MULTI-CRITERIA ABC METHOD USING HYBRID HEURISTIC MODELS

Bo Li¹, Lianghao Li¹ and Ping Chen¹

¹College of Management and Economics, Tianjin University,
Weijin road, Nankai str, Tianjin, 300072, China
E-mail: bingling.0815@163.com

Inventory classification using ABC method is the most widely techniques in enterprises. A hybrid model of rough sets (RS) and Support Vector Machines (SVMs) for ABC classification of stock keeping units (SKU) has been presented in this paper. The whole process involves two stages: Firstly, the attribute reduction of RS has been applied as pre-processor so that we can delete the redundant attributes and conflicting objects from decision-making table but without loss of the efficient information. Then, we realize the classification modelling and forecasting test based on SVMs. By the hybrid heuristic method, we can greatly reduce the dimension of data and highly decrease the complexity in the process of SVMs classification but obtain the good classification performance. Finally, a numerical simulation shows the effectiveness of the proposed hybrid method.

Keywords: Multi-criteria classification, rough sets, support vector machine, decision rules

1. Introduction

ABC classification is a classical scheme to be used to inventory classification. In this scheme, items are ordered in descending order based on some criteria, such as annual dollar usage values, and then divided stock keeping units into three classes: 20% of the first items belong to Class A, which control 80% of the total annual dollar usage of all items; 50%-60% of the last items constitute Class C, which control a relatively small portion of the total annual dollar usage; Items between the two classes are Class B (about 20%-30%). The division of these classes is designed artificially according to the control policy of companies. One company may have the different control policy to the different classes. Generally, Class A has a tight management control and Class C receives a loose one, Class B should have an intermediate control between them. The criterion used in the classical ABC classification is the annual dollar usage value, however, there are many other criteria which can be affect the inventory classification management, such as, the lead time, the rate of obsolescence, the order size requirement etc..

To deal with the multi-criteria ABC classification problems, a joint criteria matrix for two criteria has been proposed by [1], but it cannot be extended to more than two criteria and its applications are limited. The Analytic Hierarchy Process (AHP) is an alternative option for the multi-criteria ABC classification [2]. The approach can incorporate many relevant qualitative and quantitative criteria, but when constructing the pair wise comparison matrix, the preferences of the decision maker have involved in the matrix and the quality of outputs depends directly on this matrix. However, different decision makers may have different views. Therefore, it is necessary to introduce the AI-based methods to reduce the amount of human involvement in the classification process. Guvenir and Erel [3] proposed a method using Genetic Algorithm to multi-criteria classification, and Partovi, et. al. [4] showed that an artificial neural network with BP algorithm could achieve the higher predictive accuracy as a multi-criteria inventory classifier. Wan [5] gave a simple model for multiple criteria inventory classification model which converted all criteria measures of an inventory item into a scalar score. The classification based on the calculated scores using ABC principle is then applied. Further, Ozan and Mustafa [6] integrated fuzzy concepts with real inventory data and designed a decision support system assisting a sensible multi-criteria inventory classifications, they applied it to a small electrical appliances company to validate the design of the proposed multi-criteria inventory classification system. However, their methods have some common drawbacks that may cause difficulties in practice. One is the number of variables which can be input into these models are limited because of the great amount of calculation. However, to satisfy the various demands of markets, companies have an increasing trend of carrying a great number of different items in inventory. The other is that the models cannot directly deal with the qualitative variables, such as, the manager’s experience, knowledge and judgment, and it is difficult to explain the relationships between criteria and classification for extracting simple rules. Recently, Yu [7] compared artificial intelligence (AI)-based classification techniques with traditional multiple discriminate analysis (MDA).
Operation Research and Decision Making Models

The results show that AI-based techniques demonstrate superior accuracy to MDA. Statistical analysis reveals that Support Vector Machines (SVMs) enables more accurate classification than other AI-based techniques. Referring to this result, we adopt SVMs to inventory classification in this paper.

