Proceedings of the International MultiConference of Engineers and Computer Scientists 2011 Vol I,
IMECS 2011, March 16 - 18, 2011, Hong Kong

Construction and Evaluation of a Robust Classification Model for Multi-objective Problems

Hung-Yi Lin and Yu-Han Lai, *Member, IAENG*

Abstract—**Classification by using of multiple variables is a frequently encountered data mining problem. Conventional classification models can either suffer from insufficient data collection or be burdened with overabundant data. In this paper, we propose a novel model in generating a robust multivariate classifier to solve the overabundance case. The classification problems with multiple objectives can be supported by a subset of effective variables identified by our scheme. Traditional Gini index and principal component analysis (PCA) are integrated to complete our classification model. Some experiments based on practical databases are conducted to verify the robustness of our method.**

Index Terms—**Multi-objective classification, Gini index, PCA, multivariate classifier**

I. Introduction

The ubiquitous information tools and technology allows us to collect and store a variety of data for various application domains. The ripe development of multi-dimensional database management, artificial intelligence, expert system, and data mining techniques makes easy implementation of various observation, examination, study, or analysis on the stored data. Based on these advancements of computer technologies and computational techniques, decision makers have gained more momentum to tackle difficulties and even to make precise predictions.

Classification is a frequently encountered problem where a categorical dependent variable needs to be predicted based on a subset of independent variables. Many classification problems including web page classification [25], web spam detection [3, 6], intrusion detection [7], mobile commerce behavior [19], fraud detection [5], bankruptcy prediction [29], medical diagnosis [9, 12], and crime activity analysis [10], have attracted many attentions and encouraged new research stream.

In order to simplify data processing procedure and in turn promote classification performance, data dimension reduction is essential before data analysis is preceded. Especially when high dimensional features are considered, the use of proper features is crucial to subsequent handles. Feature extraction and feature selection are two familiar strategies in achieving data dimension reduction. The methods of feature extraction transform or arrange some original features into a single new feature which is more capable of classification task [20, 22, 27]. The methods of feature selection generally adopt some specific criteria to evaluate original features and then select a subset of proper features for classification task [17, 23, 28, 30]. In this paper, feature selection and feature extraction are integrated in our two-stage scheme for efficient classification.

Dataset size and data diversity in modern digital applications keep growing and changing [13], the handle of multi-objective problems necessitates accurate data analysis and effective data processing. Many studies are dedicated to analyze various types of data and develop new split criteria for achieving adequate categorization. Multivariate analysis methods with the integration of the effective use of multiple independent variables and the collaboration of suitable classifying scheme have verified their validations [11, 21, 24, 34]. A new PCA-based multivariate classification method which takes data granularity into account is proposed to manipulate multi-objective problems in this paper.

The remainder of this paper is organized as follows. In Section 2, some related studies are discussed. The third section proposes a new model with two stages for the generation of new multivariate classifier. The experimental results are given in Section 4. We conclude this paper in section 5.

II. Related Work and Background

Complexity, efficiency, and accuracy are three principles in assessing classifiers. The number of variables selected and the association handles based on these selected variables are two important factors impact the complexity of classifiers. And typically, the complexity of one classifier should be disproportional to its classification efficiency. So, one classifier with simple handle can process classification with efficiency. Nevertheless, efficiency does not grantee accuracy. Many classification methods devote so many efforts on data processing and analyzing that they sacrifice classification efficiency in return for higher classification accuracy. Identifying a subset of useful variables and taking effectual handles are no doubt the way to promoting efficiency and accuracy.

On another hand, in order to collect sufficient information, every instance usually includes a lot of categorical and/or

This work was supported by the Nation Science Council of ROC under Grant No. 99-2221-E-025-012. Hung-Yi Lin is now an associate professor with the Department of Distribution Management, National Taichung Institute of Technology, Taiwan R.O.C., (e-mail: linhy@ntit.edu.tw).

Yu-Han Lai is now a graduate student with the Department of Distribution Management, National Taichung Institute of Technology, Taiwan R.O.C., (e-mail: guiwui77@gmail.com).

numerical features (or factors). These features generally reflect independent data elements. For clustering problem, all features are considered independently. For classification problem, one or few features are taken as dependent for prediction purpose. On the other hand, in order to induce data trend and then deduce future data, it is required to accumulate an enough amount of instances for analysis. In this scenario, overabundance becomes the bottleneck when filtering and condensing a huge amount of data. Useful features and typical instances are essential to remedy such overabundance problem. This paper focuses on the issue of useful features.

