Identification of Reliable Information for Classification Problems

Kuang Yu Huang
Department of Information Management
Ling Tung University
#1 Ling Tung Road, Taichung City 408
Taiwan
kyhuang@mail.ltu.edu.tw

Hung-Yi Lin
Department of Distribution Management
E-mail: linhy@ntit.edu.tw
National Taichung University of Technology
129, Sanmin Rd., Sec. 3, Taichung, Taiwan (R.O.C.)

Abstract: - A novel information identification model is proposed to support accurate classification tasks with mixtures of categorical and real-valued attributes. This model combines the advantages of rough set theory and cluster validity method to promote the classification quality to the higher levels. Real-valued attribute values are pre-processed by fuzzy c-means clustering method and then analyzed by variable precision rough set theory. Our cluster validity index finalizes the information system with the feasible cluster number for each attribute. In the case that a considerable amount of ambiguous instances is included, the experimental results show that our model can explicitly improve traditional classifiers in the aspects of classification accuracy and discrimination power. This paper provides a better solution for the generation of reliable decision rules for classification problems with attribute mixtures.

Key-Words: - Reliable information, classification problems, fuzzy c-means, variable precision rough set, cluster validity index, discrimination power

1 Introduction

Information and knowledge discovered from the tremendous amount of data have constantly encouraged the data mining techniques to step forward a higher level. These techniques including association rules, classification models, cluster analyses, sequential patterns, and time sequence methods [27]. Many modern decision makers concentrate more and more their attentions on how to interpret the historic transaction data. For the sake of making right strategies, they are in the urgent need of learning past experiences and inducing the useful results. Among a variety of different data mining techniques, classification is the typical inductive learning model toward supervising and exploring the relationships between a number of attributes and the target class. A classification algorithm capable of implementing a learning model can be used to generate a specific classifier. The well-known classifiers include decision trees [30], rule-based algorithm [9], neural networks (NN) [40], support vector machines (SVM) [33], conformal predictors [35], Bayesian classifiers [41], logistic regression [25], linear discriminant analysis (LDA) [4], nearest neighbor methods [10], random forests [31], and so forth. In addition, some classification methods [8, 16, 18, 26, 37] are proposed for the problems of text classification.

The data types of attributes fall into two broad categories: nominal and numeric. Nominal (also termed as categorical) attributes generally comprise text-based data and they can either be unordered (for example, male or female) or ordered (for example, high or low). Numeric attributes are usually preserved with ordinal data such as integer or real-valued data. Many classifiers can have more favor to numeric attributes than nominal attributes. For example, to achieve high quality of classification results, classifiers such as SVM, NN, and LDA prefer to handle real-valued attributes. However, in case that datasets equipped with a plentiful amount of nominal attributes or the crucial information is identified in some nominal attributes, these classifiers are unable to draw the suffice information in their supervised learning processes. Namely, many classifiers are likely to suffer from the situation of attribute mixtures and in turn deteriorate their classification performance. Although decision trees [14], nearest neighbor methods [39], random forests [34], and logistic regression [24] get rid of the problem of attribute mixtures, it still leaves space for further improvement. This triggers the first motivation of our method.

Ambiguous information involved in the supervised learning process readily cause overfitting problem. Many traditional classification models ignore to distinguish the typical instances from ambiguous ones. Classifiers constituted from such models could have the poor discrimination powers suffering from the overfitting problem. Rough set (abbreviated as RS) [29] is the well-known theory which can deal with the vagueness and uncertainty of categorical data. RS analyzes an information system based on the advantageous scenario of data driven, nonparametric, and less restrictive in a priori assumption. RS devotes the mathematical analyses on the data structure of the given information and its analysis of equivalence classes is helpful to distinguish indiscernibility and vagueness from unambiguous information. Such advantage in RS triggers the second motivation of our method. In other words, RS is capable of producing the better knowledge representation structure without severely suffering the indiscernible and vague information. So, the using of RS promotes a higher probability to generate the meaningful and interpretable decision rules in an intuitive and comprehensible manner [11].