The rough sets theory introduced by Pawlak [8] has proved to be an effective tool for analysis of a vague description of objects, especially studying how to express, learn and induct incomplete data and imprecise knowledge following from information granulation. It has been applied in several fields such as machine learning, data mining, decision and prediction, pattern recognition and fuzzy control [9]. For example, using rough sets, Angel and Alfonso [10] presented a real application where the information system is reduced so as to get a minimum subset of attributes without loss of quality. Chen et al. [11] proposed multiple criteria ABC analysis method that employed a dominance-based rough set approach (DRSA) to generate linguistic rules to represent a decision maker’s preferences based on the classification of a test data set. These linguistic rules have been applied to classify other SKUs. Because of the properties of these linguistic rules represented by rough sets, we investigate such an extension to the inventory classification problem based on the rough sets [12].

In this paper, a hybrid approach based on rough sets and SVMs to inventory classification is proposed. On one hand, rough set theory is applied so that the redundant criteria can be ruled out in multi-criteria information system with the qualitative representation without any information loss. On the other hand, classification modelling and forecasting test based on SVMs will further be realized. The dimension of data and the complexity can be greatly reduced in the process of SVMs classification but obtain the good classification performance. Finally, numerical simulations show the effectiveness of the proposed hybrid method.

2. Fundamentals

2.1. Attributes Reduction Based on Rough Set

Formally, a decision system, DS (or an approximation space) can be seen as a system \( < U, C \cup D, V, f > \), where \( U \) is the universe (a finite set of objects, \( U = \{x_1, x_2, \cdots, x_m\} \)), \( C \) is the finite sets of the conditional attribute and \( D \) is the finite sets of decision attribute. Each attribute \( a \in (C \cup D) \) (attribute \( a \) belonging to the considered set of attributes \( A \) ) defines an information function \( f_a : U \rightarrow V_a \), where \( V_a \) is the set of values of \( a \), called the domain of attribute \( a \). DS is written as a form of decision table where rows represent objectives and columns represent attribute. In decision table, one of the core contents is attribute reduction that reflects the natural information of a decision table. In most cases, attributes in an approximate space are not equally important and especially some ones are redundant. Redundant attributes not only waste many resources (need large storage space) but also have effect on precise and concise decision. Therefore, attribute reduction is very indispensable.

Dependence between attributes determines whether attributes can be reduced and importance of attributes in terms of attributes dependence is often important information of attribute reduction. For a set \( B \subseteq C \), approximate classification quality is defined as:

\[
r(B, D) = \frac{|pos_B(D)|}{|U|}
\]

Where \( |\cdot| \) represents the number of elements in limited sets and \( pos_B(D) \) is defined as positive region based on a set of \( B \). If \( D \) is all decision attributes and \( B \) is some conditional attribute, \( r(B, D) \) is defined as probability that can be accurately classified by \( D \) after \( B \) is used as the partition. Meanwhile \( r(B, D) \) represents the capacity of conditional attribute \( B \) describing decision attribute \( D \). Based on the dependence between attributes attribute reduction may be defined. Consider \( r(C, D) = r(B, D) \) and there is not \( B' \subseteq B \) satisfying \( r(B', D) = r(B, D) \). \( B \) is called as a reduction of \( C \), where conditional attribute \( B \subseteq D \). There are more than one reduction of attribute sets and \( Core(C) \) called as the core of attribute \( C \) is defined as the intersection of all reductions of \( C \). The importance of attribute \( a \) is defined as \( sig(a, B, D) = r(B + \{a\}, D) - r(B, D) \). In short, the more important the attribute \( a \)
Operation Research and Decision Making Models

contributing to decision attribute is, the bigger effect the attribute \( a \) has on the decision partition.

The fundamental idea of attribute reduction is to use some kind of standard to measure the importance of attributes as heuristic information and to find the optimal reduction from information system or decision system. Optimization means the minimal attributes extraction or being able to cover the largest universe space. The selection of all the standards is determined according to concrete problems. This paper uses the forward attribute reduction based on attribute approximate classification quality. The detailed procedure of which can be seen from the literature [7].