Feature selection aims at exploring the effective variables in datasets. Many statistical analyses and artificial intelligence techniques are proposed to identify the effective features. They include genetic algorithm [8, 33], support vector machine [4, 32], neural network and fuzzy [18, 31], logistic regression [14], and principal component analysis [16]. The virtues of feature selection are threefold. First is to eliminate the redundant information so that the analytical time of mining process can be reduced. Secondly, the selection of a small subset of low correlated features will facilitate data mining process since it prevents similar factors from being repeatedly examined. Thirdly, for classification problems, the relevant features to the target feature are more effective than the irrelevant ones when exploring their relations.

Many previous works are using different measures of impurity/entropy/goodness to select the split attribute in order to produce the best classification quality. It is a relatively hard task since different classification problems evolved from different application backgrounds. Gini Index [15] and Information Gain are two widely used split criteria. Gini index is a measure of statistical dispersion. Gini index is used as a measure of impurity of an independent variable and is commonly used when the dependent variable (target variable) is a categorical variable. The Information Gain function has its origin in the information theory. Decision trees (ID3, C4.5 and C5) are based on Information Gain. In [26], the behavior of Gini index and Information Gain are reported to disagree only in 2%. Without loss of generality, Gini index is adopted in our model and C4.5 is taken as the competitor for comparison analysis in this paper.

III. HEURISTIC MODEL FOR NEW CLASSIFIER

We propose a new classification model which integrates feature selection and feature extraction. A new heuristic algorithm is proposed in the first stage. The algorithm is based on Gini index and considers the correlation between features. A compact set of effective features is determined by a heuristic scheme. The second stage applies multivariate analysis on the selected features for the generation of multivariate classifier and then this classifier undertakes the task of final inductive learning.

A. Feature Selection

At first, the Gini indexes of all independent variables are measured after a collection of training data is inputted. To simplify the explanation of our design, we illustrate a small dataset with 10 instances as shown in Table I. This table investigates 10 individuals with three classes of transportation mode: bus, car and train. Independent variables are gender (x_1), car ownership (x_2=0 for none, x_2=1 for one and x_2=2 for two cares), travel cost (x_3) and income level (x_4). The classes of transportation mode (C) consist of three groups: 4 buses, 3 cars and 3 trains.

TABLE I A SMALL DATASET

Attributes				Classes
gender(x_1)	Car ownership(x_2)	travel cost(x_3)	income level(x_4)	
male	0	cheap	low	bus
male	1	cheap	medium	bus
female	0	cheap	low	bus
male	1	cheap	medium	bus
female	1	expensive	high	car
male	2	expensive	medium	car
female	2	expensive	high	car
female	1	cheap	medium	train
male	0	standard	medium	train
female	1	standard	medium	train

The Gini indexes for x_1, x_2, x_3, x_4 are 0.6, 0.45, 0.16, 0.36, respectively. The minimum gain is obtained for the split based on x_3. That is, among all features, the dataset split by "income level" has the best consistence with "transportation mode".

Generally, unifying several features may improve classification quality. For example, if the dataset is split by ($x_1 \cup x_2$) the Gini index of 0.23 is obtained which is better than those by x_1 and x_2. All six possibilities formed by extracting two features from four are listed in column (1) of Table II. In order to distinguish the correlativity between features, simple data transformations are applied on four attributes and then correlation coefficients (denoted as $|r_{ij}|$) are measured. As for gender, x_1=female is transformed into "1" and male into "2". In addition, cheap (or low), standard (or medium) and expensive (or high) are respectively transformed into 1, 2 and 3. Some phenomenons in the table are observed and depicted as following.

- Among the correlation coefficients derived from all six cases as shown in the column (2) of Table II, (x_1, x_2) owns the least value (marked as *). However, this information is useless since x_3 is not involved.
- Among the correlation coefficients derived from (x_1, x_3), (x_2, x_3) and (x_3, x_4), x_1 has the lowest correlation with x_3 even though x_1 has the highest Gini index of all.
- Three lowest Gini index products (as depicted in column (3)) happen to cases ($x_1 \cup x_3$), ($x_2 \cup x_3$) and ($x_3 \cup x_4$).
- Integrating correlation coefficient and Gini index product, (x_1, x_3) has the lowest result of all (as depicted in column (4)).
- The unifying of x_1 and x_3 results in the lowest Gini index as shown in column (5).