Although the usefulness of RS is already verified in many scientific fields such as machine learning, data mining, and pattern recognition [21], it lacks the capability in tolerating some certain extent of uncertainty or misclassification error [5, 6]. To remedy such drawback, Ziarko [42] allows some extent misclassification in RS and proposes variable precision rough set (VPRS). In VPRS, the uncertain nature of information within the interested system is handled using the concept of β-lower and β-upper approximate sets. The β-lower approximate set contains all the objects within the system which can be unambiguously ascribed to a particular target set with a certain pre-defined misclassification error, and provides the information required to extract a set of decision rules with which to classify new data arrivals within the system. On the other hand, the β-upper approximate set contains all the objects within the system which may possibly belong to a particular target set. In this paper, the VPRS's tolerant capability of uncertainty or misclassification error is adequately employed to concrete the information derived from the typical objects.

The remainder of this paper is organized as follows. Section 2 respectively reviews the fundamental designs of fuzzy c-means, VPRS, and clustering validity index function. Section 3 employs a hypothetical example and presents how to integrate these concepts to complete a novel information identification model supporting classification problems. The experimental results of classification performance regarding accuracy and discrimination power are given in Section 4. Finally, Section 5 presents some brief concluding remarks and indicates the intended direction of future research.

2 Related work

2.1 Clustering method

As is well known, cluster analyses can achieve the explorative task of assigning a set of data into clusters so that the data in the same cluster are more similar to each other than to those in other clusters. In hard clustering, data is divided into distinct clusters, where each data element belongs to exactly one cluster. In fuzzy clustering (also termed soft clustering), data elements can belong to more than one cluster. Fuzzy clustering provides the higher flexibility and thus the better analytical quality than hard clustering. So, we adopt the most widely used fuzzy clustering algorithm, i.e., fuzzy c-means (FCM) algorithm [12], in this paper. The main advantages of FCM are its simplicity and speed which allows it to run on large datasets. The detail procedures about how to implement FCM are omitted here.

2.2 VPRS model

The VPRS is a generalized model of RS that inherits all basic mathematical properties of the original RS model. RS model assumes that the universe under consideration is known and all conclusions derived from the model are only applicable to this universe. However, in practice, there is an evidence show that only a smaller set of examples suffices for generalizing the conclusions of a larger population. The VPRS model allows for a controlled degree of misclassification. If the majority of available data can be correctly classified, any partially incorrect classification rule provides valuable trend information about future test cases.

Every object (i.e., example or instance) x in a universal set U is characterized and classified by a set of conditional attributes A and a single decision attributes d. The VPRS model deals with partial classifications by introducing a precision parameter β. This parameter is the threshold of the portion of objects in a particular conditional class being classified into the same decision class. In RS model, β is equal to one. The past studies involved in VPRS do not investigate the choosing of β. The systematic method presented by [19] can determine the value of β using FCM clustering method and related fuzzy set theories. In case that an information system is processed by the VPRS model with $0.5 < \beta \le 1$, the goal of this processing is to identify the β-lower and β-upper approximate sets associated with each class of the decision attribute. In general, for a subset of objects X extracted from U (i.e., $X \subseteq U$), the β-lower approximation set under a subset of conditional attributes P (i.e., $P \subseteq A$) is denoted as $\underline{R}_P^{\beta}(X)$ and can be expressed as follows.

$$\underline{R}_P^{\beta}(X) = \{x : P(X \mid [x]_P) \ge \beta\} = \bigcup \{[x]_P : P(X \mid [x]_P) \ge \beta\} \tag{1}$$

Similarly, the β-upper approximation set is denoted as $\overline{R}_P^{\beta}(X)$ and can be given as follows.

$$\overline{R}_P^{\beta}(X) = \{x : P(X \mid [x]_P) > 1 - \beta\} = \bigcup \{[x]_P : P(X \mid [x]_P) > 1 - \beta\} \tag{2}$$

For the k-th class of the decision attribute, the accuracy denoted as α_k^{β} is measured as $\frac{|\underline{R}_P^{\beta}(X)|}{|\overline{R}_P^{\beta}(X)|}$.

In the case of $\beta = 1$, the cardinalities of $\underline{R}_P^{\beta}(X)$ and $\overline{R}_P^{\beta}(X)$ are equal to those of the lower and upper approximation sets in RS theory. That is, the effect of VPRS model is the same as that of RS model.

2.3 Cluster validity index and MV-index

Before applying VPRS model to an information system, it necessitates the number of clusters for every attribute in the dataset. Unfortunately, such information is not known a priori. Finding the optimal number of clusters for discretizing a set of real-valued attributes is a NP-hard problem [23, 32]. To prevent from falling into the NP-hard situation, cluster validity index [7] is integrated with VPRS model for the assessment of cluster quality in this paper. Namely, we choose the adequate cluster number by means of comparing distinct index values. Many cluster validity indexes have been proposed to assess the nature of fuzzy clustering methods [15, 28, 36]. The common insufficient part of them is that they do not take the complicated interrelationships among various attributes into account. In [20, 22], a heuristic VPRS-based index is proposed, which can integrate a new cluster validity index function with the VPRS model. The major contributions of this approach are threefold. First, it is very feasible to classify labeled/unlabeled datasets. Second, it provides a more reliable basis for the extraction of decision-making rules from labeled/unlabeled datasets. Third, it generates the feasible cluster number for unlabeled attributes.

In this paper, we propose a new index function modified from VPRS index and abbreviated it as MV-index. Because VPRS-based index is only feasible to categorical attributes, we utilize the VPRS-based index and our MV-index for the handle of mixture attributes. MV-index becomes a helpful complement to the VPRS-based index. The MV-index ensures a small number of compact clusters are applied to the information system and maximizes the separation distance between these resulting clusters. It can also significantly improve the accuracy of approximation when an appropriate clustering scheme is applied in advance. The design of MV-index is given as follows.

For the completeness of explanation, the notations used in the design are given as follows.

- *N*: The total number of objects in the dataset.
- $|d|$: The cardinalities of *d*. Precisely, the class number of the decision attribute *d*.
- κ: The designated cluster number used to partition real-valued attributes.
- z_i: The centroid of the lower approximate set associated with the *i*-th class for the decision attribute *d*.
- α_i^{β}: The accuracy of the *i*-th class of the decision attribute in VPRS model, $i \in \{1, 2, \cdots, |d|\}$.

For a given dataset X, the measurement of MV-index adopting a cluster number κ is by means of the computation of attribute values in all real-valued attributes and the decision attribute. The index is given as

$$MV(\kappa) = \frac{D}{\kappa \times F} \tag{3}$$

, where F measures the *compactness* inside the resulting clusters while *D* measures the *separation* outside the resulting clusters. Since the more compact individual clusters (i.e., lower measurement of *F*) and the more separable cluster centroids (i.e., higher measurement of

D) are expected, the goal in this process can be reduced to find the least value of κ which maximizing the MV-index. *F* is formulated as $\sum_{i=1}^{|d|}\sum_{j=1}^{N}\frac{\delta\times\|x_j - z_i\|}{\alpha_i^\beta}$, where δ is the dummy variable with the alternative values of 0 or 1. After the approximate sets are identified, in case that x_j is classified into the *i*-th class with centroid of z_i , δ is assigned to be 1. Otherwise δ is assigned to be 0. The calculation of z_i has to retrieve the raw data. The higher the accuracy α_i^β , the more the term *F* contributes to the MV-index. Term *D* is formulated as $\max_{\substack{i,j=1\\ i\neq j}}^{|d|}\|z_i - z_j\|$ which indicates the maximum separation distance among the centroids. Since the distances measured from categorical attributes are meaningless, the measurements of z_i and *D* in the MV-index is valid only when real-valued attributes are considered. Again, we remind that the proposed MV-index is responsible for the decision of optimal cluster number (i.e., κ) for all real-valued attributes while the handles of categorical attributes appeal to VPRS-based index.

3 Novel information identification model

In order to explicitly illustrate the computation of MV-index and in turn the determination of optimal cluster number for real-valued attributes, a simple hypothetical dataset with six objects, i.e., X={x1, x2, x3, x4, x5, x6}, as shown in Table 1 is taken to demonstrate our design step by step. This hypothetical dataset comprises two real-valued attributes (i.e., a1 and a3) and one categorical attribute (i.e., a2). The decision attribute d has the alternative value of class 1 or class 2 (respectively denoted as c1 and c2). The following three steps complete the calculation of MV-index for every round.

Step I. Cluster analysis using FCM.
Initially, the cluster number κ assigned to handle real-valued attributes using FCM is 2, i.e., $\kappa = 2$. Table 2 shows the membership values of a1 and a3 for c1 and c2. Attributes a2 and d retain their original values.

Step II: Generation of decision table and approximate sets.
By the using of index function $I_{\max}$ proposed in [22], the membership values depicted in Table 2 are taken to determine the proper cluster to which the attribute value belongs. As shown in Table 3, the results from $I_{\max}$ is consistent with an intuition that the large membership values dominate the decision in this example. The output of this step is called *decision table* where only discrete values are preserved.
For the simplicity of explanation, the resulting first and second classes of the decision attribute for the hypothetical dataset are listed in Table 4. The original data followed with dominative memberships in the parentheses are depicted in the table. The parameter β of the first class (and the second class) is selected from the minimum among the membership values (please refer to the numbers with shadow). As formulae (1) and (2) mentioned in Section 2.2, the β-lower and β-upper approximate sets associated with each class of the decision attribute can be expressed as follows.

- For the first class (i.e., $d = c_1$), $\underline{R}_A^{0.861}(X)$ = $\{x_3, x_4\}$ and $\overline{R}_A^{0.861}(X)$ = $\{x_3, x_4, x_5\}$.
- For the second class (i.e., $d = c_2$), $\underline{R}_A^{0.945}(X)$ = $\{x_1, x_2\}$ and $\overline{R}_A^{0.945}(X)$ = $\{x_1, x_2, x_6\}$.

The centroids of class one and two are calculated according to simple statistical mean and also given in the last rows of Table 4(a) and 4(b). In addition, the accuracies of class one and two are simply concluded as $\alpha_1^{0.861} = \frac{|\{x_3, x_4\}|}{|\{x_3, x_4, x_5\}|} = \frac{2}{3} = 0.667$ and $\alpha_2^{0.945} = \frac{|\{x_1, x_2\}|}{|\{x_1, x_2, x_6\}|} = \frac{2}{3} = 0.667$. These measurements also reveal that class 1 and class 2 both have 66.7% reliable information and 33.3% ambiguous information.

Step III: Calculation of MV-index.
The calculation of MV-index is sensitive to the test cluster number κ. In the case that $\kappa = 2$ is applied to the hypothetical dataset *X*, the compactness of two classes (i.e., term *F*) is measured as follows.

$$F=\sum_{i=1}^{2}\sum_{j=1}^{4}\frac{\delta\times\|x_j-z_i\|}{\alpha_i^{\beta}}=$$
$$\sum_{j=1}^{4}\frac{\delta\times\|x_j-z_1\|}{\alpha_1^{0.861}}+\sum_{j=1}^{4}\frac{\delta\times\|x_j-z_2\|}{\alpha_2^{0.945}}$$
$$=\frac{\left(0\cdot\|x_1-z_1\|+0\cdot\|x_2-z_1\|+1\cdot\|x_3-z_1\|+1\cdot\|x_4-z_1\|\right)}{\alpha_1^{0.861}}$$
$$+\frac{\left(1\cdot\|x_1-z_2\|+1\cdot\|x_2-z_2\|+0\cdot\|x_3-z_2\|+0\cdot\|x_4-z_2\|\right)}{\alpha_2^{0.945}}$$
$$=0.212+0.335$$
$$=0.547$$

On the other hand, the separation between these two classes (i.e., term *D*) is simply measured from the Euclidean distance between (1.5, 0.5) and (1, 0.25). That is, *D*=0.559. Finally, $MV(2)$ is measured as $\frac{0.559}{2\times 0.547}=0.511$.

The same procedures from step I to III are applied to X when κ increases to 3. The measurement of $MV(3)$ is 0.341 and the detailed calculations are omitted here. Now that $MV(2)$ is greater than $MV(3)$, the decision table achieved by $\kappa=2$ provides a more reliable information basis than that achieved by $\kappa=3$. We terminate all procedures for the measurements of other MV-indexes. Inversely, if $MV(2)$ is less than $MV(3)$, these procedures continue to execute and the comparison between $MV(3)$ and $MV(4)$ is used to determine the consequent handle. The following algorithm is employed to determine the optimal cluster number κ and generate the most reliable decision table for a given dataset.

Using the index function given in Section 2.1, each conditional or decision attribute cluster to which each attribute of each record belong is determined.

Algorithm $MV(\kappa)$

Input: the training dataset *X*

Output: the cluster number κ and the decision table containing discrete attribute values

1. $\kappa_1 \leftarrow 2$ /* Initiate κ_1 for the first round of MV-index calculation. */
2. $\kappa_2 \leftarrow \kappa_1+1$ /* Designate κ_2 for the second round of MV-index calculation. */
3. While $MV(\kappa_1)<MV(\kappa_2)$ and $\kappa_2\le\Psi$ do
4. $\kappa_1 \leftarrow \kappa_2$
5. $\kappa_2 \leftarrow \kappa_2+1$
6. End While
7. Return κ_1 and $MV(\kappa_1)$

In line 3 of this algorithm, Ψ is the maximum permission cluster number for testing. User can designate this number upon their practical demands. A reasonable value of Ψ is assigned to be 10.

4 Experimental results

The lower approximate sets validated by the proposed index method contain the reliable instances while exclude the ambiguous instances. The goal of such information identification is to solve the problems severely degraded by the presence of noisy information. The instances categorized in the lower approximate sets can advantage classifiers being trained under the situation with high consistent information. Experiments are conducted in this study to assess classification performances between distinct classifiers trained by the information systems with and without the prior handle of information identification. Three datasets are employed in this paper. The first dataset is *Australian credit approval* extracted from [1] and denoted as DB1.The second dataset is a subset of the 1987 *National Indonesia contraceptive prevalence survey* extracted from [2] and denoted as DB2. The third is a dataset of determinants of wages from the 1985 *current population survey* extracted from [3] and denoted as DB3. Table 5 is the abstract of these datasets.

To avoid biased condition, common 10-fold cross-validation is adopted and there are 621, 1326, and 480 training instances respectively selected from DB1 to DB3. For each round of the 10 sub-experiments, the training instances are saved into many individual files so that they could be reused for building distinct classifiers for comparison studies. All decision models were implemented in *Matlab* programming languages executed on a workstation with an Intel Core 2 dual 2.4 GHz processor. To verify our design, four classification methods including C4.5, SVM, Logistic Regression (LR), and Simple CART are selected from the 10 most influential algorithms [38] and used in our comparison experiments. Table 6 depicts the information about the quantity of reliable instances categorized in the datasets. The reduction ratios listed in the last row of the table are used to show the ambiguity degree in the training datasets. The higher the ratios, the higher amount of ambiguous instances are involved in the training process. For example, in DB2, almost a half of the amount of training instances is categorized as ambiguous. Our proposed method gets rid of these ambiguous instances and supplies the reliable and concise information for the training of robust classifiers. Inversely, if these ratios are near zero, our proposed method will have limited contribution to classification performance. For example, only 1% amount of instances are categorized as ambiguous in DB1. We can also conclude the instances preserved in this dataset are so typical that they all have helpful contribution to the classification task.

To verify the effectiveness of the proposed method, the original datasets (denoted as Ω) and the datasets refined by MV-index (denoted as Ω') are separately employed to train distinct classifiers. In Table 7, classification accuracy is investigated and the ratio of improved accuracy denoted as Δ is also listed in parentheses. As shown in Table 6, because a little amount of ambiguous instances is categorized in DB1, only slight accuracy improvements are achieved for four classifiers. Since instances in DB1 appear to be consistent so that they constitute the reliable information basis for classification task. In view of all the experimental results shown in Table 7, DB1 really has the best accuracy performance over other datasets. Regarding DB2 or DB3, a considerable amount of ambiguous instances is detected and removed from the original dataset. The accuracies obtained from Ω' generally outperform those from Ω and the improvements are more significant. The last column of Table 7 shows that an average near 10% promotion is achieved by the MV-index in DB2 and DB3. According to overall observation, it is worthy to note that the performance of C4.5 algorithm can be highly sensitive to the presence of noisy and our method can benefit C4.5 more than other algorithms.

The discrimination degree of classifiers is measured by the ROC area [17]. The ROC area is directly represented by plotting the fraction of true positives out of the positives (TPR = true positive rate) vs. the fraction of false positives out of the negatives (FPR = false positive rate). It is a comparison of two operating characteristics (TPR & FPR) as the criterion changes and therefore measures the discrimination capability of the classifier. The closer the curve is to the upper left-hand corner of the graph, the greater the area and the higher the discrimination capability [13]. The range of the ROC area is 0 to 1.

Finally, Table 8 lists the ROC areas achieved by Ω and Ω' for the four classifiers and three datasets with the best results being highlighted. Our method does not seem to favor any classifier since the difference is insignificant. Although there still retains some improvement space, it is consistent with our previous assertion that the MV-index explicitly promote the discrimination powers of four classifiers.

5 Conclusion

MV-index proposed in this paper addresses three contributions to classification problems. First, all attributes are equally respected and the MV-index is good to handle a mixture of real-valued and categorical attributes without any selective priority. This eases the overhead costs for the processing of feature selection before implementing classification tasks. Second, the ambiguous instances are detected and removed from the original dataset. The MV-index can supply training tasks with reduced datasets and in turn reduce the time complex of training process. Third, the reliable information basis derived from MV-index can support the construction of robust classifiers and promote the classification performance of classifiers in terms of accuracy and discrimination power.

Future researches should be directed to the following aspects. The first is to apply different cluster analyses to different attributes and use a distinct number of clusters to assess every single attributes. We are motivated to raise the discrimination power of attributes to an even higher level. In addition, the cooperation of information identification with attribute reduction could be another success in improving classification problems. Finally, the innovation of an all new classifier capable of fully boosting the superiorities of the MV-index is helpful to advance the data mining technique and greet the future challenges.

References:

[1] http://archive.ics.uci.edu/ml/datasets/Statlog+(Australian+Credit+Approval)
[2] http://archive.ics.uci.edu/ml/datasets/Contraceptive+Method+Choice
[3] http://lib.stat.cmu.edu/datasets/CPS_85_Wages
[4] S. An, W. Liu, S. Venkatesh, H. Yan, Unified formulation of linear discriminant analysis methods and optimal parameter selection, Pattern Recognition, 44 (2011) 307-319.
[5] M. Beynon, Reducts within the variable precision rough sets model: a further investigation, European Journal of Operational Research, 134 (2001) 592-605.
[6] M.J. Beynon, Stability of Continuous Value Discretisation: an Application within Rough Set Theory, International Journal of Approximate Reasoning, 35 (2004) 29-53.
[7] J.C. Bezdek, Cluster validity with fuzzy sets, J. Cybernet, 3 (1974) 58-74.
[8] C.M. Bishop, Pattern Recognition and Machine Learning, Springer Verlag (2006).
[9] P. Clark, T. Niblett, The CN2 induction algorithm, Machine Learning, Kluwer Academic Publishers, 3(4) (1989) 261-283.
[10] H.-L. Chen, B. Yang, G. Wang, J. Liu, X. Xu, S.-J. Wang, D.-Y. Liu, A novel bankruptcy prediction model based on an adaptive fuzzy k-nearest neighbor method, Knowledge-Based Systems, 24 (2011) 1348-1359.
[11] M. Daubie, P. Levecq, N. Meskens, A comparison of rough sets and recursive partitioning induction approaches: An application to commercial loans, International Transactions in Operational Research, 9 (2002) 681-694.
[12] J.C. Dunn, A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, J. Cybernet, 3 (1973) 32-57.
[13] T. Fawcett, An introduction to ROC analysis, Pattern Recognition Letters, 27 (2006) 861-874.
[14] B. J. Felix, C. Weber, Gas storage valuation applying numerically constructed recombining trees, European Journal of Operational Research, 216 (2012) 178-187.
[15] A. Ghosh, N.S. Mishra, S. Ghosh, Fuzzy clustering algorithms for unsupervised change detection in remote sensing images, Information Sciences, 181 (2011) 699-715.
[16] D. J. Hand, Construction and Assessment of Classification Rules, John Wiley & Sons (1997).
[17] D.J. Hand, J.T Robert, A Simple Generalization of the Area under the ROC Curve for Multiple Class Classification Problems, Machine Learning, 45 (2001) 171-186.
[18] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Springer (2009).
[19] K.Y. Huang, Application of VPRS model with enhanced threshold parameter selection mechanism to automatic stock market forecasting and portfolio selection, Expert Systems with Applications, 36(9) (2009) 11652-11661.
[20] K.Y. Huang, Applications of an Enhanced Cluster Validity Index method based on the Fuzzy C-means and Rough Set Theories to Partition and Classification, Expert Systems With Applications. 37 (2010) 8757-8769.
[21] K.Y. Huang, S. Wan, Application of enhanced cluster validity index function to automatic stock portfolio selection system, Information Technology and Management, 12(3) (2011) 213-228.
[22] K.Y. Huang, A hybrid particle swarm optimization approach for clustering and classification of datasets, Knowledge-based Systems, 24 (2011) 420-426.
[23] J. Komorowski, Z. Pawlak, L. Polkowski, A. Skoworn, Rough sets: A tutorial. In: Pal SK, Showorn A (Eds.), Rough Fuzzy Hybridization: A New Trend in Decision-Making, Springer, Berlin, 3-98, 1999.
[24] J. Li, E. K. Burke, R. Qu, Integrating neural networks and logistic regression to underpin hyper-heuristic search, Knowledge-Based Systems, 24 (2011) 322-330.
[25] H. Li, J. Sun, Empirical research of hybridizing principal component analysis with multivariate discriminant analysis and logistic regression for business failure prediction, Expert Systems with Applications, 38 (2011) 6244-6253.
[26] G. J. McLachlan, Discriminant Analysis and Statistical Pattern Recognition, John Wiley & Sons Inc, 2004.
[27] K. M. Osei-Bryson, Towards supporting expert evaluation of clustering results using a data

mining process model, Information Sciences, 180 (2010) 414-431.
[28] M.K. Pakhira, S. Bandyopadhyay, U. Maulik, Validity index for crisp and fuzzy clusters, Pattern Recognition, 37 (2004) 487-501.
[29] Z. Pawlak, Rough sets, International Journal of Information and Computer Sciences. 11(5) (1982) 341-356.
[30] J.R. Quinlan, Induction of decision trees, Machine Learning, Kluwer Academic Publishers, 1(1) (1986) 85-106.
[31] V.F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, J.P. Rigol-Sanchez, An assessment of the effectiveness of a random forest classifier for land-cover classification, ISPRS Journal of Photogrammetry and Remote Sensing, 67 (2012) 93-104.
[32] A. Roy, S.K. Pal, Fuzzy discretization of feature space for a rough set classifier, Pattern Recognition Letters, 24 (2003) 895–902.
[33] R. Stoean, M. Preuss, C. Stoean, E. El-Darzi, D. Dumitrescu, Support vector machine learning with an evolutionary engine, Journal of the Operational Research Society, 60(8) (2009) 1116-1122.
[34] A. Verikas, A. Gelzinis, M. Bacauskiene, Mining data with random forests: A survey and results of new tests, Pattern Recognition, 44 (2011) 330-349.
[35] V. Vovk, A. Gammerman, G. Shafer, Algorithmic learning in a random world, New York: Springer, 2005.
[36] W. Wang, Y. Zhang, On fuzzy cluster validity indices, Fuzzy Sets and Systems, 158 (2007) 2095-2117.
[37] A. R. Webb, K. D. Copsey, Statistical Pattern Recognition, John Wiley & Sons Inc, 2011.
[38] X. Wu, V. Kumar, J.R. Quinlan, J.Y. Ghosh, M.H. Q., G. McLachlan, et al., Top 10 algorithms in data mining, Knowledge and Information Systems, 14(1) (2008) 1-37.
[39] J. Yang, L. Zhang, J.Y. Yang, D. Zhang, From classifiers to discriminators: A nearest neighbor rule induced discriminant analysis, Pattern Recognition, 44 (2011) 1387-1402.
[40] G.P. Zhang, V.L. Berardi, Time series forecasting with neural network ensembles: an application for exchange rate prediction, Journal of the Operational Research Society, 52(6) (2001) 652-664.
[41] H. Zhu, P.A. Beling, G.A. Overstreet, A Bayesian framework for the combination of classifier outputs, Journal of the Operational Research Society, 53(7) (2002) 719-727.
[42] W. Ziarko, Variable precision rough set model, Journal of Computer and System Sciences, 46 (1993) 39-59.

