	Initialize: Randomly initialize population of feature subset, initialize E and others;					
3	While (stop if condition is not satisfied)					
4	Evaluate fitness of all feature subset encoded in the population;					
5	Find E best feature subset in the population and put them into elite pop;					
6	For (each subset in elite pop )					
7	Perform local search and replace it with new feature subset;					
8	End For					
9	Evaluate fitness of new solutions which is generated by local search;					
10	Select the best solution based on fitness function as global optimum;					
11	Perform evolutionary operators, i.e. selection, crossover, mutation;					
12	End While					
13	End					

Fig. 1: The procedure of MFS.

## 3.1 Chromosome Encoding

In the MFS, encoding solution is a binary string with a length equal to the total number of features. Each bit encodes a single feature. As shown in fig.2, bit "1" implies the corresponding feature is selected and "0" is not so that the length of the chromosome is N.



Fig.2: Solution encoding as a binary bit string.

## 3.2 Fitness Function

To evaluate the goodness of each feature subset which is generated by each chromosomewe use accuracy of classification. It can be defined based on equation (3):

$$Fit = J(fs) \tag{3}$$

Where fs denotes the corresponding selected feature subset encoded in solution, J compute the goodness of feature subset. In this paper, it is accuracy of classification. Note that if two solutions have the same fitness, that which one has smaller feature number is selected.

#### 3.3 Evolutionary Operations

In the evolution process, we use a rank based elitism roulette wheel selection which is based on the fitness of chromosomes. It should ensure that fitter chromosomes have better chance to survive. We use one-point crossover such that if two parent chromosomes  $C_1$  and  $C_2$  are selected, they perform crossover operation with a crossover probability  $P_c$  to generate two new chromosomes  $Off_1$  and  $Off_2$  with exchanging information in a randomly cut point. In mutation operator selects some positions with a mutation probability  $P_m$  randomly and invert genes at these positions.

#### 3.4 Local Search Improvement Procedure

Our local search iterate over each feature of candidate feature subset. At each step, each feature will be deleted or added in order. New feature subset is evaluated, if an improvement is achieved, new feature subset is accepted as the current one. Then the iteration continues with another feature. This process iterates for L (LS steps) times. Fig. 3shows the procedure of our local search procedure.

Pro	redure of Local Search
1	Begin
2	Input: Elite population;
3	Initialize: K;
4	For (each feature subset in elite population(E), E <sub>i</sub> )
5	For (number of K)
6	$\mathbf{E}_{best} = \mathbf{E}_{i;}$
7	Add or delete each feature in E <sub>i</sub> ;
8	calculate filter evaluation of improved feature subset fs using FE(fs);
9	If $(FE(fs) > FE(E_{best}))$
10	$E_{best} = fs;$
11	Else
12	change feature subset in original format;
13	End If
14	Replace E <sub>i</sub> withE <sub>best</sub>
15	End For
16	End For
17	End

Fig. 3: The procedure of our local search method.

## **3.5** Feature Subset Evaluation filter(FE)

In local search there is a filter evaluation function to evaluate solution which is generated during the local search procedure. In our method, for improving our work effectively, we used equation (2) which is a famous subset evaluation filter.

#### 4 Experimental Result and Discussion

In this section, our experimental result is carried out to show the effectiveness of our method. In The following subsections, a brief description of dataset benchmark is given, and then our simulations results and comparison with literature works are presented and discussed.

## 4.1 Database Description and preprocessing

We use 12 benchmark datasets which are frequently used in literatures. Our datasets are from the University of California, Irvine (UCI) repository [27]. Table 1 shows description of these datasets. They are both nominal and numerical data

Since some of these datasets have missing values or continues values in uncontrolled rang, they have a preprocessing step before they are used. For missing values, we replaced them with the most frequently used values for nominal and numeric features. To control the range of continues features we normalize them in rang [0, 1].

#### J. Basic. Appl. Sci. Res., 3(10)126-133, 2013

No.	Database	Ν	Number of instances	Number of classes
1.	Spectf	44	269	2
2.	Lymphography	18	148	4
3.	isolet	617	1559	26
4.	Arrhythmia	279	452	16
5.	Glass	9	214	7
6.	Synthetic	60	600	6
7.	Molecular splice	60	3190	3
8.	Cylinder-bands	36	540	2
9.	Audiology	69	226	24
10.	Dermatology	34	366	6
11.	Musk clean1	166	476	2
12.	Soybean	35	683	19

#### 4.2 Metrics measured

In our method, the following metrics are used:

- **Result validation**: We use accuracy of one-nearest-neighborhood-classifier (1NN) for result validation. Since it is stable and it is not sensitive to the initial setting. Other classifiers like neural networks due to producing different outputs based on initial weight settings may not be proper for performance comparison.

