Expert Systems with Applications 36 (2009) 8932–8937

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A rough set-based multiple criteria linear programming approach for the medical diagnosis and prognosis

Zhiwang Zhang [a,b,*], Yong Shi [b,c], Guangxia Gao [d]

[a] School of Information of Graduate University of Chinese Academy of Sciences, China
[b] Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences, No. 80 Zhongguancun East Road, Beijing 100080, China
[c] College of Information Science and Technology, University of Nebraska at Omaha, Omaha, NE 68182, USA
[d] Foreign Language Department, Shandong Institute of Business and Technology, Yantai, Shandong 264005, China

ARTICLE INFO

Keywords:
Data mining
Rough set
Multiple criteria linear programming
Classification

ABSTRACT

It is well known that data mining is a process of discovering unknown, hidden information from a large amount of data, extracting valuable information, and using the information to make important business decisions. And data mining has been developed into a new information technology, including regression, decision tree, neural network, fuzzy set, rough set, and support vector machine. This paper puts forward a rough set-based multiple criteria linear programming (RS-MCLP) approach for solving classification problems in data mining. Firstly, we describe the basic theory and models of rough set and multiple criteria linear programming (MCLP) and analyse their characteristics and advantages in practical applications. Secondly, detailed analysis about their deficiencies are provided, respectively. However, because of the existing mutual complementarities between them, we put forward and build the RS-MCLP methods and models which sufficiently integrate their virtues and overcome the adverse factors simultaneously. In addition, we also develop and implement these algorithm and models in SAS and Windows system platforms. Finally, many experiments show that the RS-MCLP approach is prior to single MCLP model and other traditional classification methods in data mining, and remarkably improve the accuracy of medical diagnosis and prognosis simultaneously.

1. Introduction

Data mining has been used by many organizations to extract information or knowledge from large volumes of data and then use the valuable information to make critical business decisions. Consequently, analysis of the collected history data in data warehouse or in data mart can gain better insight into your customers and evaluation of the medical diagnosis and prognosis (Mangasarian et al., 1995), improve the quality of decision-making and effectively increase the opportunity of the curability for these vital illness.

From the aspect of methodology, data mining can be performed through association, classification, clustering, prediction, sequential patterns, and similar time sequences (Han & Kamber, 2001). For classification, data mining algorithms use the existing data to learn decision functions that map each case of the selected data into a set of predefined classes. Among various mathematical tools including statistics, decision trees, fuzzy set, rough set and neural networks, linear programming (Dantzig & Thapa, 1997; Pardalos & Hearn, 2005; Yosukizu, Hirotaki, & Tanino, 1985) has been initiated in classification for more than 20 years (Freed & Glover, 1981). Given a set of classes and a set of attribute variables, one can use a linear programming model to define a related boundary value separating the classes. Each class is then represented by a group of constraints with respect to a boundary in the linear program. The objective function minimizes the overlapping rate of the classes or maximizes the distance between the classes (Kou et al., 2003; Shi, Wise, Luo, & Lin, 2001). The linear programming approach results in an optimal classification. It is also flexible to construct an effective model to solve multi-class problems.

However, the MCLP model is not good at dimensional reduction and at removing information redundancy, especially facing many attributes with a large number of data. To our joy, rough set can find the minimal attribute set and efficiently remove redundant information (Pawlak, 1982). Consequently, the developing approach of RS-MCLP to data mining is promising to overcome these disadvantages.

In this paper, we will give a full description of the rough set-based MCLP method and model for classification in data mining. First a detailed introduction of MCLP model and rough set in the related work section is given, including the algorithms of the MCLP

* Corresponding author. Address: Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences, No. 80 Zhongguancun East Road, Beijing 100080, China. Tel./fax: +86 010 82680698.
E-mail addresses: zzwmis@163.com (Z. Zhang), yshi@gucas.ac.cn (Y. Shi), Gaogungxia2006@126.com (G. Gao).

0957-4174/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.11.007

model, rough set for feature selection and their virtues of classification. Then we put forth the methodology of the rough set-based MCLP model after the analysis of their deficiencies, respectively, and implement the combinational model in SAS and Windows platform. And then we describe the advantages of the RS-MCLP model. Finally we present a comprehensive example in different data set and experimental conclusions.

