OPEN ACCESS

International
Journal
Of Modern Engineering Research (IJMER)

A Threshold fuzzy entropy based feature selection method applied in various benchmark datasets using Ant-miner algorithm

P. Devipriya[1], P.Divya[2], R.Srivinitha[3], J.Joshimadevi[4], S.Saravanan[5], R. G. Suresh Kumar[6]

[1,2,3,4] Department of Computer Science, Rajiv Gandhi College of Engineering and Technology, Pondicherry, India

[5,6] Department of Computer Science, Rajiv Gandhi College of Engineering and Technology, Pondicherry, India

ABSTRACT: *Large amount of data have been stored and manipulated using various database technologies. Processing all the attributes for the particular means is the difficult task. To avoid such difficulties, feature selection process is processed.In this paper,we are collect a eight various benchmark datasets from UCI repository.Feature selection process is carried out using fuzzy entropy based relevance measure algorithm and follows three selection strategies like Mean selection strategy,Half selection strategy and Neural network for threshold selection strategy. After the features are selected, they are evaluated using Radial Basis Function (RBF) network,Stacking,Bagging,AdaBoostM1 and Ant-miner classification methodologies.The test results depicts that Neural network for threshold selection strategy works well in selecting features and Ant-miner methodology works best in bringing out better accuracy with selected feature than processing with original dataset.The obtained result of this experiment shows that clearly the Ant-miner is superiority than other classifiers.Thus, this proposed Ant-miner algorithm could be a more suitable method for producing good results with fewer features than the original datasets.*

Keywords: *Ant Colony Optimization (ACO), AdaBoostM1, Ant-miner, Bagging, Radial Basis Function (RBF), Stacking*

I. Introduction

Data mining, *the extraction of hidden predictive information from large databases*, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools [3] predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining [8,9] move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve. It is interactive and iterative, involving the following steps:

- *Step 1.* Application domain identification: Investigate and understand the application domain and the relevant prior knowledge. In addition, identify the goal of the KDD from the administrators' or users' point of view.
- *Step 2.* Target dataset selection: Select a suitable dataset, or focus on a subset of variables or data samples where data relevant to the analysis task are retrieved from the database.
- *Step 3.* Data Preprocessing: the DM basic operations include 'data clean' and 'data reduction': In the 'data clean' process, we remove the noise data, or respond to the missing data field. Inthe 'data reduction' process, we reduce the unnecessary dimensionality or adopt useful transformation methods. The primary objective is to improve the effective number of variables under consideration.
- *Step 4.* Data mining: This is an essential process, where AI methods are applied in order to search for meaningful or desired patterns in a particular representational form, such as association rule mining, classification trees, and clustering techniques.
- *Step 5.* Knowledge Extraction: Based on the above steps it is possible to visualize the extracted patterns or visualize the data depending on the extraction models. Besides, this process also checks for or resolves any potential conflicts with previously believed knowledge.

- *Step 6*. Knowledge Application: Here, we apply the found knowledge directly into the current application domain or in other fields for further action.
- *Step 7*. Knowledge Evaluation: Here, we identify the most interesting patterns representing knowledge based data on some measure of interest. Moreover, it allows us to [2] improve the accuracy and efficiency of the mined knowledge.

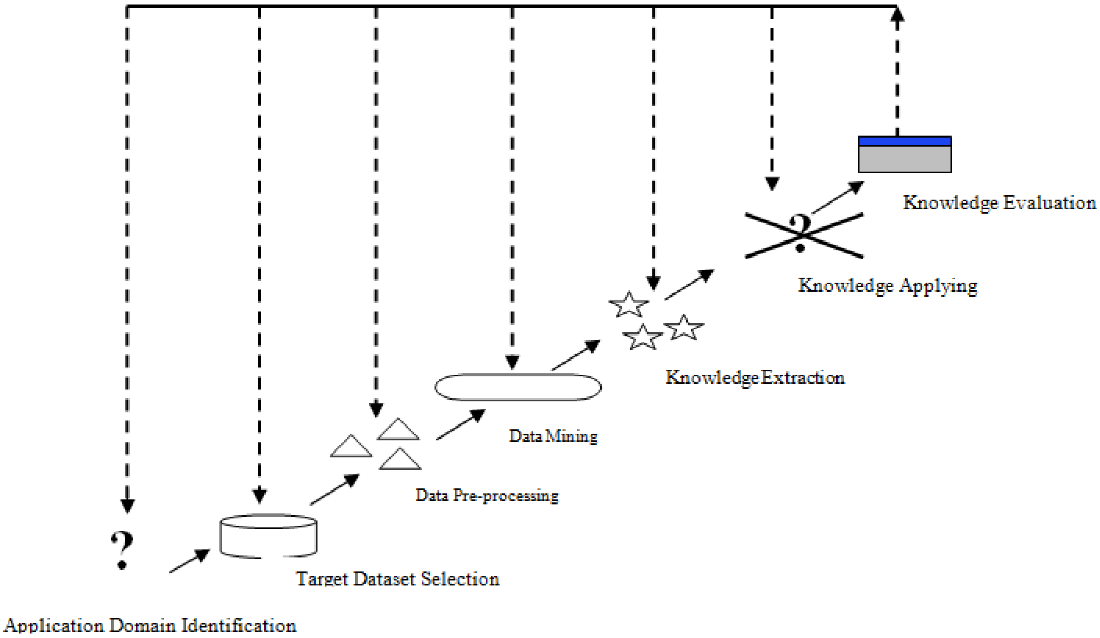


Fig.1. Data mining phases

The revolution in database technologies has resulted in an increase of data accumulation in many areas, such as financial, marketing, sports and the biological and [29] medical sciences. Feature selection has many advantages such as shortening the number of measurements, reducing the execution time and improving transparency and compactness of the suggested diagnosis.