Table 1. Simple hypothetical dataset with four objects.

Object-id	a_1	a_2	a_3	d
x_1	0.90	1	0.30	c_2
x_2	1.10	1	0.20	c_2
x_3	1.45	2	0.45	c_1
x_4	1.55	2	0.55	c_1
x_5	1.55	1	0.55	c_1
x_6	1.55	1	0.55	c_2

Table 2. Membership values for a_1 and a_3.

Object-id	a_1		a_2	a_3		d
	c_1	c_2		c_1	c_2	
x_1	0.977	0.023	1	0.043	0.957	c_2
x_2	0.945	0.055	1	0.023	0.977	c_2
x_3	0.027	0.973	2	0.861	0.139	c_1
x_4	0.002	0.998	2	0.995	0.005	c_1
x_5	0.002	0.998	1	0.995	0.005	c_1
x_6	0.002	0.998	1	0.995	0.005	c_2

Table 3. Decision table.

Object-id	a_1	a_2	a_3	d
x_1	1	1	2	c_2
x_2	1	1	2	c_2
x_3	2	2	1	c_1
x_4	2	2	1	c_1
x_5	2	1	1	c_1
x_6	2	1	1	c_2

Table 4. First and second classes for the hypothetical dataset.

(a) $d = c_1$, $\beta = 0.861$

	a_1	a_3
x_3	1.45 [0.973]	0.45 [0.861]
x_4	1.55 [0.998]	0.55 [0.995]
z_1	1.5	0.5

(b) $d = c_2$, $\beta = 0.945$

	a_1`	a_3
x_1	0.9 [0.977]	0.3 [0.957]
x_2	1.1 [0.945]	0.2 [0.977]
z_2	1	0.25

Table 5. Abstract of the three datasets.

Datasets	DB1	DB2	DB3
# of instances	690	1473	534
# of real-valued attributes	6	2	4
# of categorical attributes	8	7	6
# of target classes	2	3	2

Table 6. Abstract of the three datasets.

Datasets	DB1	DB2	DB3
# of instances classified in lower approximate sets	614	721	368
# of training instances	621	1326	480
Reduction ratio	1%	46%	23%

Table 7. Classification accuracies and improvements (%).

	C4.5	SVM	LR	Simple CART	$\overline{\Delta}$
	Ω Ω' (Δ)	Ω Ω' (Δ)	Ω Ω' (Δ)	Ω Ω' (Δ)	
DB1	86.4 87.2 (0.9)	84.8 87.1 (2.7)	77.3 77.6 (0.4)	84.0 86.7 (3.2)	1.8
DB2	47.8 53.6 (12.1)	54.2 58.9 (8.7)	55.0 58.5 (6.4)	53.6 57.0 (6.3)	8.4
DB3	43.0 49.0 (14.0)	64.4 68.9 (7.0)	56.1 57.4 (2.3)	43.0 49.0 (14.0)	9.3

Table 8. ROC areas (%) for four classifiers.

	C4.5		SVM		LR		Simple CART	
	Ω	Ω'	Ω	Ω'	Ω	Ω'	Ω	Ω'
DB1	84.1	86.1	84.6	87.0	84.1	84.2	84.3	87.0
DB2	62.5	73.9	66.8	70.9	72.4	75.9	70.1	78.4
DB3	49.3	59.6	60.1	72.2	54.7	60.4	49.3	59.6