2. Related work

2.1. MCLP approach for classification

A general problem of data classification by using multiple criteria linear programming can be described as the following: given a set of n variables or attributes in database $A = (a_1, a_2, \ldots, a_n)$, let $A_i = (a_{i1}, a_{i2}, \ldots, a_{in}) \in R^n$ be the sample observations of data for the variables, where $i = 1, 2, \ldots, l$ and l is the sample size. If a given problem can be predefined as s different classes, $C_1, C_2, \ldots, C_s$, then the boundary between the jth and $(j+1)$ th classes can be b_j, $j = 1, 2, \ldots, s-1$. Then we determine the coefficients for an appropriate subset of the variables which can be represent the whole of decision space, denoted by $X = (x_1, x_2, \ldots, x_m) \in R^m$ $(m \leqslant n)$, and scalars b_j such that the separation of these classes can be described as follows:

$$A_iX \leqslant b_1,\ \forall A_i \in C_1 \text{ and } b_{k-1} \leqslant A_iX \leqslant b_k,\ \forall A_i \in C_k,\ k = 2, \ldots, s-1, \tag{1}$$

and $A_iX \geqslant b_{s-1}, \forall A_i \in C_s$ where $\forall A_i \in C_j,\ j = 1, 2, \ldots, s$, means that the data case A_i belongs to the class C_j.

For a binary classification, we need to choose a boundary b to separate two classes: G (Goods) and B (Bads); For the purpose of simplification, we present only descriptions about binary classification, which can be extended easily into multiple classification circumstances. That is

$$A_iX \leqslant b,\ A_i \in G \text{ and } A_iX \geqslant b,\ A_i \in B, \tag{1'}$$

where A_i are the vector value of the subset of the variables.

For better separation of Goods and Bads, someone considered the two measurements of the overlapping degree with respect to A_i and the distance where A_i departed from its adjusted boundary b, respectively (Freed & Glover, 1981). Subsequently, Glover introduced the two factors above in models (Glover, 1990). Consequently, we have the following conclusions.

Let α_i be the overlapping degree as described above, and we want to minimize the sum of α_i, then the primal linear programming can be written as

$$\text{Minimize} \sum_i \alpha_i, \text{subject to} : A_iX \leqslant b + \alpha_i,\ A_i \in G \text{ and}$$
$$A_iX \geqslant b - \alpha_i,\ A_i \in B. \tag{2}$$

Let β_i be the distance as defined above too, and we want to maximize the sum of β_i, then the primal linear programming can be expressed as

$$\text{Maximize} \sum_i \beta_i,\ \text{subject to} : A_iX \geqslant b - \beta_i,\ A_i \in G \text{ and}$$
$$A_iX \leqslant b + \beta_i,\ A_i \in B. \tag{3}$$

If considering the two measurements in classification simultaneously, we will get hybrid multiple criteria linear programming model as follows:

$$\text{Minimize} \sum_i \alpha_i \text{ and Maximize} \sum_i \beta_i,\ \text{subject to} : A_iX$$
$$\leqslant b + \alpha_i - \beta_i, A_i \in G \text{ and } A_iX \geqslant b - \alpha_i + \beta_i,\ A_i \in B, \tag{4}$$

where A_i are given, X and b are unrestricted, and α_i and $\beta_i \geqslant 0$.

Furthermore, the compromise solution approach has been used to improve the above model (4) in business practices (Shi and Yu, 1989). It is assumed that the ideal value of $-\sum_i \alpha_i$ be $\alpha^*(\alpha^* > 0)$, at the same time, the ideal value of $\sum_i \beta_i$ be β^* $(\beta^* > 0)$. Then, if $-\sum_i \alpha_i > \alpha^*$, the regret measure is defined as $-d_\alpha^+ = \alpha^* + \sum_i \alpha_i (d_\alpha^+ \geqslant 0))$; otherwise, it is 0. If $-\sum_i \alpha_i < \alpha^*$, the regret measure is also written as $d_\alpha^- = \alpha^* + \sum_i \alpha_i (d_\alpha^- \geqslant 0)$; otherwise it is 0.