1.1 Feature selection

Feature selection is the process of selecting a subset of'd' features of the set D, such that d<D. The primary purpose of feature selection is to reduce the computational cost and to improve the performance of the learning algorithm. Feature selection algorithms [30, 31, 32] deal with different evaluation criteria and generally, are classier into filter and wrapper models[31]. The filter model evaluates the general characteristics of the training data to select a feature subset without relation to any learning algorithms; thus, it is computationally economical.

1.2 Feature selection strategies

This subsection explains three different criteria for the feature selection process. The features are regulated with respect to decreasing values of the fuzzy entropy. A feature in the first position is the most relevant and the one in the last position is the least relevant in the resulting rank vector.

Mean Selection (MS) Strategy: A feature $f \in F$ is selected if it satisfies the following condition:

$$\sigma(f) \geq \sum_{f \in F} \frac{\sigma(f)}{|F|}$$

Where $\sigma(f)$ is the relevance value of the features, which is selected if it is greater than or equivalent to the mean of the relevant values. This strategy will be useful in examining the suitability of the fuzzy entropy relevance measure.

Half Selection (HS) Strategy: The half selection strategy aims to reduce feature dimensionality to select approximately 50% of the features in the domain. The feature $f \in F$ is selected if it satisfies the following condition:

$$p_a \geq \frac{|F|}{2}$$

Where p_a is the position of the feature in the rank vector. It represents the selected features having a relevance value higher than a given threshold, which is calculated as |F|/2 . This strategy does produce great reductions, close to 50%. At the same time, some of the selected features are irrelevant despite them passing the threshold.
Neural Network for Threshold Selection (NNTS): an ANN is one of the well-known machine learning techniques and it can be used in a variety of applications in data mining. The ANN provides a variety of feed forward networks that are generally called back propagation networks. The output value can be considered as a threshold value of the given fuzzy entropy.

II. Literature Review

Recently ,a number of researchers have focused on several feature selection methods and most of them have reported their good performance in database classification. Battiti [7] proposes a method called Mutual-Information-based Feature Selection(MIFS),in which the selection criterion is based on maximizing the mutual information between candidate features and the class variables and minimizing the redundancy between candidate features and the selected features. Hanchuan et al. [8] follow a similar technique to MIFS ,which has been called the minimal-redundancy-maximal-relevance(mRMR)criterion.

It eliminates them annually tuned parameter with cardinality of the features already selected. Pabloetal. [9] present a Normalized Mutual Information Feature Selection algorithm. YuandLiu [10] developed a correction-based method for relevance and redundancy analysis and then removed redundant features using the Markov Blanket method. In addition ,features election methods are analyzed by a number of techniques. Abdel-Aal [1] developed a novel technique for feature ranking and selection with the group method of data handling. Feature reduction of more than 50% could be achieved and improved in the classification performance. Sahanetal. [11] built a new hybrid machine learning method for a fuzzy-artificial immune system with ak-nearest neighbor algorithm to solve medical diagnosis problems, which demonstrated good results. Jaganathanetal. [12] applied a new improved quick reduct algorithm, which is a variant of quick reduct for feature selection and test editon a classification algorithm called Ant Miner. Sivagami- nathanetal. [13] proposed a hybrid method combining Ant Colony Optimization and Artificial Neural Networks(ANNs)to deal with feature selection, which produced promising results. Linetal. [14] proposed Simulated Annealing approach for parameter setting in Support Vector Machines, which is compared with a grid search parameter setting and was found to produce higher classification accuracy. Lin etal. [15] applied a Particle-Swarm-Optimization-based approach to search for appropriate parameter values for a back-propagation network to select the most valuable subset of features to improve classification accuracy. Unleretal. [16] developed a modified discrete particle swarm optimization algorithm for the feature selection problem and compared it with tabuand scatter search algorithms to demonstrate it effectiveness. Changetal. [17] introduced a hybrid model for integrating a case-based reasoning approach with a particle swarm optimization model for feature subset selection in medical database classification. Salamoetal. [18] evaluated a number of measures for estimating feature elevance based on rough set theory and also proposed three strategies for feature selection in a Case Based Reasoning classifier. Qasemetal. [19] applied a time variant multi- objective particle swarm optimization to an RBF Network for diagnosing medical diseases. The use of discriminant analysis is ever increasing (Wilson and Sharda, 1994; Altman, Haldeman, and Narayanan, 1977).The Altman Z-score model is predominately used in discriminant analysis (Jo, Han and Lee, 1997).

Work	Classifier	No.of datasets
R.E.Abdel-Aal(2005)	GMDH	6
M.F.Akay(2009)	SVM	2
Chin-Yuan Fan(2011)	HM	1
Huan Liu(2005)	IFSA	4
R.Battiti(1994)	NN	4
Pablo A.Extevez(2009)	NN	2
Seral Sahan(2007)	K-nn	3
P.Jagaathan(2007)	ACO	1
Rahul Karthik Sivagaminathar (2007)	ACO	1
Shih-Wei Lin(2008)	SVM	4

Zne-Jung Lee(2008)	SVM	4
Tsung-Yuan Tseng(2008)	SVM	4
Alper Unler(2010)	Swarm optimization	3
Pei-Chann Chang(2012)	Swarm optimization	3
Maria Salamo(2011)	Feature selection	2
Sultan Noman Qasem(2011)	RBF	1
H.M.Lee(2001)	Feature selection	3
S.Rajasekaran(2010)	NN	
Resul Das(2009)	NN	4
Hanchuan Peng (2005)	Feature selection based on mutual information	8
Chin-Yuan Fan(2011)	hybrid model	1
Yang C, Sudderth J(2009)	Akt signaling	69
Y. Saeys, I. Inza,(2007)	Feature selection	19
Parpinelli, R(2002)	ant colony optimization	4
Ooghe, H. and Spaenjers(2010)	BFPM	1
Fuellerer, G(2009)	Ant colony optimization	3
Jia Yu, Yun Chen(2011)	ant colony optimisation	35
Gao X, Bu D(2009)	*BMC Structural Biology*	1
Stoothoff W(2009)	*Neurochem*	11
Brown G,(2012)	Feature selection	27
Dramiski M,(2008)	Feature selection	110
Sun X, Liu Y, Li(2012)	Feature evalution	45

SVM= Support Vector Machine,DT= Decision Tree,LR=Logistic Regression,NN=Neural Networks,UCI repository machine,ACO=Ant Colony Optimization,RBF=Radial Bias Function,GMDH=Group Method of Data Handling based future ranking &selection method,IFSA=Integrating Feature Selection Algorthim,HM=Hybrid Model,BFPM=Business Focused Project Model.

