Expert Systems with Applications xxx (2008) xxx–xxx

Contents lists available at ScienceDirect

Expert Systems with Applications

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

A novel decision rules approach for customer relationship management of the airline market

James J.H. Liou *

Department of Air Transportation, Kainan University, No. 1, Kainan Road, Luchu, Taoyuan 338, Taiwan

ARTICLE INFO

Keywords:
Dominance
Rough set
Customer relationship management
Data mining
Factor analysis

ABSTRACT

Customer churn means the loss of existing customers to a competitor. Accurately predicting customer behavior may help firms to minimize this loss by proactively building a lasting relationship with their customers. In this paper, the application of the factor analysis and the Variable Consistency Dominance-based Rough Set Approach (VC-DRSA) in the customer relationship management (CRM) of the airline market is introduced. A set of "if...then..." decision rules are used as the preference model to classify customers by a set of criteria and regular attributes. The proposed method can determine the competitive position of an airline by understanding the behavior of its customers based on their perception of choice, and so develop the appropriate marketing strategies. A large sample of customers from an international airline is used to derive a set of rules and to evaluate its prediction ability.

1. Introduction

Customer relationship management (CRM) is crucial in today's airline business because of globalization, increasing competition, market saturation and rapid advances in technology. The aim of CRM is to understand the profitability of their customers and to retain the profitable ones. Therefore, many firms need to be able to determine the value of their customers in order to retain or even cultivate the potential profit of customers (Hawkes, 2000). CRM is a dynamic process of managing a customer–company relationship such that customers elect to continue mutually beneficial commercial exchanges and at the same time are dissuaded from participating in exchanges that are unprofitable to the company (Bergeron, 2002). CRM is a key business strategy in which a firm needs to stay focused on the needs of its customers and must integrate a customer-oriented approach throughout the organization.

The trend of increasing competition and decreasing customer loyalty have led to the emergence of concepts that push from a product orientation to a customer orientation and that define their market strategy from the outside-in and not from the inside-out. The focus here is on customer needs rather than on product features (Ozgener & Iraz, 2006). This shift in organizational culture challenges airlines to revise their organizational system and processes, identify customer-related metrics, and identify areas of strategic advantage. To address this customer focus, discussion on data management, availability, data warehousing, and data mining are occurring at various levels within the airline companies, from booking, check-in, cabin service, customer complaint handling to frequent flyer incentives. An important driver of this change is the advent of CRM, which is underpinned by the information and communication technologies (Ryals & Knox, 2001). Thus, a clear shift toward data-based decision making, using so-called data mining or knowledge discovering techniques is evident.

Data mining – the extraction of hidden predictive information from a large database – is a useful tool for airlines that can identify valuable customers, predict future behaviors, and enables firms to make proactive, knowledge-driven decisions. The Variable Consistency Dominance-based Rough Set Approach (VC-DRSA) originally developed by Greco, Matarazzo, and Slowinski (1998, 2000) and extended by Blaszczynski, Greco, and Slowinski (2007) is a relatively new approach in data mining, and is very useful for data reduction in both quantitative and qualitative analysis. The decision rule preference model resulting from the VC-DSRA can even represent inconsistent preferences (Blaszczynski et al., 2007). Unlike conventional data analysis, which uses a statistical inferential technique, the rough set approach is based on data mining techniques for discovering knowledge (Goh & Law, 2003). According to Zhu, Premkumar, Zhang, and Chu (2001), the rough set method does not require additional information about the data; it can work with imprecise values or uncertain data, and is able to discover important facts hidden in that data and express them in natural language. The rough set theory has been successfully applied in a variety of fields, including: evaluation of bankruptcy risk (Slowinski & Zopounidis, 1995), business failure prediction (Beynon & Peel, 2001), travel demand analysis (Goh & Law, 2003), mining stock prices (Wang, 2003), insurance market (Shyng, Wang, Tzeng, & Wu, 2007), accident prevention (Wong & Chung, 2007), customers'

* Tel.: +886 3 3412500x6064; fax: +886 3 3416912.
E-mail address: jhliou@mail.knu.edu.tw

0957-4174/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.05.002

classification of telecommunication services (Blaszczynski et al., 2007) etc.

The objective of this research was to apply the VC-DRSA data mining technique to investigate the behaviors of customers in the airline market, and to develop an appropriate CRM strategy for personalized marketing that could contribute to the enhancement of the long-term relationships with exiting customers. The rest of this paper is structured as follows: In Section 2, some of the important previous researches regarding CRM are summarized, and in Section 3, the basic concept of the VC-DRSA is introduced. In Section 4, an empirical example is illustrated for use in the validation of the proposed model. The results and discussions of the empirical study are presented in Section 5. Finally, in Section 6 some conclusions are drawn.

2. A brief review of CRM

With the ever-increasing competition for marketing dominance, many firms have utilized the CRM system for improved business intelligence, better decision making, enhanced customer relations, and increased quality of services and product offerings. The underpinning of the customer-oriented management concept is that identification and satisfaction of customers must lead to improved customer retention, which is based on corporate profitability (Roh, Ahn, & Han, 2005). There are various definitions of CRM in the literature. Among the most representative, Day and Van den Bulte (2002) define CRM as a cross-functional process for achieving a continuing dialogue with customers, across all their contact and access points, with personalized treatment of the most valuable customers, to increase customer retention and the effectiveness of marketing initiatives. Another one is that CRM is an active, participatory and interactive relationship between business and customer (Ozgener & Iraz, 2006).

Although CRM escalated into a topic of major importance a decade ago, its origins which involve building relationships of mutual value between companies and customers have been in existence since the start of commerce (Gronroos, 1996). CRM takes a wide view and is an attitude towards customers and to the organization itself, which dynamically integrates sales, marketing and the customer care service in order to create and add value for the company and its customers (Chalmeta, 2006). Several recent trends have impacted the ability of organizations to build more enduring relationships, especially for those businesses with a large customer base. Amongst the most important ones are: the increasing power of personal computers; the availability of increasingly sophisticated tools to undertake data mining and data analysis; the rise of e-commerce and the ability to be able to target customers via the Internet at a much lower cost; and an increased recognition of the importance of customer retention and customer lifetime value (Payne & Frow, 2004).