Thus, we have $\alpha^* + \sum_i \alpha_i = d_\alpha^- - d_\alpha^+$ and $|\alpha^* + \sum_i \alpha_i| = d_\alpha^- + d_\alpha^+$. Similarly, we have $\beta^* + \sum_i \beta_i = d_\beta^- - d_\beta^+$ and $|\beta^* - \sum_i \beta_i| = d_\beta^- + d_\beta^+$, $d_\beta^+ \geqslant 0$, $d_\beta^- \geqslant 0$. To sum up, the improved MCLP model which we use for modeling in this paper may be expressed as

$$\begin{aligned} &\text{Minimize} : d_\alpha^- + d_\alpha^+ + d_\beta^- + d_\beta^+, \\ &\text{Subject to} : \alpha^* + \sum_i \alpha_i = d_\alpha^- - d_\alpha^+ \text{ and } \beta^* + \sum_i \beta_i = d_\beta^- - d_\beta^+, \\ &\qquad A_iX = b + \alpha_i - \beta_i, A_i \in G \text{ and } A_iX = b - \alpha_i + \beta_i, A_i \in B, \end{aligned} \tag{5}$$

where A_i, α^* and β^* are given, X and b are unrestricted, and α_i, β_i, $d_\alpha^-, d_\alpha^+, d_\beta^-, d_\beta^+ \geqslant 0$.

Owing to the following characteristics, MCLP models are more popular correspondingly than traditional nonlinear models, (a) Simplicity, from algorithm to model results MCLP are very easy to understand and explain. (b) Flexibility, user may freely input different parameters to adjust model performance and get better effects. (c) Generalization, because of systematic consideration to the best trade-off between minimizing the overlapping degree and maximizing the distance departed from boundary, the model will gain better classification correct rate and generalization of training set and test set.

2.2. Rough sets-based feature selection method

On account of the deficiency which MCLP model failed to make sure and remove the redundancy in variables or attributes set. That is to say the model is not good at giving judgment on attributes which are useful and important or unnecessary and unimportant relatively. However, rough set methods have an advantage in this aspect.

Rough set theory which was developed by Pawlak is a new mathematical analysis method for dealing with fuzzy and uncertain information and discovering knowledge and rules hid in data or information (Pawlak et al., 1982). Besides, knowledge or attribute reduction is one of the kernel parts of rough sets, and it can efficiently reduce the redundancy in knowledge base or attribute set (Bhatt & Gopal, 2005; Lin & Liau, 2005; Peters & Skowron, 2004; Swiniarski & Skowron, 2003; Zhai, Khoo, & Fok, 2006).

For supervised learning, a decision system or decision table may often be the form $A = (U, A \cup \{d\})$, where U is a nonempty finite set of objects called the universe, A is a nonempty finite set of attributes, $d \notin A$ is the decision attribute. The elements of A are called conditional attributes or simple conditions.

And a binary relation $R \subseteq X \times X$ which is reflexive (i.e. an object is in relation with itself xRx), symmetric (i.e. if xRy then yRx) and transitive (if xRy and yRz then xRz) is called an equivalence relation. The equivalence class of an element $x \in X$ consists of all objects $y \in X$ such that xRy.

Let $A = (U, A)$ be an information system, then with any $B \subseteq A$ there is associated an equivalent relation $IND_A(B)$: $IND_A(B) = \{(x, x') \in U^2 | \forall a \in B, a(x) = a(x')\}$, here $IND_A(B)$ is called B-indiscernibility relation (Zhao, Yao & Luo, 2006). If $(x, x') \in IND_A(B)$, then objects x and x' are indiscernible from each other by attributes from B. Then the equivalence classes of the B-indiscernibility relation are denoted $[x]_B$. An equivalence relation induces a partitioning of the universe U. These partitions can be used to build new subsets of the universe. Subsets that are most often of interest have the same value of the outcome attribute (Komorowski & Polkowski, 1998).