Expert Systems with Applications 39 (2012) 6123–6128

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Rough set and scatter search metaheuristic based feature selection for credit scoring

Jue Wang [a,*], Abdel-Rahman Hedar [b], Shouyang Wang [a], Jian Ma [c]

[a] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, PR China
[b] Department of Computer Science, Faculty of Computers and Information, Assiut University, Egypt
[c] Department of Information Systems, City University of Hong Kong, Hong Kong

ARTICLE INFO

Keywords:
Credit scoring
Feature selection
Rough set
Scatter search
Meta-heuristics

ABSTRACT

As the credit industry has been growing rapidly, credit scoring models have been widely used by the financial industry during this time to improve cash flow and credit collections. However, a large amount of redundant information and features are involved in the credit dataset, which leads to lower accuracy and higher complexity of the credit scoring model. So, effective feature selection methods are necessary for credit dataset with huge number of features. In this paper, a novel approach, called RSFS, to feature selection based on rough set and scatter search is proposed. In RSFS, conditional entropy is regarded as the heuristic to search the optimal solutions. Two credit datasets in UCI database are selected to demonstrate the competitive performance of RSFS consisted in three credit models including neural network model, J48 decision tree and Logistic regression. The experimental result shows that RSFS has a superior performance in saving the computational costs and improving classification accuracy compared with the base classification methods.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Credit scoring is a method modeling potential risk of credit applications, which has experienced two decades of rapid growth with significant increases in auto-financing, credit card debt, and so on. The advantages of credit scoring include reducing the cost of credit analysis, enabling faster credit decisions, closer monitoring of existing accounts, and prioritizing collections (Brill, 1998). Traditionally, logistic regression (Henley, 1995) and discriminant analysis (Wiginton, 1980) are the most widely used approaches when assessing customer credit risk. And then a series of non-parametric methods from the machine learning, data mining, artificial intelligence and operations research communities have been employed, including *k*-nearest neighbor (Henley & Hand, 1996), genetic programming (Ong, Huang, & Tzeng, 2005), decision tree (Davis, Edelman, & Gammerman, 1992), support vector machines (Huang, Chen, Hsu, Chen, & Wu, 2004) and neural network (Malhotra & Malhotra, 2002; West, 2000).

With the growth of the credit industry and the large loan portfolios under management today, credit industry is actively developing more accurate credit scoring models. However, credit scoring datasets containing huge number of features are often involved, which leads to high complexity and be instable with high-dimensional data for many credit scoring methods. Hence, feature selection will be necessary for significantly reducing the burden of computing and improving the accuracy of the credit scoring models (Isabelle & Andre, 2003; Liu & Motoda, 1998). This effort is leading to the investigation of effective approach to feature selection for credit scoring applications.

Feature selection is one of the most fundamental problems in the field of machine learning. The main aim of feature selection is to determine a minimal feature subset from a problem domain while retaining a suitably high accuracy in representing the original features. Actually, feature selection is a process of finding a subset of features that ideally is necessary and sufficient to describe the target concept from the original set of features in a given data set.

Due to the abundance of noisy, irrelevant or misleading features, the ability to handle imprecise and inconsistent information in real world problems has become one of the most important requirements for feature selection. Rough sets theory, proposed by Pawlak, is a novel mathematic tool handling uncertainty and vagueness, and inconsistent data (Pawlak, 1982, 1991; Pawlak & Skowron, 2000). The rough set approach to feature selection is to select a subset of features (or attributes), which can predict or classify the decision concepts as well as the original feature set (Swiniarski & Andrzej, 2003). Obviously, feature selection is an attribute subset selection process, where the selected attribute subset not only retains the representational power, but also has minimal redundancy. There are many rough set algorithms for feature selection. The simplest approach is based on calculation

* Corresponding author.
E-mail address: wjue@amss.ac.cn (J. Wang).