2.2. Classification Mechanism of SVMs

The main idea of SVM is the model of learning from examples, which can be described as: there are \( n \) random independent identically distributed examples

\[
\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}, \quad y_i \in \{-1, +1\}, \; i = 1, 2, \ldots, n
\]

where \((x_i, y_i)\) is the \( i^{th} \) input sample, \( y_i \in \{-1, +1\} \) is the expected output of \( x_i \in R^n \). The optimal separating hyper plane is determined by giving the largest margin of separation between different classes. The optimal hyper plane bisects the shortest line between the convex hulls of the two classes. The points on the margin of the two classes are called Support Vectors (SV). A set of SV can uniquely determine a hyper plane. SVM can be tracked back on the optimal separating hyper plane on the condition of linear separability. The hyper plane of SVM is written as \( w \cdot x + b = 0 \).

This paper uses the nonlinear SVM, which satisfies the following optimization problem:

\[
\begin{align*}
\text{min} & \quad f(X) = \frac{1}{2} \|W\|^2_2 + C \sum_{i=1}^{n} \xi_i \\
\text{s.t.} & \quad y_i (W \cdot \Phi(x_i) + b) \geq 1 - \xi_i \\
& \quad \xi_i \geq 0, i = 1, \ldots, n
\end{align*}
\]

Where \( \cdot \) denotes the inner product, \( C \) are used to weight the penalizing variables \( \xi_i \) and \( \Phi(\cdot) \) is a nonlinear function which maps the input space into a higher dimensional space.

Minimizing the first term \( \frac{1}{2} \|W\|^2_2 \) in the goal function plays the role of controlling the capacity of the learning machine and avoiding the over fitting of the machine. While minimizing the second term is to minimizing the empirical risk (i.e. error rate). Here, a Lagrange method is used to solve the above problem. Then Eq. (1) can be written as

\[
\begin{align*}
\text{max} & \quad W(a) = \sum_{i=1}^{n} a_i - \frac{1}{2} \sum_{i, j=1}^{n} a_i a_j y_i y_j K(x_i, x_j) \\
\text{s.t.} & \quad \sum_{i=1}^{n} a_i y_i = 0, a_i \in [0, C], \; i = 1, \ldots, n
\end{align*}
\]

And then the decision function can be written as:

\[
f(x) = \sum_{SV} a_i y_i K(x_i, x) + b
\]

Where \( K(x, y) = \Phi(x) \cdot \Phi(y) \) is a Kernel function. Sigmoid kernel function written as
Kernel function is infinite and finite samples in the space is bound to be separable.

The classification mechanism of Binary-class problem based on SVM is discussed. For multi-class classification, three methods are employed as follows:

1) one-against-rest. For different $k$ classes, $k$ classifiers are constructed and among them the $i^{th}$ classifier is separating the $i^{th}$ class from the rest, i.e. The $i^{th}$ classifier is labelled as 1 and the rest are labeled as −1. $k$ quadric programming is running in the overall process in order to separate each sample. The shortcoming of this method is that multi-class points and none-class points are produced easily.

2) One-against-one. For any two classes, a classifier is constructed to recognize the two. $k(k − 1)/2$ classifiers are required to be constructed in the overall process. Each sample is determined to belong to which class according to the result of each classifier. Its drawback is that multi-class points may be produced.

3) hierarchy classification. The method approves the shortcoming of one-against-one. The mechanism is combining all $k$ classes into two classes and then separating each big class into two classes respectively until the basic $k$ leaf nodes are produced. Thus, many different hierarchies are produced and SVM is used for classification in each hierarchy.

3. The Hybrid Classification Method Based on RS and SVM

Fig.1 illustrates the whole modelling procedure based on the hybrid classification model. The process is composed of two steps that are pre-processing based on RS and classification modelling and forecasting test based on SVMs. Firstly, the attribute reduction of RS has been applied as pre-processor so that we can delete redundant attributes from decision table but efficient information lossless. To deal with data more effectively, this paper employs 2-dimension data-pre-processing algorithm based on attribute reduction, which need implement not only reduction lengthways, i.e. attribute reduction but also conflict check breadth, i.e. noise data bearing the same conditional attributes but different decision attributes because the results using SVM training conflict objectives may be different. In general, the number of “wrong” output value on objectives is more than the number of objectives on “right” output value, wrong classification will be gained on the basis of SVM; otherwise right classification will be achieved, but the result will lose its sense if the number between right and wrong is the same. Therefore, conflict objectives must be deleted and this is traverse reduction process.