Many studies in marketing suggest that using data mining tools in CRM can enhance a firm's performance. By means of the data mining technology large databases can be explored to find previously unknown relationships and trends that can provide support for complex decisions (Ozgener & Iraz, 2006). Gustafsson, Johnson, and Roos (2005) studied telecommunication services to examine the effects of customer satisfaction and behavior on customer retention. Results indicated a need for CRM managers to more accurately determine customer satisfaction in order to reduce customer churn. Mithas, Krishnan, and Fornell (2005) studied the effects of CRM initiatives showing that CRM efforts improved a firm's knowledge of their customers and in turn, improved customer satisfaction. They also determined that sharing CRM information with suppliers created gains in customer knowledge. Kim, Jung, Suh, and Hwang (2006) proposed a framework for analyzing customer value and segmenting customers based on their value. After segmenting customers based on their value, strategy building based on the mined information can be carried out by the company.

Despite its apparent value, the literature on data mining and its application to CRM in the airline market is virtually silent. In this emerging research area, the present study can provide insight for research and theory development. In the next section, the concept of the VC-DRSA theory will be introduced.

3. The basic concept of VC-DRSA

The rough set theory, first introduced by Pawlak (1982), is a valuable mathematical tool to deal with vagueness and uncertainty (Pawlak, 1997). For a long time, however, the use of the rough set approach and the use of data mining techniques in general were restricted to classification problems where the preference order of evaluations was not considered. This was due to the fact that the classical rough set approach cannot handle inconsistencies (i.e., preferences not satisfying the Pareto-dominance principle) as a result of the violation of the dominance principle (Greco, Matarazzo, & Slowinski, 2001). In order to deal with this particular kind of inconsistency a number of methodological changes to the original rough sets theory were necessary. Greco et al. (1998) proposed an extension of the rough set theory based on the dominance principle, which permits it to deal with inconsistencies. This method is mainly based on substituting the indiscernibility relation by a dominance relation in the rough approximation of decision classes. However, the decision rules induced from the lower approximations of the Dominance-based Rough Set Approach (DSRA) are sometimes weak in that only a few objects support them. For this reason, a variant of DSRA, called VC-DRSA, has been proposed (Blaszczynski et al., 2007). It allows some inconsistency in the lower approximations of sets by a parameter called consistency level. It is more general than the classic functional or relational model and is more understandable for users because of its natural syntax and because it considers the inconsistency of real-life. The basic concepts of the VC-DRSA are described as follows (Blaszczynski et al., 2007; Greco et al., 1998, 2000, 2001, Greco, Matarazzo, & Slowinski, 2002).

3.1. Data table

For algorithmic reasons, the information regarding the objects is supplied in the form of a data table, in which the separate rows refer to distinct objects (actions), and where the columns refer to the different attributes or criteria (attributes with preference-ordered domains) that are considered. Each cell of this table indicates an evaluation (quantitative or qualitative) of the object placed in that row by means of the attribute/criterion in the corresponding column.

Formally, a data table is the 4-tuple information system $IS = (U, Q, V, f)$, where U is a finite set of objects (universe), $Q = \{q_1, q_2, \ldots, q_m\}$ is a finite set of attributes/criteria, V_q is the domain of attribute/criterion q, $V = \cup_{q \in Q} V_q$ and $f: U \times Q \to V$ is a total function such that $f(x,q) \in V_q$ for each $q \in Q$, $x \in U$, called information function. The set Q is usually divided into set C of condition attributes and set D of decision attributes.

3.2. Rough approximation by means of the dominance relationship

Let $\succeq_q$ be an outranking relation to U with reference to criterion $q \in Q$, such that $x \succeq_q y$ means that "x is at least as good as y with respect to criterion q". Suppose that $\succeq_q$ is a complete preorder, i.e., a strongly complete (which means that for each $x, y \in U$, at least one of $x \succeq_q y$ and $y \succeq_q x$ is verified, and thus x and y are always comparable with respect to criterion q) and transitive binary relation. Moreover, let $\boldsymbol{Cl} = \{Cl_t, t \in T\}$, $T = \{1,\ldots,n\}$, be a set of classes of U, such that each $x \in U$ belongs to one and only one class $Cl_t \in \boldsymbol{Cl}$. We assume that all $r, s \in T$, such that $r > s$, each element of Cl_r is preferred to each element Cl_s. In other words, if $\succeq$ is a comprehensive outranking relation on U, then it is supposed that

$$[x \in Cl_r, y \in Cl_s, r > s] \Rightarrow x \succ y, \tag{1}$$

where $x \succ y$ means $x \succeq y$ and not $y \succeq x$.

We can define unions of classes relative to a particular dominated or dominating class – these unions of classes are called upward and downward unions of classes, defined, respectively, as:

$$Cl_t^{\geqslant} = \bigcup_{s \geqslant t} Cl_s, \quad Cl_t^{\leqslant} = \bigcup_{s \leqslant t} Cl_s. \tag{2}$$

It is said that object x P-dominates object y with respect to $P \subseteq C$ (denoted as xD_Py) if $x \succeq_q y$ for all $q \in P$, and $D_P = \cap_{q \in p} \succeq_q$, then the dominance relation D_P is a partial preorder. Given $P \subseteq C$ and $x \in U$, let

$$D_P^+(x) = \{y \in U : yD_Px\}, \tag{3}$$

$$D_P^-(x) = \{y \in U : xD_Py\} \tag{4}$$

representing the P-dominating set and the P-dominated set with respect to x, respectively. In the VC-DRSA, the sets to be approximated are upward and downward unions of classes and the items (granules of knowledge) used for this approximation are dominating and dominated sets.

For any $P \subseteq C$ we say that $x \in U$ belongs to $Cl_t^{\geqslant}$ at consistency level $l \in (0, 1]$, if $x \in Cl_t^{\geqslant}$ and at least $l \times 100\%$ of all objects $y \in U$ dominating x with respect to P also belong to $Cl_t^{\geqslant}$, more formally,

$$\frac{|D_P^+(x) \cap Cl_t^{\geqslant}|}{|D_P^+(x)|} \geqslant l \tag{5}$$

where $|\cdot|$ denotes the cardinality of a set. In other words, if $l < 1$, then $(1 - l) \times 100\%$ of all objects $y \in U$ dominating x with respect to P may not belong to $Cl_t^{\geqslant}$ and thus contradict the inclusion of x in $Cl_t^{\geqslant}$. The P-lower approximation of $Cl_t^{\geqslant}$ and $Cl_t^{\leqslant}$, $t \in \{1,\ldots,n\}$, with respect to $P \subseteq C$ at consistency level $l \in (0,1]$ (denotation $\underline{P}^l(Cl_t^{\geqslant})$ and $\underline{P}^l(Cl_t^{\leqslant})$, respectively), are defined as:

$$\underline{P}^l(Cl_t^{\geqslant}) = \left\{x \in Cl_t^{\geqslant} : \frac{|D_P^+(x) \cap Cl_t^{\geqslant}|}{|D_P^+(x)|} \geqslant l\right\}, \tag{6}$$

$$\underline{P}^l(Cl_t^{\leqslant}) = \left\{x \in Cl_t^{\leqslant} : \frac{|D_P^-(x) \cap Cl_t^{\leqslant}|}{|D_P^-(x)|} \geqslant l\right\}. \tag{7}$$

The P upper approximations of $Cl_t^{\geqslant}$ and $Cl_t^{\leqslant}$, by complementation of $\underline{P}^l(Cl_{t-1}^{\leqslant})$ and $\underline{P}^l(Cl_{t+1}^{\geqslant})$ with respect to U can be obtained as follows:

$$\overline{P}^l(Cl_t^{\geqslant}) = U - \underline{P}^l(Cl_{t-1}^{\leqslant}), \tag{8}$$

$$\overline{P}^l(Cl_t^{\leqslant}) = U - \underline{P}^l(Cl_{t+1}^{\geqslant}). \tag{9}$$

The above $\overline{P}^l(Cl_t^{\geqslant})$ can be interpreted as a set of all the objects belonging to $Cl_t^{\geqslant}$, and may possibly be inconsistent at consistency level l. The P-lower and P-upper approximations defined above satisfy the following properties for all $t \in \{1,\ldots,n\}$ and for any $P \subseteq C$:

$$\underline{P}^l(Cl_t^{\geqslant}) \subseteq Cl_t^{\geqslant} \subseteq \overline{P}^l(Cl_t^{\geqslant}), \quad \underline{P}^l(Cl_t^{\leqslant}) \subseteq Cl_t^{\leqslant} \subseteq \overline{P}^l(Cl_t^{\leqslant}). \tag{10}$$

The P-boundaries (P-doubtable region) of $Cl_t^{\geqslant}$ and $Cl_t^{\leqslant}$ are defined as:

$$Bn_P(Cl_t^{\geqslant}) = \overline{P}^l(Cl_t^{\geqslant}) - \underline{P}^l(Cl_t^{\geqslant}), \tag{11}$$

$$Bn_P(Cl_t^{\leqslant}) = \overline{P}^l(Cl_t^{\leqslant}) - \underline{P}^l(Cl_t^{\leqslant}). \tag{12}$$

We define the accuracy of the approximation of $Cl_t^{\geqslant}$ and $Cl_t^{\leqslant}$ for all $t \in \{1,\ldots,n\}$ and for any $P \subseteq C$, respectively, as

$$\alpha_p(Cl_t^{\geqslant}) = \frac{|\underline{P}^l(Cl_t^{\geqslant})|}{|\overline{P}^l(Cl_t^{\geqslant})|}, \quad \alpha_p(Cl_t^{\leqslant}) = \frac{|\underline{P}^l(Cl_t^{\leqslant})|}{|\overline{P}^l(Cl_t^{\leqslant})|}. \tag{13}$$

The ratio

$$\gamma_p(Cl) = \frac{\left|U - \left(\cup_{t\in\{2,\ldots,n\}} Bn_p(Cl_t^{\geqslant})\right)\right|}{|U|} = \frac{\left|U - \left(\cup_{t\in\{1,\ldots,n-1\}} Bn_p(Cl_t^{\leqslant})\right)\right|}{|U|} \tag{14}$$

defines the quality of the approximation of classification $\boldsymbol{Cl}$ by means of the criteria from set $P \subseteq C$, or, briefly, quality of classification. This ratio expresses the proportion of all P-correctly classified objects, i.e., all the non-ambiguous objects to all the objects in the data table. Every minimal subset $P \subseteq C$ such that $\gamma_P(\boldsymbol{Cl}) = \gamma_C(\boldsymbol{Cl})$ is called a *reduct* of C with respect to $\boldsymbol{Cl}$ and is denoted by $RED_{Cl}(P)$. It should be noted that a data table may have more than one *reduct*. The intersection of all the *reducts* is known as the *core*, denoted by $CORE_{Cl}$.

3.3. Decision rules

The end result of the VC-DRSA is a representation of the information contained in the data table considered in terms of simple "if…,then…" decision rules. For a given upward union of classes $Cl_t^{\geqslant}$ the decision rules, induced under a hypothesis that actions belonging to $\underline{P}^l(Cl_t^{\geqslant})$ are positive and all the others are negative, suggest an assignment to "at least class Cl_t". Analogously, for a given downward union $Cl_s^{\leqslant}$, the rules induced under a hypothesis that actions belonging to $\underline{P}^l(Cl_s^{\leqslant})$ are positive and that all others are negative suggest an assignment to "at most class Cl_s". On the other hand, the decision rules induced under a hypothesis that actions belonging to the intersection $\overline{P}^l(Cl_s^{\leqslant}) \cap \overline{P}^l(Cl_t^{\geqslant})$ are positive and that all the others are negative suggest an assignment to some class between Cl_s and Cl_t ($s < t$). Within the VC-DRSA, decision rules have an additional parameter α, called credibility, $l \leqslant \alpha \leqslant 1$. This parameter says to what degree the rule is certain. Please note that the DRSA rules induced from P-lower approximations of unions of decision classes are equivalent to the VC-DRSA rules induced with $\alpha = 1$.

The following three types of decision rules can be considered:

1. $D_{\geqslant}$-decision rules, which have the following form:If $f(x,q_1) \geqslant r_{q1}$ and $f(x,q_2) \geqslant r_{q2}$ and $\ldots f(x,q_p) \geqslant r_{qp}$, then $x \in Cl_t^{\geqslant}$. These rules are supported only by objects from P-lower approximations of the upward unions of classes $Cl_t^{\geqslant}$.
2. $D_{\leqslant}$-decision rules, which have the following form:if $f(x,q_1) \leqslant r_{q1}$ and $f(x,q_2) \leqslant r_{q2}$ and $f(x,q_p) \leqslant r_{qp}$, then $x \in Cl_t^{\leqslant}$. These rules are supported only by objects from the P-lower approximation of the downward unions of classes $Cl_t^{\leqslant}$.
3. $D_{\geqslant\leqslant}$-decision rules, which have the following form: if $f(x,q_1) \geqslant r_{q1}$ and $f(x,q_2) \geqslant r_{q2}$ and $\ldots f(x,q_k) \geqslant r_{qk}$ and $f(x,q_{k+1}) \leqslant r_{qk+1}$ and $\ldots f(x,q_p) \leqslant r_{qp}$, then $x \in Cl_s \cup Cl_{s+1} \cup \cdots \cup Cl_t$.

These rules are supported only by objects from the P-boundaries of the unions of classes $Cl_s^{\leqslant}$ and $Cl_t^{\geqslant}$, where $P = \{q_1, q_2,\ldots,q_p\} \subseteq C$, $(r_{q1},r_{q2},\ldots, r_{qp}) \in V_{q1} \times V_{q2} \times \cdots \times V_{qp}$ and $t \in \{1,\ldots,n\}$.

It should be noted that the set of decision rules induced by the approximations which were defined using dominance relations gives, in general, a more synthetic representation of the knowledge contained in the decision table than the set of rules induced by the original rough set theory, which uses indiscernible relations. This is due to the more general syntax of the rules ("$\geqslant$" and "$\leqslant$" are used instead of "=").

3.4. Relative strength of rules

The strength of each rule refers to the number of objects belonging to it. But, the number of each class is different. The fact that a rule has a higher strength in one class does not guarantee that its influence is greater than another rule with lower strength in another class. Therefore, the relative strength of each rule RS_ρ of rule ρ with respect to decision class Cl_t can be calculated as follows:

$$RS_\rho(Cl_t) = \frac{|Cond_\rho \cap Cl_t|}{|Cl_t|} \tag{15}$$

where $Cond_\rho$ denotes the set of objects verifying the condition part of rule ρ, and $|Cond_\rho|$ and $|Cond_\rho \cap Cl_t|$ denote the cardinalities of the corresponding set.

4. Empirical study: a case of CRM in the airline market

In order to demonstrate the effectiveness of the VC-DRSA and our proposed approach, an empirical study is presented in this section. The surveyed airline is an international airline in Taiwan flying to more than 40 destinations around the world. First, we employed a survey and a factor analysis to extract some important factors based on the considered attributes/criteria. Second, using the extracted common factor, we conducted a questionnaire about customer behaviors in the airline market using the VC-DRSA to explore the classification problem. The results showed that our approach could provide airlines with information to develop their CRM strategies and to achieve their marketing goals.

4.1. Variable extraction

The buying behaviors of airline customers are complex and are affected by a number of manageable, distinct attributes/criteria. To reduce the complexity of the considered system, we first proposed 18 related attributes/criteria by consulting with the managers of the marketing departments of airlines and through literature review (Kim, 2006; Park, 2007; Roh et al., 2005). These 18 items (Table 1) include price, seating comfort, check-in service, meal service, in-flight entertainment service, safety, schedule, on-time performance, frequent-flyer program, etc. We then conducted the first questionnaire survey. The 92 participants included tour leaders, employees of the marketing department of airlines, and airline customers from different backgrounds. The design of the questionnaire was based on multiple-item measurement scales. The measurement items were designed for the airline and the statements were measured on a five-point Likert-type scale. The collected data were then analyzed to extract the common factors, in order to simplify the system structure and to express the characteristics of the surveyed data by fewer but more representative factors. The principle component analysis was adopted in this study and six common factors were obtained from the results, as shown in Table 1. The variance explained for six common factors were 25%, 16%, 14%, 10%, 10%, and 7%, respectively. The total variance explained reached 81%. There were two common factors including three attributes, two common factors containing two attributes, and one common factor composed of seven attributes, which means that those attributes in the common factor had inter-dependent relationships. After the axis was rotated, seven attributes, A_1–A_7 had a higher loading factor in common factor 1, which was renamed as "service." Attributes A_8–A_{10} formed a new common factor called "time table." Attributes A_{11}–A_{13}, were combined as "facilities and food." And, A_{14}, A_{15}, and A_{16}, A_{17} were then combined as "reliability and safety" and "incentives," respectively. The "price" remained as it was.

4.2. Preparing the information table

Based on the above evaluation process, the results of the first questionnaire indicated that service, timetable, facilities and food, reliability and safety, incentive and price were the most important factors to influence the buying behavior of airline customers. Since demographic variables remain an important reference for airline customer behavior, they are considered as part of the 'other' attributes to be evaluated. Therefore, the second questionnaire contained three parts: (1) the demographic information about the participants, (2) the satisfaction levels of each criterion, and (3) their buying decisions regarding the surveyed airline. The attributes/criteria for customer behavior in the airline market are shown in Table 2. The preference for services, timetables, facilities and food, reliability and safety, incentives and buying behavior were set as gain while the price was set to cost (smaller the better).

Table 1
Factor analysis results after varimax rotated

Common factors	Influential parameters (attributes)	Common factors						Communality h^2
		1	2	3	4	5	6	
1. Services	1. Luggage handling (A_1)	**0.867**	0.068	0.125	0.051	0.083	−0.102	0.904
	2. Attitude of employees (A_2)	**0.833**	0.164	0.160	0.208	0.035	0.012	0.816
	3. Check-in service (A_3)	**0.818**	0.044	−0.02	0.149	0.193	0.084	0.924
	4. Customer complaint handling (A_4)	**0.807**	0.136	0.195	−0.035	0.046	−0.025	0.832
	5. Language skill of employees (A_5)	**0.789**	0.077	0.251	−0.029	−0.04	0.146	0.950
	6. Appearance of employees (A_6)	**0.760**	0.171	0.163	0.020	−0.01	0.331	0.858
	7. Reservation and booking (A_7)	**0.407**	0.234	0.161	0.319	0.050	0.306	0.896
2. Timetables	1. Scheduling (A_8)	0.127	**0.938**	0.098	0.089	0.091	0.059	0.443
	2. Routes (A_9)	0.199	**0.932**	0.096	0.098	0.142	0.051	0.862
	3. Transferring service (A_{10})	0.139	**0.908**	0.070	0.062	0.171	0.120	0.860
3. Facilities and food	1. Comfort and cleanliness of seat (A_{11})	0.163	0.046	**0.897**	0.095	0.138	−0.012	0.869
	2. Meal service (A_{12})	0.249	0.125	**0.859**	0.206	0.106	−0.029	0.791
	3. In-flight entertainment (A_{13})	0.222	0.111	**0.855**	0.168	−0.10	0.169	0.744
4. Reliability and safety	1.Safety record (A_{14})	-0.017	0.102	0.158	**0.892**	0.010	0.040	0.791
	2. On-time performance (A_{15})	0.208	0.079	0.199	**0.841**	0.017	0.138	0.722
5. Incentives	1.Service of lounge room (A_{16})	0.104	0.169	−0.009	−0.009	**0.919**	−0.037	0.712
	2.Frequent-flyer programs (A_{17})	0.043	0.198	0.157	0.047	**0.838**	0.297	0.738
6. Price	1.Ticket Price (A_{18})	0.125	0.135	0.037	0.154	0.178	**0.902**	0.885
Eigenvalue		4.417	2.847	2.584	1.801	1.735	1.216	–
Variance interpreted (%)		24.54	15.82	14.35	10.01	9.640	6.760	–
Accumulated variance (%)		24.54	40.35	54.71	64.71	74.35	81.11	–

*Extraction method: principle component analysis.
*Rotated method: varimax with Kaiser normalization.

Table 2
Specifications of attributes/criteria for customer behavior

Attribute/criterion name	Attribute/criterion value	Value set	Preference
Condition attributes			
Gender	Female; male	{F, M}	None
Marital status	Yes; no	{Y, N}	None
Age	Below 30; 30–40; 40–60; 60 and above	{1, 2, 3, 4}	None
Occupation	Government employee; employee of company; student; others	{G, P, S, O}	None
Education	High school and below; college; master and above	{H, C, M}	None
Income	NT$30,000 and below; 30,001 ~ 70,000; 70,000 and above	{1, 2, 3}	None
Service quality	Poor; medium; satisfactory	{1, 2, 3}	Gain
Timetables	Poor; medium; satisfactory	{1, 2, 3}	Gain
Facilities and food	Poor; medium; satisfactory	{1, 2, 3}	Gain
Reliability and safety	Poor; medium; satisfactory	{1, 2, 3}	Gain
Incentives	Poor; medium; satisfactory	{1, 2, 3}	Gain
Price	Low; medium; high	{1, 2, 3}	Cost
Decision criterion			
Buying behavior	Will not consider; maybe; surely purchase	{1, 2, 3}	Gain

Table 3
The backgrounds of respondents

Distribution	Sample number	Frequency (%)
Gender		
Female	264	55.8
Male	209	44.2
Marital status		
Yes	269	56.9
No	204	43.1
Age		
Less than 30	138	29.2
30–40	135	28.5
40–60	177	37.4
60-above	23	4.9
Occupation		
Government employee	118	24.9
Employee of company	142	20.0
Student	51	10.8
Others	162	44.3
Education		
High school and below	111	23.5
College	306	64.7
Master and above	56	11.8
Income (NTD/per month)		
Less than 30,000	144	30.4
30,000–70,000	225	47.6
70,000 and above	104	22.0

On the other hand, the attributes of personal profile were all set to "no preference". Four hundred and 73 participants answered the second questionnaire. The demographics of the 473 respondents are shown in Table 3.

4.3. Results of the VC-DRSA analysis

In this section, we calculated the VC-DRSA analysis using the prepared information table as described in the previous section. The results of the VC-DRSA analysis consisted of four parts: quality of approximation, rule generation, rule validation and the significance of the condition attributes/criteria.

4.3.1. Quality of approximation

To incorporate the inconsistencies of decision-makers (respondents), we set the consistent level as 0.9, which means we tolerate 10% of the inconsistencies within the considered criteria. The accuracy of the approximation for the three decision classes is shown in Table 4. The results indicate good accuracies for different classes. In general, high values for the quality of classification and accuracies mean that the criteria selected are adequate to approximate the classification. The 'at most 1' class means the class "not consider buying" which has a lower approximation of 82 and an upper approximation of 87. The boundary for 'at most 1' is 5. The accuracy of approximation for 'at most 1' is 0.943. The 'at most 2' class includes the classes "not consider buying" and "maybe", for which the accuracy reaches 0.944. On the other hand, the 'at least 3' class refers to the "surely purchase" class, and its lower and upper approximations are 171 and 188, respectively. The accuracy of 'at least 3' class is 0.909. The 'at least 2' class is composed of the "maybe" and "surely purchase" classes, and its accuracy increases to 0.987. The overall quality of approximation is calculated as $(473 - 22)/(473) = 0.953$.

Table 4
Accuracy of classification for customers' behavior

	At most 1	At most 2	At least 2	At least 3
Lower approximation	82	285	386	171
Upper approximation	87	302	391	188
Boundary	5	17	5	17
Accuracy	0.943	0.957	0.987	0.909

4.3.2. Rule generation

The algorithm for the induction of a set of decision rules was the one used by Greco, Matarazzo, Slowinski, and Stefanowski (2001). In the present study we used the JAMM (Slowinski, 2006) software to analyze the data. Since the setting of the consistency level was 0.9, we used the minimum cover rules (the set does not contain any redundant rules), and there were a total of 56 rules generated from the data. Eliminating the cover strength to less than 10% of each class, Table 5 shows the minimum cover rules. The reduced rule set contains 27 rules, in which eight rules correspond to class 1, eight rules to at most 2, eight rules to class 3, and three rules to at least 2. We can see from Table 5, that if the timetable and facilities are poor, the decision is "not consider buying" and its cover strength is 41 (rule 1). The total number for "not consider buying" is 84. This means that there are nearly half of class 1 respondents who will not consider purchase because they feel the timetable and facilities are poor. Rules 4 and 8 suggest that customers with higher education and more income do not like the service quality. As to the elderly (rule 2), there are 12 persons that do not think the reliability and safety is satisfactory and, therefore, will not consider buying. Looking at Table 3 we can see that there were only 23 respondents whose age was above 60 in the sample data. This indicates that half of the elderly were concerned with the safety

Table 5
Minimum cover rules for relative strength greater than 10% of each class

No.	Conditions	Decision	Confidence Level	Strength
1	(Facilities ⩽ 1) and (Timetable ⩽ 1)	$D \leqslant 1$	0.927	41
2	(Reliability and safety ⩽ 1) and (Age = 4)	$D \leqslant 1$	1	12
3	(Education = M) and (Occupation = O)	$D \leqslant 1$	1	11
4	(Education = M) and (Service ⩽ 1)	$D \leqslant 1$	0.947	19
5	(Age = 2) and (Income = 3) and (Reliability and safety ⩽ 1)	$D \leqslant 1$	0.923	13
6	(Occupation = P) and (Incentives ⩽ 1)	$D \leqslant 1$	0.928	14
7	(Occupation = P) and (Age = 3) and (Reliability and safety ⩽ 1)	$D \leqslant 1$	1	16
8	(Income = 3) and (Service ⩽ 1)	$D \leqslant 1$	0.903	31
9	(Timetable ⩽ 1)	$D \leqslant 2$	0.954	66
10	(Reliability and safety ⩽ 2) and (Price ⩾ 2)	$D \leqslant 2$	0.991	205
11	(Occupation = O) and (Service ⩽ 2)	$D \leqslant 2$	0.923	91
12	(Age = 3) and (Price ⩾ 2)	$D \leqslant 2$	0.927	111
13	(Education = M) and (Timetable ⩽ 2)	$D \leqslant 2$	0.976	42
14	(Occupation = G) and (Facility ⩽ 2)	$D \leqslant 2$	0.952	63
15	(Timetable ⩽ 2) and (Price ⩾ 2)	$D \leqslant 2$	0.941	203
16	(Income = 3) and (Reliability and safety ⩽ 2)	$D \leqslant 2$	0.976	83
17	(Occupation = G) and (Price ⩽ 1) and (Facilities ⩾ 3)	D ⩾ 3	0.971	35
18	(Occupation = S) and (Timetable ⩾ 3) and (Facilities ⩾ 3)	D ⩾ 3	0.917	24
19	(Income = 3) and (Service ⩾ 3) and (Facilities ⩾ 3)	D ⩾ 3	0.911	45
20	(Occupation = G) and (Timetable ⩾ 3) and (Reliability and safety ⩾ 3)	D ⩾ 3	0.903	31
21	(Age = 3) and (Price ⩽ 1) and (Timetable ⩾ 3)	D ⩾ 3	0.909	44
22	(Occupation = P) and (Income = 2) and (Price ⩽ 1)	D ⩾ 3	0.923	26
23	(Income = 3) and (Service ⩾ 3) and (Reliability and safety ⩾ 3)	D ⩾ 3	0.947	19
24	(Education = J) and (Marriage = Y) and (Facilities ⩾ 3) and (service ⩾ 3)	D ⩾ 3	0.909	22
25	(Timetable ⩾ 2)	D ⩾ 2	0.916	407
26	(Income = 2)	D ⩾ 2	0.946	223
27	(Facilities ⩾ 2)	D ⩾ 2	0.921	406

of the surveyed airline when they made their decision not to buy. If the reliability and safety was less than medium and the price was higher than medium, then the decision of 205 respondents was equal to or less than 2 (rule 10). From rule 16 we derived that if the respondents were high income and if the reliability and safety was medium or less, then their decision was at most 2, and the cover strength was 83. Rule 17 indicated that if the airline's facilities were satisfactory and the price was low, then government employees would surely purchase and the cover strength reached 35. For higher income passengers, if the airline could provide better service and facilities, they could become loyal customers for the airline (rule 19). From rules 25 and 27, if either the timetable or the facilities was greater or equal to medium, then the respondents' decision would be at least 2, and their cover strengths were 407 and 406, respectively. This indicates that timetable and facilities attract over 80% of customers' attention. In general, from the set of minimum cover rules, when it comes to customers with a higher education, those with a higher income and the elderly all focused more on reliability and safety. The participants of medium income or government employees (Government employees are an important group for the Taiwan airline market. They have a higher income and more vacation time than their counterparts in the private industry, and the Taiwan government even subsidizes their vacation.) seemed to be the major source of passengers if the timetable and the facilities were satisfactory. It is also worth noting that gender and marital status seemed to have little effect on the condition attributes.

4.3.3. Rule validation

To check the feasibility of the decision rules generated in this study, we applied the 10-fold cross-validation technique. First, we randomly chose 90% of the data to generate the decision rules. Then, the remaining 10% of the data was used to validate the hit rate of the generated decision rules, i.e., the percentage of correct predictions for each class. These procedures were repeated 10 times, and the average hit rate is shown in Table 6.

As shown in Table 6, the overall classification error was 29%, with 332 objects being the correct decision and 140 objects being an incorrect decision. The 134 ambiguous objects meant that the objects were covered by rules that suggest that they be assigned to a different union of classes. This ambiguity may be partly due to the inconsistency of the respondents. It should also be noted that an object which is ambiguous is in a 10-fold cross-validation treated as correctly classified if the interval to which it is classified does not have an empty intersection with the original class. For example, if an object from class 'good' is classified in the class ⟨moderate, good⟩ it is calculated as 0.5 correctly classified.

Table 6
Average hit rate for 10-fold cross-validation

	Class 1	Class 2	Class 3
Class 1	84/30	0/7	0/1
Class 2	0/1	209/151	0/15
Class 3	0/0	0/26	180/108
Unclassified	0		
Ambiguous decision	134		
Overall classification error	0.29		

4.3.4. Significance of the condition attributes/criteria

There is no *reduct* generated from the information system, since all of the condition attributes/criteria are essential to distinguish the classes. However, the significance of the condition attributes/criteria can be measured by their presence on the derived rules. When a condition attribute/criterion shows up more frequently in the cover rules, and the strength of the rules are higher, then it is more likely to be a key factor in customers' decisions. From the derived minimum cover rules, as shown in Table 5, the incidence of attributes/criteria and cover strength – the number of objects matching the rule – are used as a reference for evaluating the importance of the attributes/criteria. Rules number 25 and 27 indicate that the timetable and facilities are the most significant criteria for customers' decisions, with a cover strength of 407and 406, respectively. This implies that most of the participants, when making their decision depended mainly on the timetable and the facilities of the airline. In addition, reliability and safety appears

Table 7
Possible rules for relative strength greater than 40% in each class

No.	Condition	Decision	Confidence level	Strength
1	(Reliability and safety $\leqslant$ 1) and (Timetable $\leqslant$ 1)	$D \leqslant 1$	0.923	54
2	(Income = 3) and (Reliability and safety $\leqslant$ 1) and (Facilities $\leqslant$ 2)	$D \leqslant 1$	0.925	54
3	(Income = 3) and (Reliability and safety $\leqslant$ 1) and (Incentives $\leqslant$ 2)	$D \leqslant 1$	0.907	54
4	(Income = 3) and (Reliability and safety $\leqslant$ 1) and (Timetable $\leqslant$ 2)	$D \leqslant 1$	0.927	55
5	(Education = M) and (Reliability and safety $\leqslant$ 1) and (Timetable $\leqslant$ 2)	$D \leqslant 1$	0.943	35
6	(Education = M) and (Reliability and safety $\leqslant$ 1) and (Facility $\leqslant$ 2)	$D \leqslant 1$	0.942	35
7	(Reliability and safety $\leqslant$ 1) and (Service $\leqslant$ 2) and (Timetable $\leqslant$ 2)	$D \leqslant 1$	0.941	51
8	(Incentives $\leqslant$ 2) and (Price $\geqslant$ 2)	$D \leqslant 2$	0.912	238
9	(Incentives $\leqslant$ 2) and (Service $\leqslant$ 2)	$D \leqslant 2$	0.905	241
10	(Timetable $\geqslant$ 2)	D $\geqslant$ 2	0.916	407
11	(Facilities $\geqslant$ 2)	D $\geqslant$ 2	0.921	406
12	(Price $\leqslant$ 2) and (Service $\geqslant$ 2)	D $\geqslant$ 2	0.928	362
13	(Price $\leqslant$ 1) and (Reliability and safety $\geqslant$ 2) and (Service $\geqslant$ 3)	D $\geqslant$ 3	0.901	130
14	(Price $\leqslant$ 1) and (Timetable $\geqslant$ 3) and (Service $\geqslant$ 3)	D $\geqslant$ 3	0.904	104
15	(Price $\leqslant$ 1) and (Incentives $\geqslant$ 3) and (Service $\geqslant$ 3)	D $\geqslant$ 3	0.901	81

most frequently in the 'at most class 1' (will not consider). This indicates that the main reason for those persons rejecting an airline is because they do not like its reliability and safety.

Another approach for exploring the importance of attributes/criteria can be done by setting the strength level higher than a certain threshold value. That way each rule that shows up in the derived rule set will be ensured to be above a certain percentage of cover strength. Therefore, if we choose to generate all possible rules, the consistency level still be 0.9 and the relative strength of each class be greater than 40%. Some of the important rules for each decision class are shown in Table 7. The resulting rules demonstrate similar results as described earlier, and the reliability and safety and price are the most crucial condition criteria for class 1 and 3, while the timetable and facilities are criteria that are concerned by most of respondents.

5. Discussion

A novel decision rule approach, the VC-DRSA, was applied in the CRM of the airline market and presented in this paper. In contrast to the classic statistical techniques such as discriminant analysis, the strength of the rough set theory is that it requires no underlying statistical assumptions. In particular, the VC-DSRA can handle both attributes/criteria with and without preference order. It also considers the decision-makers' inconsistencies when they make a decision. Taking advantage of the VC-DSRA, this research implemented the surveyed data with either quantitative or qualitative attributes, and with or without order of preference in order to analyze the customer behavior of the airline market. The empirical results showed that it is appropriate to apply the VC-DSRA when mining for knowledge of the airline market. Although the selected attributes/criteria are strongly influenced by the local environment and culture, the proposed method can easily be transformed and extended based on the conditions and culture of other markets in any particular environment.

There are many condition factors that can influence the decision of a customer. To include all of them in the analysis is not realistic; we therefore applied the factor analysis to extract some of the independent common factors and combined those factors that had a high inter-dependence. With the help of factor analysis, we not only can use fewer factors to represent complex systems but we can also avoid too many rules generated in the following VC-DRSA analysis. Table 4 shows that the accuracy of the classification for each class was 0.9 and above, which indicates the narrow boundary between the lower approximation and the upper approximation. This was due to the fact that the VC-DRSA enables relaxation of the conditions for the assignment of objects to lower approximations. Moreover, the total number of minimum cover rules is 56, but there are only 27 rules whose strengths are greater than 10% of the relative strength of each class. This indicates that most of the data derived from this study can be classified into 27 rules with only a few unique data. Contrary to the classical rough set theory, which usually generates too many rules with only little cover strength, the VC-DSRA seems to have better decision rules with stronger cover strength. When there are too many unique rules generated, it is difficult to understand the contents of the data sets that relate to the decision rules in the condition part.

This study demonstrates the customers' behavior for the airline market in Taiwan. From the results of our survey, customers could be divided into three groups, loyal (class $\geqslant$ 3), potential (class $\geqslant$ 2), and to be developed (class $\leqslant$ 1). From the derived rules (Tables 5 and 7) for loyal customers, it can be inferred that low price is acknowledged by most of the loyal customers. We also found that government employees and medium income customers are the main basis of the loyal customers. If the airline wants to keep the loyal customers of high income, it should continue to provide them with superior services and facilities. In general, if the timetable and the facilities are better than medium, over 80% of the participants were willing to consider using the airline. If the airline wants to maintain its competitiveness, it needs to provide better services and more attractive incentives to draw potential customers. It is obvious that for the "to be developed group" reliability and safety record are the critical factors and the reason they hesitate to buy. Although price and quality of service are major concerns in the airline market, an airline's reliability and safety record are the most essential factors affecting the " to be developed " customers' decision. Observing the rules for the "to be developed" group (class $\leqslant$ 1), we can conclude that the high-income earners, the higher educated, and the elderly are the three main customer groups that the airline need to cultivate. In fact, those three groups could be the major customers of business or first class of the airline. It should also be noted that these three groups present a section in society that is growing more than any other section. However, their common concern is reliability and safety. This might be due to the fact that the less than desirable safety record of Asian airlines in general has left most business travelers in Taiwan worried about the safety record of the airline they fly with. Therefore, the key for a successful strategy of the surveyed airline is to rebuild the confidence of its high-income customers in Taiwan. The decision rules indicate that if the safety record of the airline is good, even if the price or the quality of service is less competitive, the customers will be more than happy to consider the airline.

In order to find representative rules and effectiveness of the factor structure, a possible rule generation and a 10-fold cross-validation were conducted. Both results support the decision rules and show a good agreement with an acceptable classification error. While the estimation results show that the accuracy of the approximation, the quality of the approximation and the hit rates are satisfactory, the category of condition attributes/criteria could stand to be increased so as to help the decision maker achieve a more precise judgment. More categories of condition attributes/criteria will generate more refined decision rules that in turn will improve the quality of the strategic decisions.

6. Conclusions

This study combined the factor analysis and the VC-DRSA approach as an operational tool for predicting the purchasing decision of customers in the airline market. The proposed prediction model is in the form of decision rules. This method also provides the airline with information on the strength of particular decision rules covering the considered object. The derived rules can help the airline with developing proper strategies for different classes of customers and improve the airline's CRM. Since the derived rules are supported by real examples, they describe only facts in terms of the most relevant attributes/criteria. The classical rough set theory handles attributes without preference, which may not always be true in the real world. The VC-DRSA includes the extension of the classical rough set theory to qualitative reasoning in the preference-based customers' behavior analysis by substituting the indiscernibility relation with the dominance relation. Therefore, the conflicting preference relations in the customers' behavior assessment are objectively represented without introducing the equivalence class concept of the classical rough-set theory.

An empirical example of a Taiwanese airline demonstrated the advantage of VC-DRSA over other techniques that ignore the background knowledge on preference orders in the attributes domains. This study applied the VC-DRSA in order to mine information from the surveyed data of the airline market and help airlines develop a better CRM.

Acknowledgement

This paper was supported by National Science Committee of Taiwan under Grant no. NSC 97-2919-I-009-002-A1.

References

Bergeron, B. (2002). CRM: the customer isn't always right. *Journal of Corporate Accounting & Finance, 14*(1), 53–57.

Beynon, M. J., & Peel, M. J. (2001). Variable precision rough set theory and data discretisation: An application to corporate failure prediction. *Omega, 29*(6), 561–576.

Blaszczynski, J., Greco, S., & Slowinski, R. (2007). Multi-criteria classification – A new scheme for application of dominance-based decision rules. *European Journal of Operational Research, 181*(3), 1030–1044.

Chalmeta, R. (2006). Methodology for customer relationship management. *The Journal of Systems and Software, 79*(7), 1015–1024.

Day, G. S., & Van den Bulte, C. (2002). *Superiority in customer relationship management: Consequences for competitive advantage and performance*. The Wharton School, University of Pennsylvania (pp. 1–49).

Goh, C., & Law, R. (2003). Incorporating the rough sets theory into travel demand analysis. *Tourism Management, 24*(5), 511–517.

Greco, S., Matarazzo, B., & Slowinski, R. (1998). A new rough set approach to evaluation of bankruptcy risk. In C. Zopounidis (Ed.), *Rough fuzzy and fuzzy rough sets* (pp. 121–136). Dordrecht: Kluwer.

Greco, S., Matarazzo, B., & Slowinski, R. (2000). Extension of the rough set approach to multicriteria decision support. *INFOR, 38*(3), 161–195.

Greco, S., Matarazzo, B., & Slowinski, R. (2001). Rough sets theory for multicriteria decision analysis. *European Journal of Operation Research, 129*(1), 1–47.

Greco, S., Matarazzo, B., & Slowinski, R. (2002). Rough sets methodology for sorting problems in presence of multiple attributes and criteria. *European Journal of Operational Research, 138*(2), 247–259.

Greco, S., Matarazzo, B., Slowinski, R., & Stefanowski, J. (2001). An algorithm for induction of decision rules consistent with dominance principle. In W. Ziarko & Y. Yao (Eds.), *Rough sets and current trends in computing. LNAI 2005* (pp. 304–313). Berlin: Springer-Verlag.

Gronroos, C. (1996). Relationship marketing: Strategic and tactical implications. *Management Decision, 34*(3), 5–14.

Gustafsson, A., Johnson, M. D., & Roos, I. (2005). The effects of customer satisfaction, relationship commitment dimensions, and triggers on customer retention. *Journal of Marketing, 69*(4), 210–218.

Hawkes, V. A. (2000). *The heart of the matter: the challenge of customer lifetime value*. CRM Forum Resources. pp. 1–10.

Kim, Y. S. (2006). Toward a successful CRM: variable selection, sampling, and ensemble. *Decision Support Systems, 41*(2), 542–553.

Kim, S. Y., Jung, T. S., Suh, E. H., & Hwang, H. S. (2006). Customer segmentation and strategy development based on customer lifetime value: A case study. *Expert Systems with Applications, 31*(1), 101–107.

Mithas, S., Krishnan, M. S., & Fornell, C. (2005). Why do customer relationship management applications affect customer satisfaction? *Journal of Marketing, 69*(4), 201–208.

Ozgener, S., & Iraz, R. (2006). Customer relationship management in small-medium enterprises: The case of Turkish tourism industry. *Tourism Management, 27*(6), 1356–1363.

Park, J. W. (2007). Passenger perceptions of service quality: Korean and Australian case studies. *Journal of Air Transport Management, 13*(4), 238–242.

Pawlak, Z. (1982). Rough sets. *International Journal of Computer and Information Science, 11*(5), 341–356.

Pawlak, Z. (1997). Rough sets. In T. Y. Lin & N. Cercone (Eds.), *Rough sets and data mining: analysis for imprecise data*. Norwell, MA: Kluwer Academic Publishers.

Payne, A., & Frow, P. (2004). The role of multichannel integration in customer relationship management. *Industrial Marketing Management, 33*(6), 527–538.

Roh, T. H., Ahn, C. K., & Han, I. (2005). The priority factor model for customer relationship management system success. *Expert Systems with Applications, 28*(4), 641–654.

Ryals, L., & Knox, S. (2001). Cross-functional issues in the implementation of relationship marketing through customer relationship management. *European Management Journal, 19*(5), 534–542.

Shyng, J. Y., Wang, F. K., Tzeng, G. H., & Wu, K. S. (2007). Rough set theory in analyzing the attributes of combination values for the insurance market. *Expert Systems with Applications, 32*(1), 56–64.

Slowinski, R. (2006). The international summer school on MCDM 2006. Class note. Kainan University, Taiwan. Software available on line: <http://idss.cs.put.poznan.pl/site/software.html>.

Slowinski, R., & Zopounidis, C. (1995). Application of the rough set approach to evaluation of bankruptcy risk. *International Journal of Intelligent System in Accounting, Finance and Management, 4*(1), 27–41.

Wang, Y. F. (2003). Mining stock price using fuzzy rough set system. *Expert System with Application, 24*(1), 13–23.

Wong, J. C., & Chung, Y. S. (2007). Rough set approach for accident chains exploration. *Accident Analysis and Prevention, 39*(3), 629–637.

Zhu, D., Premkumar, G., Zhang, X., & Chu, C. H. (2001). Data mining for network intrusion detection: A comparison of alternative methods. *Decision Science Journal, 32*(4), 635–660.