The feature that is first identified as effective will become the causing factor for the low unifying Gini indexes in the subsequent. Data correlativity is another important factor affecting classification quality when several features are considered. As table II shown, x_1 has the poor classification quality while it can efficiently promote the classification quality achieved by x_3. It is worth to notice that low $|r_{ij}|$ value can indicate the low reduplication of classification effect between x_i and x_j.

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

In the following, we propose an algorithm which proposes a heuristic scheme in arranging a compact subset of features for high classification quality. The following notations are used in the algorithm.

TABLE II MEASURE FOR INTEGRATED GINI

(1) $(x_i \cup x_j)$	(2) $\lvert r_{ij}\rvert$	(3) $\text{Gini}(x_i)\times\text{Gini}(x_j)$	(4) $(2)\times(3)$	(5) $\text{Gini}(x_i \cup x_j)$
$(x_1 \cup x_2)$	0.1429*	0.2700	0.0386	0.23
$(x_1 \cup x_3)$	0.2294	0.0960	$\boxed{0.0220}$	0.10
$(x_1 \cup x_4)$	0.3162	0.2160	0.0683	0.25
$(x_2 \cup x_3)$	0.6227	0.0720	0.0448	0.13
$(x_2 \cup x_4)$	0.6776	0.1620	0.1098	0.25
$(x_3 \cup x_4)$	0.7255	0.0576	0.0418	0.13

- A: The set contains all original independent attributes for a specific dataset.
- B_1, B_2: The sets used to collect the effective attributes.
- α, β, γ: The first, second and third effective attributes extracted from set A.
- $r(\beta, \gamma)$: The correlation coefficient between attributes β and γ.
- PCA(S): After applying principle component analysis over a set S, the eigenvalue of the first component is outputted.

Algorithm
Input: The attribute set A.
Output: A subset of attributes extracted from A.

1. $B_1 \leftarrow \phi$; $B_2 \leftarrow \phi$.
2. Pick the attribute "α" with the least Gini index among all attributes. Remove it from A and add to B_1.
3. Choose the next attribute "β" from A which minimizes $\text{Gini}(\beta)\times|r(\alpha,\beta)|$. Remove β from A and add to B_1.
4. Similarly, attribute γ is picked from A for the reason that it minimizes $\text{Gini}(\gamma)\times|r(\alpha,\gamma)\cdot r(\beta,\gamma)|$. Delete γ from A and $B_2 \leftarrow B_1 \cup \{\gamma\}$.
5. While PCA(B_2) > PCA(B_1) do
6. Begin
7. Choose attribute x from A which minimizes $\text{Gini}(x)\times|\prod_{y\in B_2} r(x,y)|\cdot$
   $B_1 \leftarrow B_2$; $B_2 \leftarrow B_2 \cup \{x\}$.
8. End
9. Return B_2.

We notice that $\text{Gini}(x)\times|\prod_{y\in B_2} r(x,y)|$ in step 7 is used to measure the enhanced effect of feature x on B_2. For instance, suppose set A contains x_p, x_q and x_r and set B contains y_i and y_j, then three measurements $\text{Gini}(x_p)\times|r(x_p,y_i)\cdot r(x_p,y_j)|$, $\text{Gini}(x_q)\times|r(x_q,y_i)\cdot r(x_q,y_j)|$ and $\text{Gini}(x_r)\times|r(x_r,y_i)\cdot r(x_r,y_j)|$ are compared for the next authenticated attribute.

B. Feature Extraction

To promote the classification quality as far as possible, principal component analysis (PCA) is applied on the authenticated attributes in B_2. All dta corresponding to the authenticated attributes are preprocessed into quantitative values for PCA. The first component with the maximum variance (eigenvalue) is taken to extract the major essence of attributes. We neglect other components for their minor data variance explanation rates. Continue the previous example with table I, $y = -0.1858\ x_1' + 0.9826\ x_3'$ is generated after PCA is applied. Table III lists the preprocessed data and transformed data corresponding to every original data.