Redundant attributes and conflict objectives will be deleted from a decision table through reduction. Then the algorithm will transform the front into the back, i.e. transform attributes reduction on RS into the training and test on SVM. As SVM mainly deals with binary-class classification, multi-class classification can be implemented on the foundation of $k$-class algorithms such as one-against-one, one-against-rest and two-prong tree.

RS is used for the fore processor of SVM. Input numbers into SVM are decreased to some extent by way of attributes reduction and conflict reduction based on RS. Thus the computation in the process of SVM classification is diminished and much training time is saved largely. Moreover, the over-fitting is avoided in the way of the hybrid algorithm and classification performance is not decreased.

4. Empirical Comparisons

To demonstrate the application of rough sets and SVMs for inventory classification, we have performed the simulation experiments using the real world data obtained from the reference [9]. A small
sample of 95 data sets from a large pharmaceutical company located in Northeastern United States was used to realize the classification tasks.

### 4.1. Representing Inventory Items

The classification items with regard to four criteria are represented and the four types of information are unit price ($/unit from 0.1 to 7132), ordering cost ($/order from 0.05 to 300), demand range (units/year from 1-75 units) and lead time (days from 1-70days). Based on the four criteria, the inventory managers have used ad hoc techniques to integrate the above criteria for ABC classifications. The detail materials can refer to the reference [4].

In our experiments, the classification table of inventory items with regard to the above four criteria forms a decision system, where unit price, ordering cost, demand range and lead time are the conditional attributes and the ABC classifications constitute the decision attribute. The system concerns 95 items described by means of this four attributes. Then the decision system $S$ can be expressed:

- $U = \{1, 2, \ldots, 95\}$,
- $C = \{\text{unit-price, ordering-cost, demand-range, lead-time}\}$,
- $D = \{\text{class}\}$

$V_j = [0.1, 7132], V_2 = [0.05, 300], V_3 = [1, 75], V_4 = [1, 70]$

The information function

$$f(x, c_j) \in V_j, x \in U, c_j \in C$$

and $f(x, d) \in D = \{\text{ClassA or ClassB or ClassC}\}$.

We must note that rough sets are concerned with discrete values, it is necessary to transform quantitative attributes (or continuous attributes) into qualitative representations based on some rules, experience or prior knowledge. Therefore, rough sets method is fit for inventory classification problem, because there are many qualitative elements involved in the inventory classification, and in the real application the decision maker always make some judgment according to the vague value, such as, the unit price of some item is high, medium or low etc.. In our experiments, as two criteria of unit price and ordering cost are given the quantitative values, rough sets only can deal with the discrete values, so we need perform some transformations from the quantitative forms into the qualitative values. The simplest rule is the equal interval method, but if the distribution of the criterion values in their value domain is not uniform, the results obtained by this method are not good. Here we use a clustering partition algorithm to realize the discrete process of the quantitative criterion values. The main idea is: first, cluster the classes according to the same values or similar values of the criterion of all inventory items, and then divide all items into several classes; finally, mark the interval labels to represent a qualitative value of this criterion. The number of clustering partition can affect the results of decision rules. In real application, decision maker can determine the suitable interval size based on their experiments, judgment and knowledge. Generally, the important criterion should be partitioned into the more intervals. We give the transformed rules shown in the table 1 after a series of tests.