- *K*-Fold Cross Validation: This metric estimate how a predictive model performs accurately and computes the probability of correctly classifying a sample. In *k*-fold cross-validation, all samples are partitioned into *k* subsamples randomly. A single subsample is used as the test data and the remaining k-1 subsamples are used as training data. This process is repeated *k* times, until each of the *k* subsamples used exactly once as the validation data. After that *k* results are averaged to produce a single estimation. In our study, due to being randomness, run 10 times and at each time a 10-fold cross validation which is commonly used is used [33], and the final results were their average values (10-10 fold CV).

#### 4.3 Performance evaluation

In this section, we present an experimental study of MFSon commonly used UCI data sets. We employed a population size of 30 and generation number is 200. Crossover rate, Pc, and mutation rate. Pm, are 0.6 and 0.1, respectively. Table 2 shows the best and average accuracy of 5 runs of MFSon defined databases. Because MFSis a random search algorithm, different results may be obtained at every run. We have run this algorithm on 5 runs and record average of them. Table 2: performance of MFS (1NN, 10-10 fold CV, Unit: %)

		1	/	
No.	Database	Accuracy of Unselected features(%)	Best accuracy(%)	Average accuracies(%)
1	Spectf	82.89	85.39	84.65
2	Lymphography	83.78	87.23	85.34
3	Isolet	85.12	89.22	71.80
4	Arrhythmia	68.80	74.05	73.38
5	Glass	79.90	73.08	72.60
6	Synthetic	98.16	99.22	99.06
7	Molecular splice	74.67	89.75	89.54
8	Cylinder-bands	75.74	77.67	76.69
9	Audiology	74.04	80.53	79.89
10	Dermatology	95.36	97.21	97.05
11	Musk clean1	89.28	92.65	91.66
12	Soybean	91.35	93.87	93.00

#### 4.4 Comparison Of literature Works

We empirically evaluated the performance of MFSby comparing with recently methods, Ref. [35]and Ref. [36].We have compared our method with two typical feature selectors: Relief F and IG. Relief F [37] is a popular instance-based feature weighting algorithms and Information Gain (IG) measures the decrease in entropy when the features are presented [38]. They are all well-known methods and have excellent performance. For ReliefF, 5 neighbors and 30 instances throughout the experiments are used as suggested by Robnik-Sikonja and Kononenko[39], which is also used in the literature [40]. The results of comparisons are reported in Table3. In each row best results are bolded. As we can see in this table, in most cases the presented method (MFS) has better results which are considerable in some databases.

We also illustrate the results of comparison in a bar graph (Fig. 4). As we can see in this figure, the proposed method, MFS, in most cases has better results than the other previous works.



\* Ref. [35]: Spectf, Lymphography, Isolet, Arrhythmia, Glass, Synthetic, and Molecular splice databases Ref. [36]: Cylinder-bands, Audiology, Dermatology, Musk clean1, and Soybean databases.

NO.	Database	MFS		Results obtained from other methods		ReliefF	IG
1	Spectf	MFS+1NN 10-10 fold CV	85.39 (84.65)	CoFS+1NN 10-fold CV Ref.[35]	85.91	85.51	85.51
2	Lymphography	MFS+1NN 10-10 fold CV	87.23 (85.34)	CoFS+1NN 10-fold CV Ref.[35]	79.95	78.00	82.23
3	Isolet	MFS+1NN 10-10 fold CV	89.22 (71.80)	CoFS+1NN 10-fold CV Ref.[35]	66.52	54.65	53.62
4	Arrhythmia	MFS+1NN 10-10 fold CV	74.05 (73.38)	CoFS+1NN 10-fold CV Ref.[35]	71.23	71.20	71.45
5	Glass	MFS+1NN 10-10 fold CV	73.08 (72.60)	CoFS+1NN 10-fold CV Ref.[35]	78.49	69.92	69.92
6	Synthetic	MFS+1NN 10-10 fold CV	99.22 (99.06)	CoFS+1NN 10-fold CV Ref.[35]	92.00	86.17	83.00
7	Molecular splice	MFS+1NN 10-10 fold CV	89.75 (89.54)	CoFS+1NN 10-fold CV Ref.[35]	90.25	82.88	90.25

Fig. 4: The visual comparison of MFS with previous works.

T-1-1-

2.41.

#### J. Basic. Appl. Sci. Res., 3(10)126-133, 2013

8	Cylinder-bands	MFS+1NN 10-10 fold CV	<b>77.67</b> (76.69)	DMIFS+1NN 10-10 fold CV Ref. [36]	77.55	75.79	75.70
9	Audiology	MFS+1NN 10-10 fold CV	80.53 (79.89)	DMIFS+1NN 10-10 fold CV Ref. [36]	74.04	72.94	72.32
10	Dermatology	MFS+1NN 10-10 fold CV	97.21 (97.05)	DMIFS+1NN 10-10 fold CV Ref. [36]	92.18	83.96	81.37
11	Musk clean1	MFS+1NN 10-10 fold CV	92.65 (91.66)	DMIFS+1NN 10-10 fold CV Ref. [36]	87.34	84.10	83.53
12	Soybean	MFS+1NN 10-10 fold CV	93.87 (93.00)	DMIFS+1NN 10-10 fold CV Ref. [36]	88.19	88.40	90.47

#### 5. Conclusion

In this paper, we have proposed a novel method based on MA to find an efficient feature subset. We used filter method as local search heuristic in the MA. The goal of MFS was to improve classification performance and accelerate the search methodology to identify important feature subsets. MFS was based on a heuristic approach which can search the solution space effectively by appropriate exploring and exploiting. Exploration and exploitation are important properties in heuristic algorithms. MFS could do a trade of between exploitation and exploration. Exploration was done by genetic operators and exploitation by our local search. Local search heuristics improve the quality of the candidate solution to produce a better candidate solution at each step. Besides, intensive local search can trap algorithm into local optimum but MFS controls this issue, we have done a study on number of iteration in local search heuristic (parameter k and E). MFS which was a wrapper-filter method was compared to well known methods. Empirical study of MFS on commonly used data sets from UCI data sets indicates that it outperformed recent methods.

#### ACKNOWLEDGMENT

The authors declare that they have no conflicts of interest in the research.

## REFERENCES

- [1] Alpaydin, E., Introduction to Machine Learning 2004: The MIT Press
- [2] Baghshah, M.S. and S.B. Shouraki, *Kernel-based metric learning for semi-supervised clustering*. Neurocomputing 2010. **73** p. 1352-1361.
- [3] Duda, R.O., P.E. Hart, and D.G. stork, *pattern classification*Secon Edition ed. 2001: Wiley.
- [4] Ferreira, A.J. and M.A.T. Figueiredo, *An unsupervised approach to feature discretization and selection*. Pattern Recognition, 2012. **45**(9): p. 3048-3060.
- [5] Doquire, G. and M. Verleysen, A graph Laplacian based approach to semi-supervised feature selection for regression problems. Neurocomputing, 2013, (0).
- [6] Yang, J., et al., A new feature selection based on comprehensive measurement both in inter-category and intracategory for text categorization. Information Processing & Management, 2012. **48**(4): p. 741-754.
- [7] Vieira, S.M., J.M.C. Sousa, and U. Kaymak, *Fuzzy criteria for feature selection*. Fuzzy Sets and Systems, 2012. 189(1): p. 1-18.
- [8] García-Pedrajas, N., A. de Haro-García, and J. Pérez-Rodríguez, A scalable approach to simultaneous evolutionary instance and feature selection. Information Sciences, 2013. 228(0): p. 150-174.
- [9] Hastie, T., R. Tibshirani, and J. Friedman, The Elements of Statistical Learning. Springer, 2001.
- [10] Miller, A.J., Subset Selection in Regression. 0002-Revised ed. English, ed. 2. 2002: CRC Press. 256.
- [11] Dash, M. and H. Liu, *Feature selection methods for classifications*. Intelligent Data Analysis: An International Journal, 1997. **1**(3).
- [12] Doak, J., An evaluation of feature selection methods and their application to computer security. 1992: Davis CA: University of California, Department of Computer Science.
- [13] Liu, H. and H. Motoda, Less Is More, in Feature Extraction, Construction and Selection: A Data Mining PerspectiveH.L.a. H, Editor. 1998, Kluwer Academic. p. 3 12.
- [14] Covões, T.F. and E.R. Hruschka, *Towards improving cluster-based feature selection with a simplified silhouette filter*. Information Sciences, 2011. **181**(18): p. 3766-3782.
- [15] Sun, X., et al., Using cooperative game theory to optimize the feature selection problem. Neurocomputing, 2012.
  97(0): p. 86-93.
- [16] Ferreira, A.J. and M.A.T. Figueiredo, *Efficient feature selection filters for high-dimensional data*. Pattern Recognition Letters, 2012. 33(13): p. 1794-1804.