TABLE III MEASURE FOR INTEGRATED GINI

Attributes				Component	Classes
x_1	x_3	x_1'	x_3'	$y = -0.1858\ x_1' + 0.9826\ x_3'$	
male	cheap	2	1	0.6110	bus
male	cheap	2	1	0.6110	bus
female	cheap	1	1	0.7968	bus
male	cheap	2	1	0.6110	bus
female	expensive	1	3	2.7620	car
male	expensive	2	3	2.5762	car
female	expensive	1	3	2.7620	car
female	cheap	1	1	0.7968	train
male	standard	2	2	1.5936	train
female	standard	1	2	1.7794	train

Final inductive learning completes this example, the classification rule is concluded as that entities with component values fall in [0, 1], [1, 2], or [2,3] are respectively categorized into class "bus", "train", or "car". The accuracy of this example is 0.9 since the eighth entity is the unique mismatched record. More datasets are employed in the next section for the verification of our model.

IV. EXPERIMENTAL RESULTS

Three multi-objective datasets are adopted in our experiments. The first dataset is German credit approval (DB1) which is extracted from the UCI [2]. The second and third dataset are residential quality (DB2) and wealth (DB3) which are taken from [1]. Table IV lists the related information including the total number of instances (N), the number of input attributes (A) and the number of outcomes for the target attribute (R). We adopted 10-fold cross-validation to train and test these datasets for 10 times. So, the experimental results with numeric measures were averaged to produce a single estimation. In following subsections, statistical analyses, classification effectiveness and classification performance are measured for detail verification.

TABLE IV THREE DATASETS

	DB1	DB2	DB3
N	1000	329	252
A	20	9	15
R	4	4	5

A. Statistical analyses

Our first experiment measures the correlation coefficients between attributes in order to observe the reduplication degree from different attributes. In Table V, the higher correlations are found in the entropy-based model C4.5 while this suffering has less impact on our model.

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

TABLE V

	C4.5	Our
DB1	0.05	0.03
DB2	0.48	0.20
DB3	0.84	0.26

The results of second experiment as shown in Table VI list the varied multivariate attributes $y_1^{(k)}$ for three datasets, $k \in \{2,3,4,5\}$. The eigenvalue λ for each classifier reveals the total explained variance caused by the component attributes. The eigenvalues with shadow mean the current classifiers are good enough to accept. Although the more component attributes lead to the higher λ and thus the higher explanation ratio, the slowing-down growth indicates that the last appended attribute has the constricted enhancement of classification effect to the previous attributes. Our heuristic algorithm ceases to execute when the growth of eigenvalues is below 30% of the previous measure. For instance, DB3 adopts $y_1^{(4)}$ as its classifier rather than $y_1^{(5)}$ since the eigenvalue of $y_1^{(5)}$ enhances only 26% growth of the total explained variance when compared to that of $y_1^{(4)}$, i.e. (3.956-3.134)/3.134 $\approx$ 0.26.

TABLE VI VARIOUS MULTIVARIATE ATTRIBUTE

	Multivariate attribute	λ
DB1	$y_1^{(2)}=-0.0031\,a_{18}+a_{12}$	1.074
	$y_1^{(3)}=0.0112a_{18}-0.08446a_{12}+0.9964a_1$	1.554
	$y_1^{(4)}=0.0111a_{18}-0.0854a_{12}+0.9963a_1-0.0053\,a_8$	1.554
DB2	$y_1^{(2)}=0.5488a_7+0.8536a_9$	1.429
	$y_1^{(3)}=0.6796a_7+0.2463a_9+0.6910a_5$	1.854
	$y_1^{(4)}=0.1652a_7+0.6741a_9+0.6537a_5+0.3018a_1$	1.913
DB3	$y_1^{(2)}=0.7568a_7+0.6537a_3$	1.392
	$y_1^{(3)}=0.7067a_7+0.1406a_3+0.6934a_{11}$	2.085
	$y_1^{(4)}=-0.5799a_7-0.0412a_3-0.5500a_{11}-0.5996a_4$	3.134
	$y_1^{(5)}=-0.5128a_7-0.0193a_3-0.4736a_{11}-0.5287a_4-0.4826a_{13}$	3.956

B. Classification effectiveness

To explore whether our multivariate classifier would support effective classification, analysis of variance (ANOVA) is applied to test if there was any significant difference between different classes for three datasets. Table VII shows F-test and p values verify the soundness of our model. The accuracies of our model for three datasets shown in table VIII outperform those of C4.5. Although only little superiority is made by our model, we remind the readers that our model only executes one round of inductive learning while C4.5 needs to execute many rounds. The decision levels used in C4.5 are listed in the parentheses behind the accuracies.