**TABLE 1. The transformation rules**

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Qualitative Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit-price</td>
<td>$[0.1,1] \cup [1,10]; [10,12] \cup [12,20]; [20,30] \cup [30,48]; [48,65] \cup [65,116]; [116,161]; [161, +\infty]$</td>
</tr>
<tr>
<td>Ordering-cost</td>
<td>$[0.2,]; [2,8]; [8,150]; [150,300]; [300, +\infty]$</td>
</tr>
<tr>
<td>Demand-range</td>
<td>$[1,2]; [3,10]; [11,20]; [21,40]; [41,75]$</td>
</tr>
<tr>
<td>Lead-time</td>
<td>$[1,2]; [3,5]; [6,14]; [15,35]; [36,70]$</td>
</tr>
</tbody>
</table>

To produce unbiased estimates for the probability of misclassification, 2-fold cross validation for re-sampling method is used. The experiment data is split into two sets, 48 items are used for training and 47 items are used for testing.

Running the criterion reduction algorithm over the inventory classification dataset has resulted in
the three attributes being selected for ABC classification problem. The reduction experiments have been performed for two repetitions using 2-fold cross validation. Table 2 lists the reduction attributes and the quality of approximation at every step and gives the final reduction results. They are unit-price, demand-range and lead-time, and the decision table consisting of 65 objects has been reduced to that of 57 objects.

**TABLE 2.** The results after reduction of criteria

<table>
<thead>
<tr>
<th>Exp. No.</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality in step 1</td>
<td>{U} 0.0625</td>
<td>(U) 0.4583</td>
</tr>
<tr>
<td>Quality in step 2</td>
<td>{U, L} 0.625</td>
<td>(U,L) 0.8333</td>
</tr>
<tr>
<td>Quality in step 3</td>
<td>{U, L, D} 0.8750</td>
<td>(U, L, D) 0.8750</td>
</tr>
<tr>
<td>Reduction results</td>
<td>{U, L, D}</td>
<td>{U, L, D}</td>
</tr>
</tbody>
</table>

**4.2. The Classification and testing by SVMs**

The 65 samples are used to train and the rest 30 samples are used to test. The experiments have been performed by SVMs and SVMs+RS. The parameters of SVMs are adjusted accordingly, the results are shown as Table 3.

In the experiment, the first row results are trained and tested by SVM individually. The second row results are trained and tested by RS and SVM. And the three models, One-against-one, One-against-rest and hierarchy, have been applied to prove the proposed method. As shown in table 3, RS is used for feature selection from conditional and decision attributes; the results of reduction are trained and tested by SVMs. By this method, the generalization performance classified by SVMs+RS can be increased compared with SVMs alone.

**TABLE 3.** The comparisons of the three classification precisions

<table>
<thead>
<tr>
<th>Param.</th>
<th>One-against-one</th>
<th>One-against-rest</th>
<th>Hierarchy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SVM</td>
<td>SVM+RS</td>
<td>SVM</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>P</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>P</td>
<td>T</td>
</tr>
<tr>
<td>$\sigma^2 = 0.1$</td>
<td>0.77</td>
<td>0.67</td>
<td>0.88</td>
</tr>
<tr>
<td>C=300</td>
<td>$\sigma^2 = 1$</td>
<td>0.71</td>
<td>0.60</td>
</tr>
<tr>
<td>C=300</td>
<td>$\sigma^2 = 5$</td>
<td>0.71</td>
<td>0.60</td>
</tr>
<tr>
<td>C=300</td>
<td>$\sigma^2 = 10$</td>
<td>0.71</td>
<td>0.60</td>
</tr>
</tbody>
</table>
| C=1000 | Where “T” and “P” represents the training and the testing phase respectively.

**5. Conclusion**

A hybrid heuristic model for multi-criteria classification of inventory items is presented in this paper. The whole procedure includes two steps. The first one is to apply the advantages of rough sets (RS) to handle the qualitative criteria and realize the reduction of multi-criterion attributes without information loss. The second one is to further classify the items of SKU based on SVMs. SVMs have been proven to successfully solve many real world applications. SVMs can find maximal margin boundaries between the classes, minimize the structural risk, and therefore endow the classifier with the excellent generalization performance. Finally, the comparable simulation experiments for the real inventory classification problem have been performed using the method proposed in this paper. It is validated that the use of the hybrid model is a promising analytical tool in the real decision-making systems.
References


Received on the 17th of December 2013