III. Background Details

3.1 Fuzzy Entropy Relevant Measure

In information theory, the Shannon entropy measure is generally used to characterize the impurity of a collection of samples. Assuming X as a discrete random variable with a finite set of n elements, where X={x1, x2, x3,…,xn}, then if an element xi occurs with probability $p(x_i)$ [20], the entropy H(X) of X is defined as follows:

$$H(X) = -\sum_{i=0}^{n} p(x_i) log_2 p(x_i)$$

where n denotes the number of elements.

An extension of Shannon entropy with fuzzy sets, which is to support the evaluation of entropies, is called fuzzy entropy. The proposed fuzzy entropy method is based on the utilization of the Fuzzy C-Means Clustering algorithm (FCM), which is used to construct the membership function of all features. The data may belong to two or more clusters simultaneously and the belonging of a data point to the clusters is governed by the membership values [24]. The FCM algorithm is explained as follows [25].

Step 1: assume the number of clusters (C), where $2 \le c \le N$, C - number of clusters and N - number of data points

Step 2: calculate the jth cluster center C_j using the following expression

$$C_j = \frac{\sum_{i=1}^{N} \mu_{ij}^{g} \ X_{ij}}{\sum_{i=1}^{N} \mu_{ij}^{g}}$$

where g≥1 is the fuzziness coefficient and μij is the degree of membership for the ith data point X_i in cluster j.

Step 3: calculate the Euclidean distance between the ith data point and the jth cluster center as follows:

$d_{ij} = |C_j\text{-}x_i|$

Step 4: update the fuzzy membership values according to d_{ij}. If $d_{ij} \ge 0$, then

$$\mu = \frac{1}{\sum_{m=1}^{C}\left(\frac{d_{ij}}{d_{im}}\right)^{\frac{2}{g-1}}}$$

d=0, then the data point coincides with the jth cluster center (C) and it will have the full membership value, i.e.,$\mu_{ij} = 1.0$

Step 5: repeat Steps 2–4 until the changes in [μ] are less than some pre-specified values.

The FCM algorithm computes the membership of each sample in all clusters and then normalizes it. This procedure is applied for each feature. The summation of membership of feature 'x' in class 'c', divided by the membership of feature 'x' in all 'C' classes, is termed the class degree $CD_c(\tilde{A})$[26], which is given as

$$CD_c(\tilde{A}) = \frac{\sum_{x \in c}\mu_{\tilde{A}}(x)}{\sum_{x \in C}\mu_{\tilde{A}}(x)}$$

where$\mu_{\tilde{A}}$ denotes the membership function of the fuzzy set and $\mu_{\tilde{A}}(X_i)$ denotes the membership grade of x belonging to the fuzzy set $\tilde{A}$.

The fuzzy entropy $FE_c(\tilde{A})$ of class 'c' is defined as

$$FE_c(\tilde{A}) = -CD_c(\tilde{A})\log_2 CD_c(\tilde{A})$$

The fuzzy entropy $FE(\tilde{A})$ of a fuzzy set X is defined as follows:

$$FE(\tilde{A}) = \sum_{c \in c} FE_c(\tilde{A})$$

The probability p(x_i) of Shannon's entropy is measured by the number of occurring elements. In contrast, the class degree $CD_c(\tilde{A})$ in fuzzy entropy is measured by the membership values of the occurring elements. Suppose the elements are divided into a number of intervals: I_1, I_2, then the Shannon entropy of interval I_1 is equal to the interval I_2. However, fuzzy entropy can indicate that interval I_1 is distinguishable from interval I_2. The difference between Shannon's entropy and the proposed fuzzy entropy is illustrated in Fig. 1, where the symbols 'M' and 'F' denote male and female samples, respectively. The computation of Shannon's P.

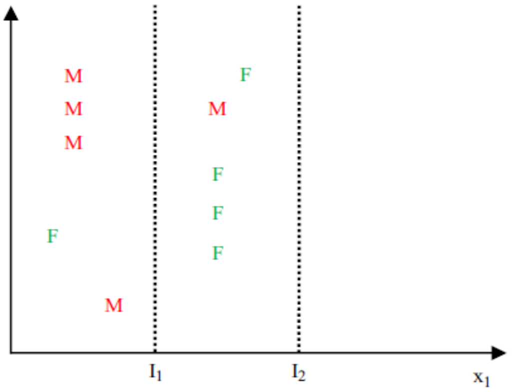


Fig 1. Distribution of samples

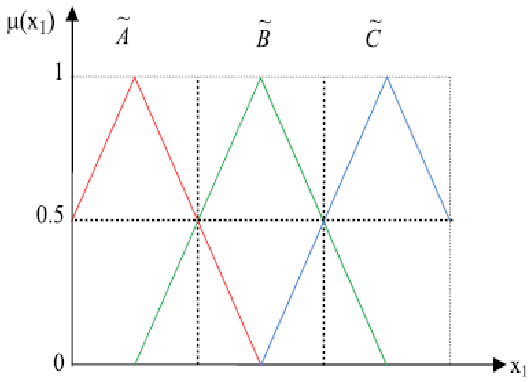


Fig 2. A feature $x_1$with fuzzy sets $\tilde{A}, \tilde{B}$ *and* $\tilde{C}$

entropy in the intervals I_1 and I_2 is as follows.

$$H(I_1) = -(p(M)\log_2 p(M) + p(F)\log_2 p(F))$$

$$H(I_1) = -\left(\frac{4}{5}\log_2\frac{4}{5} + \frac{1}{5}\log_2\frac{1}{5}\right) \cong 0.72$$

$$H(I_2) = -\left(\frac{1}{5}\log_2\frac{1}{5} + \frac{4}{5}\log_2\frac{4}{5}\right) \cong 0.72$$

Based on Eq. (5), the class degrees of 'M' and 'F' are

$$CD_M(\tilde{A}) = \frac{3.75}{3.75 + 0.75} = 0.833$$

$$CD_F(\tilde{A}) = \frac{0.75}{3.75 + 0.75} = 0.167$$

The fuzzy entropy FE($\tilde{A}$) of a fuzzy set $\tilde{A}$ is calculated as follows:

$$FE(\tilde{A}) = FE_M(\tilde{A}) + FE_F(\tilde{A})$$

$$FE(\tilde{A}) = -\left(CD_M(\tilde{A})\log_2 CD_M(\tilde{A}) + CD_F(\tilde{A})\log_2 CD_F(\tilde{A})\right)$$

$$FE(\tilde{A}) = -(0.833\log_2 0.833 + 0.167\log_2 0.167) \cong 0.651$$

Similarly, the fuzzy entropy of the fuzzy set $\tilde{B}$ is as follows:

$$\sum_{x=M} \mu_{\tilde{B}}(x) = 1 + 0.25 = 1.25$$

$$\sum_{x=F} \mu_{\tilde{B}}(x) = 1 + 1 + 1 + 0.75 = 3.25$$

The class degree of 'M' is 1.25/(1.25+3.75)=0.25 and 3.75/ (1.25+3.75)=0.75.
The fuzzy entropy FE($\tilde{B}$) is

$$FE(\tilde{B}) = -(0.25log_2 0.25 + 0.75log_2 0.75) \cong 0.811$$

From the above results, Shannon's entropy of interval I_1 is equal to interval I_2. Nonetheless, the fuzzy entropy indicates that interval I_1 is distinct from interval I_2. The highest fuzzy entropy value of the feature is regarded as the most informative one.

IV. Workflow Of Ant-Miner Algorithm

The flowchart of the ant-miner algorithm is given below. The algorithm starts with obtaining a training set which consists of training cases. After that, the main loop will be executed to discover one rule per iteration: (1) It begins with initializing the index of [11,12,13,14] ant (t), the index of converge, ConvergeIndex (j) which is used to test convergence of ants paths and pheromone on all trails. Convergence of path is an important indicator to check if a steady path chosen by ant colony has formed. It tests if ants take the same path one after another and record how many ants take this path. (2) A sub-loop is executed to discover classification rules by a number of artificial ants (No_of_ants) who explore paths in turn. And the discovery process consists of three main steps: rule generation, rule pruning and pheromone updating.

The sub-loop will terminate under the condition that all ants have taken their exploration (t≥No_of_ants), or the convergence state has been reached (j≥No_rule_converge). No_rule_converge is the threshold of ConvergIndex (j). It means if there are No_rule_converge ants take the same path one after another, the path is qualified for a discovered rule candidate. If the current ant has constructed a rule that is exactly the same as the rule constructed by previous ants, then it is said the ants have converged to a single rule (path) and the value of ConvergeIndex (j) will increase by 1. (3) The main loop selects the best rule from the discovered rules according to their qualities. (4) The training cases which are covered by the best rule need to be removed from the training set. In other words, the number of training cases in the training set is gradually decreased with continuous matching with the best rules.

4.1 Datasets Description

The proposed system is evaluated by eight benchmarks datasets: diabetes, tic-tac-toe, chess, car, hepatitis, fiber, liver disorder, heart statalog which is taken from various fields life, game, industry of UCI Machine learning repository.

Dataset	Area	No. of instances	No. of features	No. of classes
Diabetes	Life	768	8	2
Tic-tac-toe	Game	958	9	2
Chess	Game	42	6	2
Car	industry	1728	6	2
Hepatitis	Life	155	19	2
Fiber	Industry	48	4	2
Liver disorder	Life	345	7	2
Heart statlog	Life	270	13	2

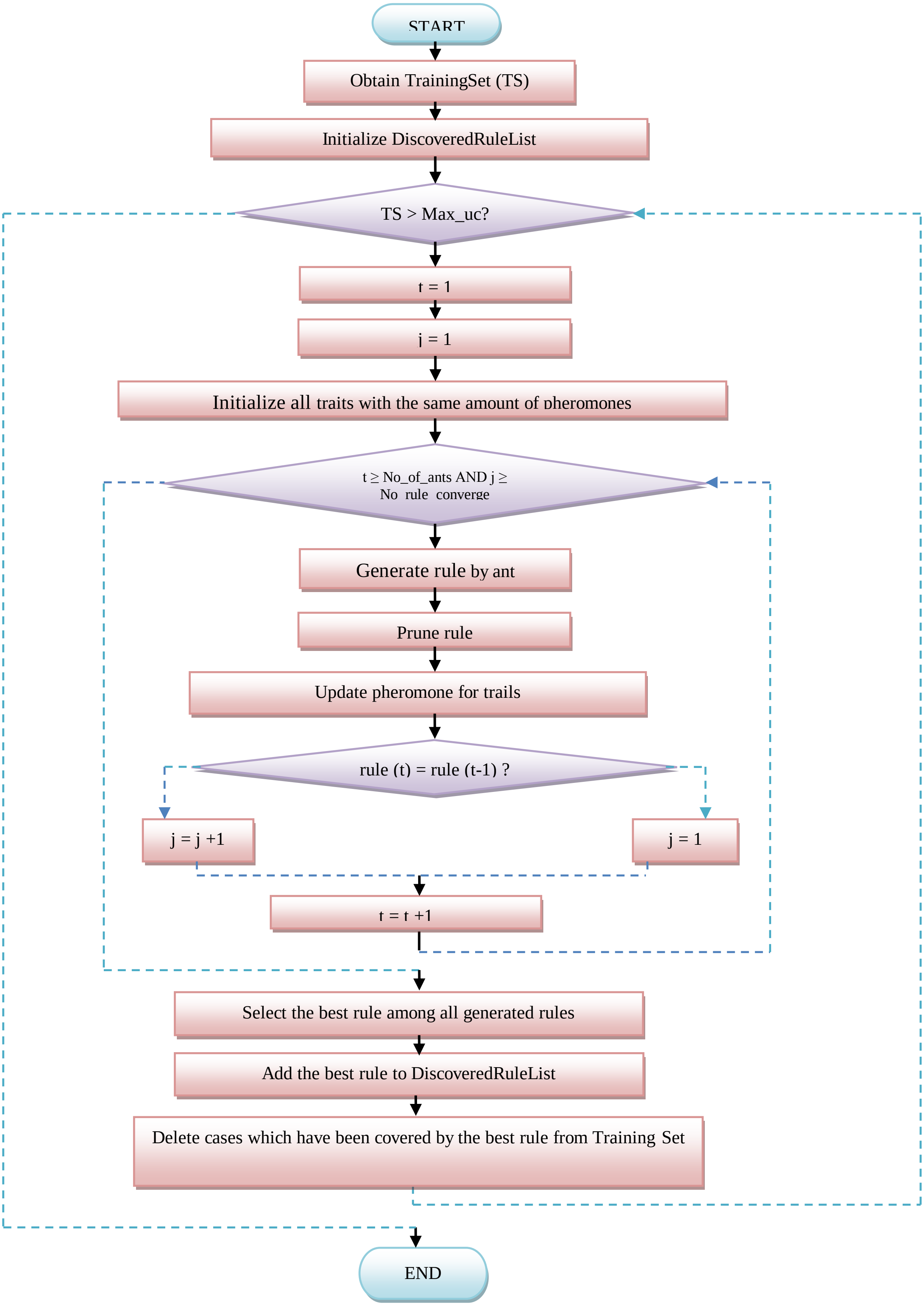


Fig.4. Work flow of Ant-Miner Algorithm for classification rule discovery

V. Results and Discussions

The classification performance of the proposed feature selection method is measured by using an RBF Network classifier, bagging, AdaBoostM1 and stacking. An RBF network is a type of ANN, which is a simple network structure with better approximation capabilities.

5.1 Pima Indians Diabetes

It contains 768 instances described by eight features used to predict the presence or absence of diabetes. The features are as follows: (1) number of pregnancies, (2) plasma glucose concentration, (3) diastolic blood pressure, (4) triceps skin fold thickness,(5) serum insulin, (6)body mass index, (7)diabetes pedigree function and (8)age in years. It is found that Ant-Miner yields the highest accuracy of 99.73 when the[3] Neural Network for Threshold Selection strategy employed which is shown in table result1 .

5.2 Hepatitis

There are 19 features (age, sex, steroid, antivirals, fatigue, malaise, anorexia, liver big, liver film, spleen palpable, spiders, ascites, varices, bilirubin, alk phosphate, SGOT, albumin, protime and histology). It is found that Ant-Miner yields the highest accuracy of 98.70 when Neural Network for Threshold Selection strategy was performed which is shown in table result1.

5.3 Heart-Statlog

It is described by 13 features (age, sex, chest, resting blood pressure, serum cholesterol, fasting blood sugar, resting electro cardio graphic, maximum heart rate, exercise induced angina, old peak, slope, number of major vessels and thal). It is found that Ant-Miner yields the highest accuracy of 99.25 when the Neural Network for Threshold Selection strategy was employed which is shown in table result1.

5.4 Liver-disorder

This dataset contains mcz, alkphos, sgpt, sgot, gammagt, drinks as 6 attributes for 187 instances with two classes 1 or 2. It is found that Ant-Miner yields the highest accuracy of 98.93 when Neural Network for Threshold Selection strategy was performed in table result1.

5.5 Tic-tac-toe

This dataset contains 51 instances with 9 attributes like top-left-square, top-middle-square, middle-left-square, middle-middle-square, middle-right-square, bottom-left-square, bottom-middle-square and bottom-right-square with two classes. It is found that Ant-Miner yields the highest accuracy of 97.41 in Neural Network for Threshold Selection strategy which is shown in table result1.

5.6 Chess

This dataset consist of 6 attributes namely 1)White_king_file, (2)White_king_rank (3)White_rook_file, (4) White_rook_rank, (5) Black_king_file ,(6) Black_king_rank and two classes like win or lose for 42 instances. It is found that Ant-Miner yields the highest accuracy of 97.61 in Neural Network for Threshold Selection strategy which is shown in table result2.

5.7 Car

This contains 1117 instances with 6 attributes like buying, maint, doors, persons, lug_boot and safety with two classes. It is found that Ant-Miner yields the highest accuracy of 97.61 in Neural Network for Threshold Selection strategy which is shown in table result2.

5.8 Fiber

The classification result for the Car dataset is shown in Table8. The classified sensitivity of 94.11, specificity of 96.77 is arhieved with the Neural Network for Threshold Selection strategy in Ant-Miner. This dataset contains two classes yes or no with 4 attributes. The attributes considered are cracker, diet, and subject and digested. It is found that Ant-Miner yields the highest accuracy of 97.91 in Neural Network for Threshold Selection strategy shown in table result2.

The result obtained after processing with several strategies and classification methodologies the Accuracy can be shown in table

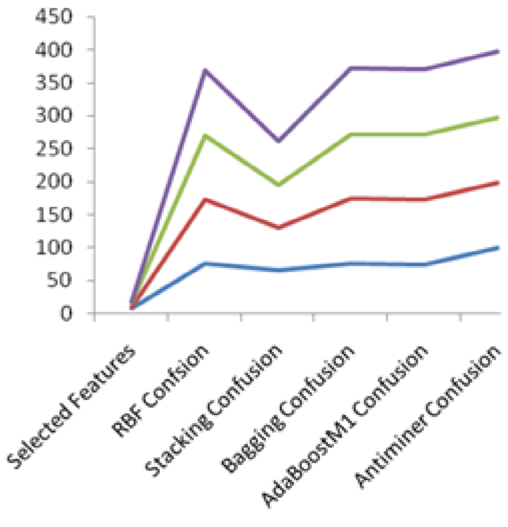
Table result 1

Sl. no	Dataset Name	Strategy	Selected Features	RBF confusion	Stacking confusion	Bagging confusion	AdaBoostM1 confusion	Ant-Miner confusion
1.	Diabetes	None	All	75.39	65.10	75.39	74.34	**99.60**
		MS	2,3	98.30	65.10	99	**99.86**	99.47
		HS	1,4,5,6, 7,8	96.09	65.10	98	**99.86**	98.17
		NNTS	2,3	98.30	65.10	99	99.26	**99.73**
2.	Hepatitis	None	All	85.80	79.35	81.29	82.58	**98.06**
		MS	1,15,16, 18	94.65	57.21	**97.42**	96.79	97.41
		HS	2,3,4,5, 6,7,8,9, 10,11,12, 13,14,17, 19	94.65	57.21	**97.32**	96.79	96.12
		NNTS	1,15,16, 18	84.51	79.35	89.67	89.03	**98.70**
3.	Heart Stat log	None	All	84.07	55.55	81.11	80	**98.88**
		MS	1,4,5,8	96.51	55.55	97.62	97.62	**97.77**
		HS	2,3,6,7, 9,10, 11,12,13	98.51	55.55	**99.62**	**99.62**	97.03
		NNTS	1,4,5,8	98.51	55.55	**99.62**	**99.62**	99.25
4.	Liver Disorder	None	All	63.10	57.21	70.88	67.91	**97.32**
		MS	1,2	94.65	57.21	**97.90**	96.79	97.86
		HS	3,4,5,6	94.65	57.21	**98.41**	96.79	98.39
		NNTS	1,2,3,4	95.72	57.21	97.32	96.79	**98.93**
5.	Tic-Tac-Toe	None	All	90.07	56.86	90.11	89.41	**90.19**
		MS	2,4,5,7	91.07	56.86	**92.20**	92.03	92.15
		HS	1,3,6,8, 9	88.23	56.86	**94.62**	93.03	94.11
		NNTS	All	84.51	79.35	89.67	89.03	**97.41**

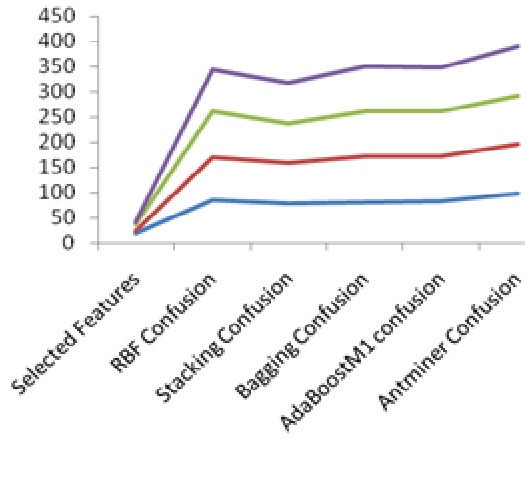
Table result 2

Sl. no	Dataset Name	Stratey	Selected Features	RBF confusion	Stacking confusion	Bagging confusion	AdaBoostM1 confusion	Ant-Miner confusion
6.	Chess	None	All	**93.52**	71.42	93.15	92.60	92.85
		MS	1,46	88.09	71.42	94.75	**95.85**	95.23
		HS	2,3,5	88.09	71.42	**95.75**	94.85	95.23
		NNTS	4,6	88.09	71.42	96.52	96.23	**97.61**
7.	Car	None	All	90.06	68.66	**98.90**	95.16	98.83
		MS	3,4	**98.95**	68.66	96.50	96.50	98.92
		HS	1,2,5,6	94.89	68.66	**97.96**	96.50	97.94
		NNTS	All	88.09	71.42	96.92	96.92	**97.61**
8.	Fiber	None	All	56.25	64.58	56.25	64.58	**93.75**
		MS	4	91.66	64.58	**95.91**	94.91	95.83
		HS	1,2,3	91.66	64.58	92.91	**93.91**	93.75
		NNTS	3,4	91.66	64.58	95.91	95.91	**97.91**

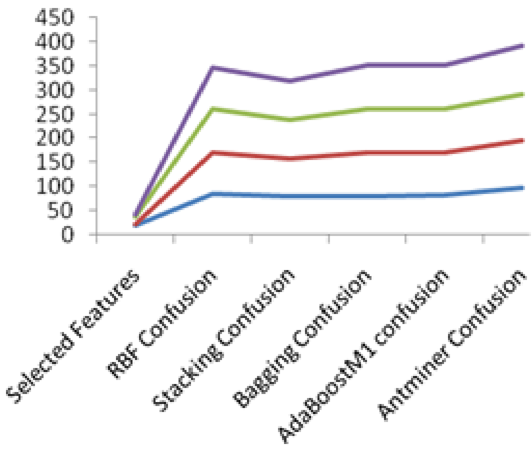
Fig.4. Work flow of Ant-Miner Algorithm for classification rule discovery (Jia Yu et al., (2011)



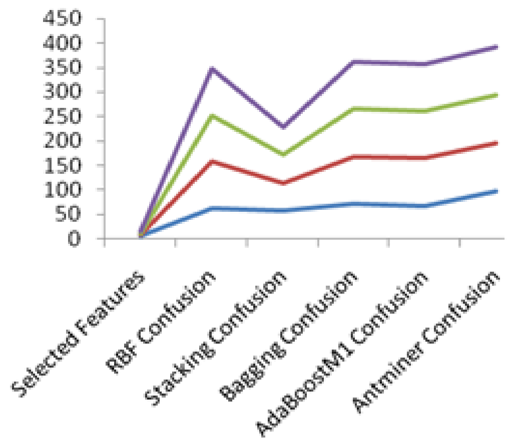
Diabetes



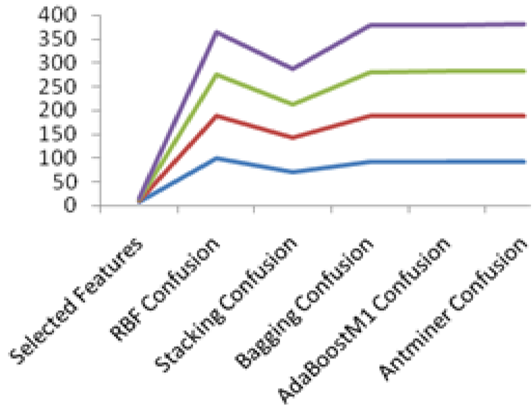
Hepatitis



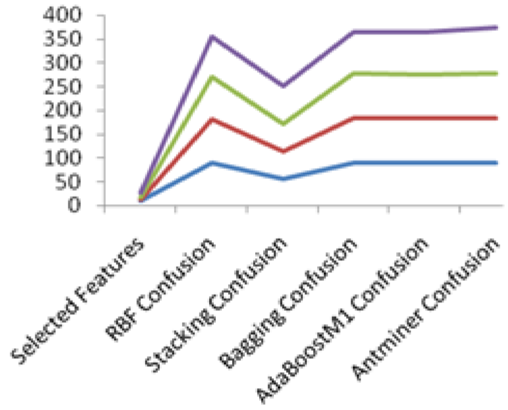
Heart-statlog



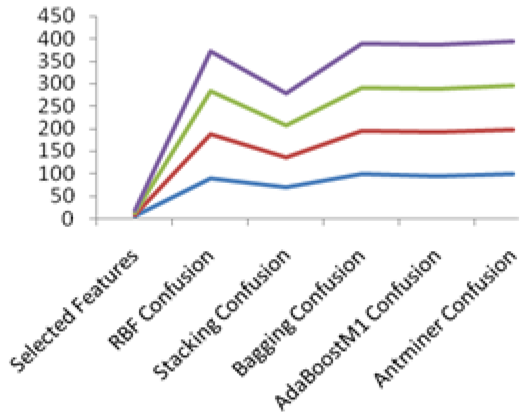
Liver-Disorder



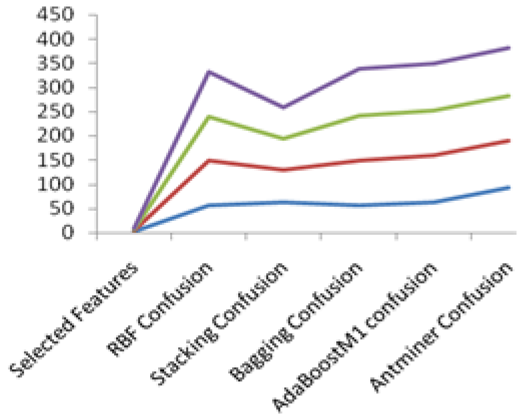
chess



Tic-tac-toe



Car



Fiber

Where,

NNTS HS MS None

This paper demonstrated the Ant-miner based data mining approach to discover decision rules from experts' decision process. This study is the first work on Ant-miner for the purpose of discovering experts.Four data mining techniques which are Radial Basis Function (RBF) network,Stacking,Bagging,AdaBoostM1 and Ant-miner classification methodologies, are applied to compare their performance with that of the ant-miner method. The results of NNTS shows the better accuracy between experts and data-mining techniques ANT-MINER method is perfectly matched and significantly better than RBF network,Stacking,Bagging and AdaBoostM1 methods.

VI. Conclusion

Data mining has been widely applied to various bench mark datasets. This work proposes a new method of classification rule discovery for accuracy prediction by using ant-miner algorithm.In performance terms, the techniques obtain different results depending on the performance factors chosen. Some techniques generate less rules and shows poor accuracy, but Ant-miner provides more rules and gives better predictive accuracy when compare to other techniques. This study has conducted a case study using public dataset for analysis and benchmark approval from UCI repository.Finally this paper suggests that Ant-miner could be a more suitable method than the other classifiers like Radial Basis Function (RBF) network,Stacking,Bagging,AdaBoostM1 and Ant-Miner.In future research, additional artificial techniques could also be applied. And certainly researchers could expand the system with more dataset.

REFERENCES

[1] Chen, M.-C., & Huang, S.-H. *Credit scoring and rejected instances reassigning through evolutionary computation techniques.* Expert Systems with Applications, 24(4) (2003) 433–441.

[2] M.Zmijewski, Methodological issues related to the estimation of financial distress prediction models, *Journal of Accounting Research* 22 (1) (1984) 59–82.

[3] J. Kennedy, R.C. Eberhart, *Particle swarm optimization, in: Proceedings of the IEEE International Conference on Neural Network*, vol. 4, 1995, 1942–1948.

[4] Tam, K., & Kiang, M. Managerial applications of neural networks:The case of bank failure prediction. Management Science, 38(7) (1992) 926–947.

[5] Etheridge, H., & Sriram, R. A comparison of the relative costs of financial distress models: Artificial neural networks, logit and multivariate discriminant analysis. Intelligent Systems in Accounting, Finance and Management, 6,(1997) 235–248.

[6] Lee, T.-S., Chiu, C.-C., Chou, Y.-C., & Lu, C.-J. Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. Computational Statistics and Data Analysis, 50 (2006) 1113–1130.

[7] Salchenberger, L., Cinar, E., & Lash, N. Neural networks: A new tool for predicting thrift failures. Decision Sciences, 23 (1992) 899–916. .

[8] Han, J., & Kamber, M. Data mining: Concepts and techniques. San Francisco,CA, USA: Morgan Kaufmann. (2001).

[9] Huang, M. J., Chen, M. Y., & Lee, S. C. *Integrating data mining with casebased reasoning for chronic diseases prognosis and diagnosis*. Expert Systems with Applications, 32(3) (2007) 856–867.

[10] Brachman, R. J., Khabaza, T., Kloesgen, W., Piatesky-Shpiro, G., & Simoudis, E. Mining business databases. Communication of the ACM, 39(11) (1996) 42–48.

[11] Alupoaei, S., & Katkoori, S. Ant colony system application to marcocell overlap removal. *IEEE Transactions Very Large Scale Integration* (VLSI) Systems, 12(10) (2004)1118–1122.

[12] Bianchi, L., Gambardella, L. M., & Dorigo, M. An ant colony optimization approach to the probabilistic traveling salesman problem. In J. J. Merelo, P. Adamidis, H. G. Beyer, J. L. Fernández-Villacanas, & H. P. Schwefel (Eds.), Proceedings of PPSN-VII, *seventh international conference on parallel problem solving from nature*. Lecture Notes in Comput Sci (2002) 883–892. Berlin: Springer.

[13] Demirel, N. Ç., & Toksari, M. D. Optimization of the quadratic assignment problem using an ant colony algorithm. Applied Mathematics and Computation, 183(1), (2006) 427–435.

[14] Lina, B. M. T., Lub, C. Y., Shyuc, S. J., & Tsaic, C. Y. Development of new features of ant colony optimization for flowshop scheduling. *International Journal of Production Economics*, 112(2),(2008) 742–755.

[15] Ramos, G. N., Hatakeyama, Y., Dong, F., & Hirota, K. Hyperbox clustering with ant colony optimization (HACO) method and its application to medical risk profile recognition. Applied Soft Computing, 9(2),(2009) 632–640.

[16] Parpinelli, R. S., Lopes, H. S., & Freitas, A. A. An ant colony algorithm for classification rule discovery. In H. A. Abbass, R. A. Sarker, & C. S. Newton (Eds.), Data mining: A heuristic approach (2002a) 191–208. London: Idea Group Publishing.

[17] Parpinelli, R. S., Lopes, H. S., & Freitas, A. A. Data mining with an ant colony optimization algorithm. *IEEE Transactions on Evolutionary Computation*, 6(4), (2002b)

[18] Jiang, W., Xu, Y., & Xu, Y. A novel data method based on ant colony algorithm. Lecture Notes in Computer Science,(2005) 3584, 284–291.

[19] Jin, P., Zhu, Y., Hu, K., & Li, S. Classification rule mining based on ant colony optimization algorithm. In D. S. Huang, K. Li, & G. W. Irwin (Eds.), Intelligent control and automation, lecture notes in control and information sciences (2006). 654–663). Berlin: Springer.

[20] Liu, B., Abbass, H. A., & Mckay, B. Classification rule discovery with ant colony optimization. *IEEE Computational Intelligence Bulletin*, 3(1),(2004) 31–35.

[21] Nejad, N.Z., Bakhtiary, A.H., & Analoui, M. Classification using unstructured rules and ant colony optimization. In: *Proceedings of the international multi conference of engineers and computer scientists* Vol. I, (2008) 506–510. Hong Kong.

[22] Roozmand, O., & Zamanifar, K.Parallel ant miner 2. In L. Rutkowski et al.(Eds.), Artificial intelligence and soft computing – ICAISC (2008) 681–692 Berlin: Springer.

[23] Liu, B., Abbass, H.A., & Mckay, B. Density-based heuristic for rule discovery with Ant-Miner. In: The 6th Australia–Japan joint workshop on intelligent and evolutionary system (2002) 180–184 Canberra, Australia.

[24] Wang, Z., & Feng, B. Classification rule mining with an improved ant colony algorithm. In G. I. Webb & X. Yu (Eds.), AI 2004: Advances in artificial intelligence, lecture notes in computer science (2005) 357–367 Berlin: Springer.

[25] Ooghe, H. and Spaenjers, C A note on Performance measures for business failure prediction models, applied economic letters, 17(1) (2010) 67-70.

[26] Altman, E. and Narayanan, P. An International survey of business failure classification models, financial markets, institutions and instruments, 6(2) (1997) 1-57.

[27] R.E.Abdel-Aal,GMDH based feature ranking and selection for improved classification medical data,J.Biomed.Inform.38(6)(2005)456–468.

[28] M.F.Akay,Support vector machines combined with feature selection for breast cancer diagnosis,Int.J.ExpertSyst.Appl.36(2)(2009)3240–3247.

[29] Chin-YuanFan, Pei-ChannChang, Jyun-JieLin, J.C.Hsieh, A hybrid model combining case-based reasoning and fuzzy decision tree for medical data classification, Int.J.Appl.SoftComput.11(1)(2011)632–644.

[30] Huan Liu, LeiYu, Toward integrating feature selection algorithms for classification and clustering,*IEEETrans.Knowl.DataEng*.17(4)(2005)491–502.

[31] R.Kohavi, George H.John, Wrappers for feature selection subset selection, Artif. Intell.97(1–2) (1997)273–324.

[32] II-SeokOh, Jin-SeonLee, Byung-RoMoon, Hybrid genetic algorithm for feature selection, *IEEETrans.PatternAnal.Mach.Intell*.26(11)(2004)1424–1437.