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.11.011

of a core for discrete attribute data set containing strongly relevant features, and reducts contain a core plus additional weakly relevant features. Mutual information and discernibility matrix based feature selection methods have been proposed in some literatures. In addition, many researchers have endeavored to develop some global optimization algorithms based on genetic algorithm, ant colony optimization, simulated annealing, Tabu search and others (Jensen & Shen, 2003, 2004; Jelonek, Krawiec, & Slowinski, 1995; Hedar, Wang, & Fukushima, 2008; Tan, 2004; Zhai et al., 2002). These techniques have been successfully applied to data reduction, text classification and texture analysis (Lin, Yao, & Zadeh, 2002).

In this paper, a novel method of feature selection based on rough sets and scatter search (RSFS) is proposed for credit scoring data. Scatter search meta-heuristic is an artificial-evolutionary-based algorithm and lies among memory-based heuristics (Glover, 1977, 1998; Glover, Laguna, & Mart, 2003; Glover & Kochenberger, 2003). However, the contributions of memory-based heuristics to information systems and data mining applications are still limited compared with other computing intelligence tools like evolutionary computing and neural networks (Osman & Kelly, 1996; Rego & Alidaee, 2005). RSFS uses a binary representation of solutions in the process of feature selection. The conditional entropy is invoked to measure the quality of solution and regarded as a heuristic. The numerical results from two credit datasets indicate that the proposed method shows a superior performance in saving the computational costs and improving the accuracy of several credit models including neural network, logistic regression and J48 decision tree.

The rest of the paper is organized as follows. In the next section, we briefly give the principles of rough set. In Section 3, we highlight the main components of RSFS and present the algorithm formally. In Section 4, numerical results of RSFS are illustrated, and the classification results are presented comparing with the models without feature selection. Finally, the conclusion makes up Section 5.

2. Rough sets preliminaries

An information system is a formal representation of a dataset to be analyzed, and it is defined as a pair $S = (U, \mathbb{A})$, where U is a non-empty set of finite objects, called the universe of discourse, and $\mathbb{A}$ is a non-empty set of attributes. With every attribute $a \in \mathbb{A}$, a set of its values V_a is associated (Pawlak, 1991). In practice, we are mostly interested in dealing with a special case of information system called a decision system. It consists of a pair $S = (U, \mathbb{C} \cup \mathbb{D})$, where $\mathbb{C}$ is called a conditional attributes set and $\mathbb{D}$ a decision attributes set.

The RS theory is based on the observation that objects may be indiscernible (indistinguishable) because of limited available information. For a subset of attributes $P \subseteq \mathbb{A}$, the indiscernibility relation is defined by $IND(P)$ (Pawlak, 1991):

$$IND(P) = \{(\xi, \eta) \in U \times U | \forall a \in P, a(\xi) = a(\eta)\}.$$

It is easily shown that $IND(P)$ is an equivalence relation on the set U. The relation $IND(P)$, $P \subseteq \mathbb{A}$, constitutes a partition of U, which is denoted $U/IND(P)$. If $(\xi, \eta) \in IND(P)$, then ξ and η are indiscernible by attributes from P. The equivalence classes of the P-indiscernibility relation are denoted $[\xi]_P$. For a subset $\Xi \subseteq U$, the P-lower approximation and P-upper approximation of Ξ can be defined as follows, respectively,

$$\underline{P}\Xi = \{\xi | [\xi]_P \subseteq \Xi\},$$
$$\overline{P}\Xi = \{\xi | [\xi]_P \cap \Xi \neq \varnothing\}.$$

As an illustrative example, Table 1(i) shows a dataset which consists of three conditional attributes $\mathbb{C} = \{a, b, c\}$, one decision attribute $\mathbb{D} = \{d\}$, and six objects $U = \{e1, e2, \ldots, e6\}$. If $P = \{a, b\}$, then objects $e1$, $e2$, and $e3$ are indiscernible, and so are objects $e4$ and $e6$. Thus, $IND(P)$ yields the following partition of U:

Table 1
An example of reducts.