TABLE VII ANOVA

	Classifier	F-test	P value
DB1	$y_1^{(3)}$	231.94	0.000
DB2	$y_1^{(3)}$	67.92	0.000
DB3	$y_1^{(4)}$	107.39	0.000

TABLE VIII ACCURACY

	C4.5	Our
DB1	0.450(10)	0.675
DB2	0.742(3)	0.773
DB3	0.550(4)	0.560

C. Classification performance

The percentages of feature reduction are shown in table IX. The numbers of features involved are listed as well in the parentheses behind the percentages. Notice that DB1 has the highest feature number of all and our model achieves the most economic classification performance by appealing to the least amount of effective features. Combine the results from tables VIII and IX, we assert that our model can complete classification task without the loss of accuracies as compared to C4.5.

TABLE IX FEATURE REDUCTION

	DB1	DB2	DB3
Our	85%(3)	67%(3)	73%(4)
C4.5	25%(15)	44%(5)	20%(12)

Finally, the amounts of data involved in the training process for three datasets are counted for efficiency analysis. For each try in the 10-fold cross-validation experiment, all features (input and target) and all instances have to be taken for feature evaluation. Namely, at least 21×1000×0.9, 10×329×0.9, and 16×252×0.9 data entries must be involved for three datasets. C4.5 algorithm needs to execute several rounds of inductive learning and feature evaluation in order to set the proper features in the sequent decision nodes. This is why table X shows our model imposes on less data amount than C4.5. The last row of table X verifies that time efficiency achieved by our model.

TABLE X THE NUMBER OF DATA INVOLVED

	DB1	DB2	DB3
Our	18900	2961	3629
C4.5	96792	4815	7974
ratio	5.12	1.63	2.20

V. CONCLUSION

A new heuristic algorithm and PCA-based analysis are successfully integrated into a novel classification model. The main contributions of this paper are twofold. First is the compact subset of attributes with enough classification capability. Such procedures make sure that the useful information are taken into account and prevent the overabundant data from processing. Second is the multivariate attribute achieved by PCA. Our final classifier is trained by only one round of inductive learning. Our experimental results support the significant improvement of classification performance. Although there is much improvement room for classification accuracy, our classification accuracy rates are still competitive to C4.5. Keeping growing classification accuracy without the loss of performance is our future research direction.

REFERENCES

[1] http://lib.stat.cmu.edu/datasets
[2] UCI Machine Learning Repository. Available: http://archive.ics.uci.edu/ml/

[3] L. Becchetti, C. Castillo, D. Donato, R. Baeza-Yates and S. Leonardi, "Link analysis for Web spam detection," *ACM TWEB Transactions on the Web*, vol. 2, no. 1, 2008.

[4] T. Bellotti and J. Crook, "Support vector machines for credit scoring and discovery of significant features," *Expert Systems with Applications*, 36(2), 3302-3308, 2009.

[5] R. J. Bolton and D. J. Hand, "Statistical fraud detection: a review," *Statistical Science*, vol. 17, no. 3, pp. 235–255, 2002.

[6] C. Castillo, D. Donato, A. Gionis, V. Murdock and F. Silvestri, "Know your neighbors: web spam detection using the web topology," in *Conf. Annual ACM Conference on Research and Development in Information Retrieval*, Netherlands Amsterdam, 2007, pp. 423-430.

[7] V. Chandola, A. Banerjee and V. Kumar, "Anomaly detection: A survey," *ACM CSUR Computing Surveys*, vol. 41, no. 3, 2009.

[8] L. Chen, Q. Ye and Y. Li, "Research on GA-based bank customer's credit evaluation," *Computer Engineering*, 32(3), 70-72, 2007

[9] D. Delen, G. Walker and A. Kadam, "Predicting breast cancer survivability: a comparison of three data mining methods," *Artificial Intelligence in Medicine*, vol. 34, no. 2, pp. 113-127, 2005.

[10] V. Estivill-Castro and I. Lee, "Data mining techniques for autonomous exploration of large volumes of geo-referenced crime data," in *Proc. 6th International Conf. on Geocomputation*, 2001.

[11] J. A. Etzel, V. Gazzola and C. Keysers, "Testing simulation theory with cross-modal multivariate classification of fMRI data," *PLoS ONE*, vol. 3, no. 11, 2008.

[12] Y. Feng, Z. Wu, X. Zhou, Z. Zhou and W. Fan, "Knowledge discovery in traditional Chinese medicine: State of the art and perspectives," *Artificial Intelligence in Medicine*, vol. 38, no. 3, pp. 219-236, 2006.

[13] J. F. Gantz. (2007, March). *A Forecast of Worldwide Information Growth Through 2010*. Available: http://www.emc.com/collateral/analyst-reports/expanding-digital-idc-white-paper.pdf.

[14] A. Genkin, D.D. Lewis and D. Madigan, "Large-scale Bayesian logistic regression for text categorization," *Technometrics*, 49(3), 291-304, 2007.

[15] C. Gini, "Variabilità e mutabilità," Reprinted in Memorie di metodologica statistica (Ed. Pizetti E, Salvemini, T), Rome: Libreria Eredi Virgilio Veschi, 1912.

[16] X. Hao, W. Deng-sheng and X. Yang-qun, "Study on enterprise credit evaluation based on PCA/FCM," *Technology Economics*, 3, 2007.

[17] M. M. Kabira, M. M. Islamb and K. Murase, "A new wrapper feature selection approach using neural network,", *Neurocomputing*, vol. 73, pp. 3273-3283, 2010.

[18] A. Khashman, "Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes," *Expert Systems with Applications*, 37(9), 6233-6239, 2010.

[19] H.-W. Kim, H. C. Chan and S. Gupta, "Value-based adoption of mobile internet: An empirical investigation," *Decision Support Systems*, vol. 43, no. 1, pp. 111-126, 2007.

[20] N. Kwak and J.-W. Lee, "Feature extraction based on subspace methods for regression problems," *Neurocomputing*, vol. 73, pp. 1740-1751, 2010.

[21] S. K. Lee, "On generalized multivariate decision tree by using GEE," *Computational Statistics & Data Analysis*, Vol. 49, Issue 4, no. 15, pp. 1105-1119, Jun. 2005.

[22] G.-C. Luh and C.-Y. Lin, "PCA based immune networks for human face recognition," *Applied Soft Computing*, 2010.

[23] S. Maldonado, R. Weber and J. Basak, "Simultaneous feature selection and classification using kernel-penalized support vector machines," *Information Science*, vol. 18, pp. 115-128, 2011.

[24] C. Orsenigo and C. Vercellis, "Multivariate classification trees based on minimum features discrete support vector machines," *IMA Journal of Management Mathematics*, vol. 14, no. 3, pp. 221-234, 2003.

[25] X. Qi and B. D. Davison, "Web page classification: Features and algorithms." *ACM CSUR Computing Surveys*, vol. 41, no. 2, 2009.

[26] L. E. Raileanu and K. Stoffel, "Theoretical Comparison between the Gini Index and Information Gain Criteria," *Annals of Mathematics and Artificial Intelligence*, vol. 41, no. 1, pp. 77-93, 2004.

[27] M. Šušteršič, D. Mramor and J. Zupan, "Consumer credit scoring models with limited data," *Expert Systems with Applications*, vol. 36, no. 3, pp. 4736-4744, 2009.

[28] C. F. Tsai and Y. C. Hsiao, "Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches," *Decision Support Systems*, vol. 50, pp. 258-269, 2010.

[29] C. F. Tsai, "Feature selection in bankruptcy prediction," *Knowledge-Based Systems*, vol. 22, no. 2, pp. 120-127, 2009.

[30] C. M. Wang and Y. F. Huang, "Evolutionary-based feature selection approaches with new criteria for data mining: A case study of credit approval data," *Expert Systems with Applications*, vol. 36, pp. 5900-5908, 2009.

[31] C. Wu and H. Xia, "Study of personal credit evaluation under C2C environment based on support vector machines ensemble," *International Conference on Management Science and Engineering* (pp. 25-31). CA: Long Beach, 2008.

[32] L. Yu, W. Yue, S. Wang and K. K. Lai, "Support vector machine based multiagent ensemble learning for credit risk evaluation," *Expert Systems with Applications*, 37(2), 1351-1360, 2010.

[33] H. Zhao, "A multi-objective genetic programming approach to developing Pareto optimal decision trees. *Decision Support Systems*, 43(3), 809-826, 2007.

[34] M. Zucknick, S. Richardson and E. A. Stronach, "Comparing the characteristics of gene expression profiles derived by univariate and multivariate classification methods," *Statistical Applications in Genetics and Molecular Biology*, vol. 7, no. 1, 2008.

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)