- [17] Macaš, M., et al., Wrapper feature selection for small sample size data driven by complete error estimates. Computer Methods and Programs in Biomedicine, 2012. 108(1): p. 138-150.
- [18] Foithong, S., O. Pinngern, and B. Attachoo, Feature subset selection wrapper based on mutual information and rough sets. Expert Systems with Applications, 2012. 39(1): p. 574-584.
- [19] Chuang, L.Y.; Yang, C.H.; Li, J.C., Chaotic maps based on binary particle swarm optimization for feature selection. Applied Soft Computing Journal 2011. 11 p. 239-248
- [20] Guyon, I., et al., Feature Extraction Foundations and Applications. 2006: Springer.
- [21] Yu, L. and H. Liu, Efficient Feature Selection via Analysis of Relevance and Redundancy. Journal of Machine Learning Research, 2004: p. 1205–1224.
- [22] Langley, P. Selection of relevant features in machine learning. in AAAI Fall Symposium on Relevance. 1994: AAAI Press.
- [23] O'Brien, R.F.J., Ant Algorithm Hyperheuristic Approaches for Scheduling Problems. 2007, Nottingham.
- [24] Ang, J.H., K.C. Tan., and A.A. Mamun, An evolutionary memetic algorithm for rule extraction Expert Systems with Applications 2010. 37 p. 1302–1315
- [25] Moscato, P., On Evolution, Search, Optimization, Genetic Algorithms and Martial Arts: Towards Memetic Algorithms, in Tech. Rep. Caltech Concurrent Computation Program, Report. 1989, California Institute of Technology: Pasadena, California, USA.
- [26] Merz, P. and B. Freisleben. A comparison of memetic algorithms, Tabu search, and ant colonies for the quadratic assignment problem. in congress on evolutionary computation 1999.
- [27] D. J. Newman, S.H., C. L. Blake, and C. J. Merz, UCI repository of machine learning databases. 1998.
- [28] Press, W.H., et al., *Numerical recipes in C. The art of scientific computing*. 1988, Cambridge, UK: Cambridge University Press.
- [29] Ghiselli, E.E., Theory of psychological measurement. 1964, New York: McGraw-Hill Education.
- [30] Hogarth, R.M., *Methods for aggregating opinions*. Decision making and change in human affairs 1977: Dordrecht, Holland: Reidel Publishing Co.
- [31] Hall, M.A., *Correlation-based Feature Subset Selection for Machine Learning*, in *Department of Computer Science*. 1999, University of Waikato: Waikato, N.Z.
- [32] Amaldi, E. and V. Kann, On the approximation of minimizing non zero variables or unsatisfied relations in linear systems. Theoretical Computer Science 1998. 2009 p. 237–260.
- [33] GJ, M.K.A.D., C. Ambroise Analyzing microarray gene expression data. 2005: Wiley Series in Probability and Statistics.
- [34] Ong Y. S., Keane A. J., Meta-Lamarckian Learning in Memetic Algorithms, IEEE Transaction On Evolutionary Computation, VOL. 8, NO. 2, April 2004.
- [35] Sun, X., Liu, Y., Li, J., Zhu, J., Chen, H., & Liu, X, Feature evaluation and selection with cooperative game theory, pattern recognition, 2012.
- [36] LiuH., Sun J., Liu L., Zhang H., Feature selection with dynamic mutual information. Pattern Recognition 42(7): 1330-1339 2009.
- [37] Kononenko, I. , *Estimating attributes: analysis and extensions of relief*, in: Proceedings of the 1994 European Conference on Machine Learning, 1994, pp. 171–182.
- [38] Lee, C., Lee, G.G., Information gain and divergence based feature selection for machine learning based text categorization, Information Processing & Management 42 (1) (2006) 155–165.
- [39] M. Robnik-Sikonja, I. Kononenko, Theoretical and empirical analysis of ReliefF and RReliefF, Machine Learning 53 (1) 23–69 2003.
- [40] Yu, L., Liu, H., Efficient feature selection via analysis of relevance and redundancy, Journal of Machine Learning Research 5 1205–1224 2004.