(i) A dataset					(ii) A reduced dataset				(iii) A reduced dataset			
U	a	b	c	d	U	a	c	d	U	b	c	d
$e1$	0	0	0	1	$e1$	0	0	1	$e1$	0	0	1
$e2$	0	0	1	0	$e2$	0	1	0	$e2$	0	1	0
$e3$	0	0	2	0	$e3$	0	2	0	$e3$	0	2	0
$e4$	1	0	0	1	$e4$	1	0	1	$e4$	0	0	1
$e5$	1	1	1	1	$e5$	1	1	1	$e5$	1	1	1
$e6$	1	0	2	0	$e6$	1	2	0	$e6$	0	2	0

$$U/IND(P) = \{\{e1, e2, e3\}, \{e4, e6\}, \{e5\}\}.$$

For example, if $\Xi = \{e1, e4, e5\}$, then $\underline{P}\Xi = \{e5\}$, $\overline{P}\Xi = \{e1, e2, e3, e4, e5, e6\}$; if $\Xi = \{e2, e3, e6\}$, then $\underline{P}\Xi = \phi$, $\overline{P}\Xi = \{e1, e2, e3, e4, e6\}$.

For an information system $S = (U, \mathbb{A})$ and a partition of U with classes X_i, $1 \leqslant i \leqslant n$, the entropy of attribute set $B \subseteq A$ is defined as (Pal, Uma Shankar, & Mitra, 2005)

$$H(B) = -\sum_{i=1}^{n} (p(X_i)) log(p(X_i))$$

where $IND(B) = \{X_1, X_2, \ldots, X_n\}$, $p(X_i) = |X_i|/|U|$, and $|\cdot|$ is the cardinality.

The conditional entropy of an attribute set D with reference to another attribute set B is defined as follows:

$$H(D|B) = -\sum_{i=1}^{n} (p(x_i)) \sum_{j=1}^{m} (p(Y_j|X_i)) log(p(Y_j|X_i))$$

where $p(Y_j|X_i) = |Y_j \cap X_i|/|X_i|$, and $U/IND(D) = \{Y_1, Y_2, \ldots, Y_m\}$, $U/IND(B) = \{X_1, X_2, \ldots, X_n\}$, $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant m$. And some theorems have been proved in some literatures.

Theorem 2.1. *$H(D|B) = H(D \cup B)) - H(B)$.*

Theorem 2.2. *If P and Q are the attribute sets of an information system S = (U,A) and IND(Q) = IND(P), then H(Q) = H(P).*

Theorem 2.3. *If P and Q are the attribute sets of an information system S = (U,A), $P \subseteq Q$ and H(Q) = H(P), then IND(Q) = IND(P).*

Theorem 2.4. *An attribute r in an attribute set $R \subseteq C$ is reducible if and only if $H(r|R - \{r\}0 = 0$.*

Theorem 2.5. *Given a relatively consistent decision table S = (U, C ∪ D), an attribute set R is relatively independent if and only if H(D|R) = H(D|R − r) for every $r \in R$.*

Theorem 2.6. *Given a relatively consistent decision table S = (U, C ∪ D), an attribute set $R \subseteq C \subseteq A$ of an information system S = (U,A) is a reduct of B if and only if*

(1) *H(R) = H(C).*
(2) *The attribute set R is relatively independent.*

Consider the dataset shown in Table 1(i), and let $P = \{a,b\}$ and $Q = \{d\}$. Then

$$U/IND(Q) = \{\{e1, e4, e5\}, \{e2, e3, e6\}\},$$
$$U/IND(P) = \{\{e1, e2, e3\}, \{e4, e6\}, \{e5\}\},$$
$$H(Q|P) = 0.7925.$$

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
✓ امکان دانلود نسخه ترجمه شده مقالات
✓ پذیرش سفارش ترجمه تخصصی
✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
✓ دانلود فوری مقاله پس از پرداخت آنلاین
✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات