Springer

Author Query Form

Journal: APIN **Article ID:** APIN465	**Please send your responses together with your list of corrections via web (preferred), or send the completed form and your marked proof to:** Akademijos 4, LT-08412 Vilnius, Lithuania fax: +370 5 2784 091 e-mail: vtexspr-corrections@vtex.lt

Dear Author,

During the preparation of your manuscript for typesetting, some questions have arisen. These are listed below.

Queries and/or remarks

Location in article (page/line)	Query / remark	Response
	IMPORTANT!!! Please note that your MS Word manuscript was converted to LaTeX in the appropriate Springer layout and style. However, due to the numerous ambiguous places in your manuscript, the final proof might have some flaws. Kindly please check your proof carefully in this regard (especially p. 2, lines 281-324) Authors' biographies and photos are required in this journal, please supply. The article cannot be published without biographies and photos. If for some reason you do not want to have them included, please clearly indicate this together with your corrections to the article.	

Many thanks for your assistance

Metadata of the article that will be visualized in Online First

Please note: Images will appear in color online but will be printed in black and white.

Journal Name	Applied Intelligence	
Article Title	A rough set-based association rule approach implemented on exploring beverages product spectrum	
Copyright holder	Springer Science+Business Media New York This will be the copyright line in the final PDF.	
Corresponding Author	Family name	Liao
	Particle	
	Given Name	Shu-Hsien
	Suffix	
	Division	
	Organization	Tamkang University
	Address	New Taipei City, Taiwan, Republic of China
	E-mail	michael@mail.tku.edu.tw
Schedule	Received	
	Revised	
	Accepted	
Abstract	When items are classified according to whether they have more or less of a characteristic, the scale used is referred to as an ordinal scale. The main characteristic of the ordinal scale is that the categories have a logical or ordered relationship to each other. Thus, the ordinal scale data processing is very common in marketing, satisfaction and attitudinal research. This study proposes a new data mining method, using a rough set-based association rule, to analyze ordinal scale data, which has the ability to handle uncertainty in the data classification/sorting process. The induction of rough-set rules is presented as method of dealing with data uncertainty, while creating predictive if—then rules that generalize data values, for the beverage market in Taiwan. Empirical evaluation reveals that the proposed Rough Set Associational Rule (RSAR), combined with rough set theory, is superior to existing methods of data classification and can more effectively address the problems associated with ordinal scale data, for exploration of a beverage product spectrum.	
Keywords	Data mining – Rough set – Association rule – Rough set association rule – Ordinal scale data processing – Product spectrum	
Footnotes		

Appl Intell
DOI 10.1007/s10489-013-0465-1

A rough set-based association rule approach implemented on exploring beverages product spectrum

Shu-Hsien Liao

Abstract When items are classified according to whether they have more or less of a characteristic, the scale used is referred to as an ordinal scale. The main characteristic of the ordinal scale is that the categories have a logical or ordered relationship to each other. Thus, the ordinal scale data processing is very common in marketing, satisfaction and attitudinal research. This study proposes a new data mining method, using a rough set-based association rule, to analyze ordinal scale data, which has the ability to handle uncertainty in the data classification/sorting process. The induction of rough-set rules is presented as method of dealing with data uncertainty, while creating predictive if—then rules that generalize data values, for the beverage market in Taiwan. Empirical evaluation reveals that the proposed Rough Set Associational Rule (RSAR), combined with rough set theory, is superior to existing methods of data classification and can more effectively address the problems associated with ordinal scale data, for exploration of a beverage product spectrum.

Keywords Data mining · Rough set · Association rule · Rough set association rule · Ordinal scale data processing · Product spectrum

1 Introduction

When items are classified according to whether they have more or less of a characteristic, the scale used is referred to as an ordinal scale. The main characteristic of the ordinal scale is that the categories have a logical or ordered relationship to each other. These types of scale permit the measurement of degrees of difference, but not the specific amount of difference, such as market segmentation. Thus the ordinal scale data processing is very common in marketing, satisfaction and attitudinal research. Any questions that ask the respondent to rate something are using ordinal scales. Likert scales are commonly used in attitudinal measurements. This type of scale uses a five-point scale ranging from strongly agree, agree, neither agree nor disagree, disagree, strongly disagree to rate people's attitudes. Although some researchers treat them as an *interval scale*, however we do not really know that the distances between answer alternatives are equal. Hence only the *mode* and *median* can be calculated, but not the *mean*. The *range* and percentile ranking can also be calculated. Ordinal measurements describe order, but not relative size or degree of difference between the items measured. In this scale type, the numbers assigned to objects or events represent the rank order (1st, 2nd, 3rd, etc.) of the entities assessed. In mathematical order theory, an ordinal scale defines a total preorder of objects (in essence, a way of sorting all the objects, in which some may be tied). The scale values themselves (such as labels like "great", "good", and "bad"; 1st, 2nd, and 3rd) have a total order, where they may be sorted into a single line with no ambiguities. If numbers are used to define the scale, they remain correct even if they are transformed by any monotonically increasing function. This property is known as the order isomorphism [10].

In decision making, many classical representations of preferences are cardinal (typically expected utility, or more elaborated models as Choquet expected utility based on monotone measures). They deal with utility functions which are real-valued, and use standard operations of arithmetic such as addition and multiplication [57]. However, it is nei-

S.-H. Liao (✉)
Tamkang University, New Taipei City, Taiwan, Republic of China
e-mail: michael@mail.tku.edu.tw

ther always easy nor desirable to deal with cardinal utility functions. A first noticing is that ordinal information is easier to get than cardinal one, and moreover, there are many situations where only order is relevant, cardinals being merely used by tradition and convenience. More fundamentally, the dominant viewpoint in economics, introduced by Hicks and Allen [27], is that utility is related to observable choices (revealed preferences), and any introspective judgments on intensity of preference should be discarded since meaningless. As a consequence, if the only purpose of utility is to explain choices, then utility is ordinal in nature [1]. In artificial intelligence, ordinal scale data processing on ranking, preference, ordinal measurement, classification, and categorization are implemented on different problem domains, such as: decision model [4, 7, 20, 30, 55], finite ordinals category on chronology [50], multi-criteria development [23, 47], measurement scale development [12, 54], medical study [9] and evaluation [37].

On the other hand, in physics, a spectrum is a series of colored bands of light, diffracted and arranged in the order of their respective wave lengths, produced by the passage of white light through a prism, or other diffracting medium. A spectrum may include many smaller spectrums; for example, the electromagnetic radiation spectrum includes the light spectrum, radio spectrum, infrared spectrum, etc. Beyond physics, a spectrum is a condition that is not limited to a specific set of scales or values, but can vary infinitely, within a continuum or sequence. Since the word saw its first scientific use within the field of optics, to describe the rainbow of the colors of visible light, when separated by a prism, it has been applied in many other fields. Thus, one might talk about a spectrum of political opinion, or the spectrum of activity of a drug, autism spectrum, or specific market segmentation. In these cases, values within the spectrum are not necessarily discrete numbers, as in optics: exact values within this type of spectrum are not precisely quantifiable. Such use implies a broad range of conditions, or behaviors, grouped together and studied under a single title, for ease of discussion. In most modern usages of the word, "spectrum", there is a unifying theme, between extremes at either end, the ordinal events set out below. Accordingly, to obtain a spectrum, the measured function must be transformed into independent scales/variables, with frequencies and the dependent variable must be reduced to the regions, over which the independent variable extends [39].

Consumers prefer certain products, so there is an associated decision-making spectrum. An effective visualization tool, especially for stakeholders, or managers, is a brand/product spectrum diagram, which highlights where the company's brands and products are situated, compared to other competitors. Some businesses have difficulty in understanding their brand attributes and how their products fit into the retail landscape. Often, when questioned, companies espouse a wish to fulfill all promises to all people. However, this approach is often limiting, as a strategy, as it is lacking in targeted vision and segmentation. Therefore, it must be asked whether a business can better understand its consumers, by realizing its own position in the industry, with respect to its specific product segmentation. However, this is easier said than done, since customers' opinions are known only to customers. The information is available, but difficult to access, and without an effective method there is little hope of exploring the full volume of data that might be collected. Thus, the effective processing and use of this data is increasingly important [51, 56].

Data mining is the process of discovering significant knowledge, such as patterns, associations, changes, anomalies and significant structures, from the large amounts of data stored in databases, data warehouses, or other information repositories [39]. Therefore, knowledge of customers, extracted through data mining, can be combined with customer profiles, purchased preferences, records of purchased products and marketing knowledge, from research. This knowledge then provides an understanding of consumers, as well as the product spectrum in a market. Association rules are a data mining method. Previous studies in mining association rules have had two deficiencies. Firstly, the discovery of rules from ordinal data has been ignored. Secondly, the discovery of rules from imprecise data has also been ignored [14]. Corporations, ranging from Coca-Cola, Nestle and McDonald's to Disney and Sony, have invested millions of dollars in developing their corporate image. However, the biggest threats to brand equity are not likely to be trademark or patent infringements, but rather the firm's own actions, or those of its myriad of agents, joint venture/alliance partners, suppliers and subsidiaries [21]. In commerce, businesses use branding to differentiate their products and services, or offerings from those of their competitors [6, 34]. The brand incorporates a set of product or service features that are associated with that particular brand name and attribute [11]. Data classification is used to reduce the large number of conditional attributes, based on the value of the decisional attribute, as well as to extract the key characteristics of certain groups from a wide spectrum of customer attributes [17]. An effective visualization tool, especially for stakeholders or managers, is a brand spectrum diagram, which highlights where the company's brands and products are situated, compared to other competitors. Some businesses have difficulty in understanding their brand attributes and how their products fit into the retail landscape. Often, when questioned, companies espouse a wish to fulfill all promises to all people. However, this approach is often limiting, as a strategy, lacking in targeted vision and segmentation [39].

Accordingly, this study investigates the concept of the product spectrum, by the analysis of algorithms, to sort consumer product preferences and then provide the appropriate decisions. The empirical evaluation reveals that the proposed Rough Set Associational Rule (RSAR), combined with the rough set theory, is superior to existing methods, for ordinal scale data classification, and can more effectively address the marketing issue associated with the investigation of product spectrum on the beverage market in Taiwan. The rest of this paper is organized as follows. Section 2 reviews literature relevant to this research. Section 3 considers the ordinal scale data. Section 4 uses a rough set method to generate associational rules. Computational experiments and conclusion are presented in Sects. 5 and 6.

2 Research background

2.1 Literature review of the development of RST

Rough set theory (RST) was originally proposed by Pawlak, in the 1980's, as a mathematical approach to aid decision making in the presence of uncertainty. It can be used not only as the basis of formal reasoning, with uncertain information, machine learning, knowledge extraction and demand forecasting [31, 44, 65], but also as for a tool for data analysis and autonomous decision-making, and has been used to extract knowledge from datasets [16, 44]. It classifies imprecise, uncertain, or incomplete information, expressed in terms of data acquired from experience [5, 29, 33]. The Pawlak rough set model provides a mathematical tool for the determination of data dependencies and the reduction of the number of features contained in a dataset, using purely structural methods. RST is a theory for the study of intelligent systems, which are characterized by inexact, uncertain, or vague information. In less than two decades, rough sets theory has rapidly established itself in many real-life applications [31]. Presently, rough set theory is used in many fields, such as learning, intelligent systems, inductive reasoning, pattern recognition, image processing, signal analysis, knowledge discovery, decision analysis and environment quality [24, 42, 44, 52, 53, 59]. It has become a key topic in the research area of information science [31, 62]. An information system is a quadruple $S = \{U, A, V, f\}$, where U is a finite set of objects, called the universe, A is a finite set of attributes, V is a domain of attribute a and $f : U \times A \to V$ is called an information function, such that $f(x, a)$ [40]. Any union of elementary sets is called a crisp set and other sets are referred to as rough sets. Ziarko [64] describes the technique as non-statistical and notes that it has been developed with full mathematical rigor, within the realm of logic and set theory. In one sense, this is strength, given that there are no explicit distributional assumptions and no requirements for selection of functional forms. Rough set theory allows easy acquisition knowledge from data, even when the operator has limited prior knowledge. Additionally, the model has the ability to reduce superfluous variables, is easily commanded with 'IF THEN' statements and can be easily modified [38]. Therefore, rough sets can be considered as uncertain or imprecise as the following [15, 25].

An attribute a is a mapping $a : U \to Va$ where U is a non-empty finite set of objects (called the universe) and Va is the value set of a. An information system is a pair $T = (U, A)$ of the universe U and a non-empty finite set A of attributes. Let B be a subset of A. The B-indiscernibility relation is defined by an equivalence relation I_B on U such that $I_B = \{(x, y) \in U_2 \mid \forall a \in B \cdot a(x) = a(y)\}$. The equivalence class of I_B for each object $x(\in U)$ is denoted by $[x]B$. Let X be a subset of U. We define the lower and upper approximations of X by $B(X) = \{x \in U | [x]B \subseteq X\}$ and $B(X) = \{x \in U \mid [x]B \cap X \neq \varnothing\}$. A subset B of A is a reduct of T if $I_B = I_A$ and there is no subset B of B with $I_B = I_A$ (i.e., B is a minimal subset of the condition attributes without losing discernibility). A decision table is an information system $T = (U, A \cup \{d\})$ such that each $a \in A$ is a condition attribute and $d \notin A$ is a decision attribute. Let V_d be the value set $\{d_1, \ldots, d_u\}$ of the decision attribute d. For each value $d_i \in V_d$, we obtain a decision class $U_i = \{x \in U \mid d(x) = d_i\}$ where $U = U_1 \cup \cdots \cup U \mid V_d|$ and for every $x, y \in U_i$, $d(x) = d(y)$. The B-positive region of d is defined by $P_B(d) = B(U_1) \cup \cdots \cup B(U \mid V_d|)$. A subset B of A is a relative reduct of T if $P_B(d) = P_A(d)$ and there is no subset B of B with $P_B(d) = P_A(d)$. We define a formula $(a_1 = v_1) \wedge \cdots \wedge (a_n = v_n)$ in T (denoting the condition of a rule) where $a_j \in A$ and $v_j \in Va_j$ $(1 \le j \le n)$. The semantics of the formula in T is defined by $[\![(a_1 = v_1) \wedge \cdots \wedge (a_n = v_n)]\!]T = \{x \in U | a_1(x) = v_1, \ldots, a_n(x) = v_n\}$. Let ϕ be a formula $(a_1 = v_1) \wedge \cdots \wedge (a_n = v_n)$ in T. A decision rule for T is of the form $\phi \to (d = d_i)$, and it is true if $[\![\phi]\!]T \subseteq [\![(d = d_i)]\!]T (= U_i)$. The accuracy and coverage of a decision rule r of the form $\phi \to (d = d_i)$ are respectively defined as follows:

$$accuracy(T', r, U_i) = \frac{|U_i \cap [\![\phi]\!]_{T'}|}{|[\![\phi]\!]_{T'}|}$$

$$accuracy(T', r, U_i) = \frac{|U_i \cap [\![\phi]\!]_{T'}|}{|U_i|}$$

In the evaluations, $|U_i|$ is the number of objects in a decision class U_i and $[\![\phi]\!]_{T'}|$ is the number of objects in the universe $U = U_1 \cup \cdots \cup U|V_d|$ that satisfy condition ϕ of rule r. Therefore, $|U_i \cap [\![\phi]\!]_{T'}|$ is the number of objects satisfying the condition ϕ restricted to a decision class U_i.

2.2 Previous research of RST functionalities

Kryszkiewicz [35, 36] assigned a null value, to replace all the incomplete attribute values. The null value represents all of the possible values attainable by the attribute with an incomplete value. Felix et al. [18, 19] solved the incomplete information problem by introducing rough discernibility relations, i.e. surely discernible and possibly indiscernible. These relations were used for the derivation of rules, to replace the original indiscernibility relation, in an incomplete information system. Some studies [13, 26, 31–33, 43, 49, 60, 61] used a hybrid approach to deal with incomplete data.

Walczak and Massart [58] mentioned that the "application of RST to qualitative attributes is straightforward. For nominal attributes, RST offers evident advantages when compared with other classifiers". Previous studies of the application of rough sets in intelligent systems, focused on classification accuracy and on preserving the information, or the order generated by the ordinal decision classes. Huang et al. [28] used a RST approach for intelligent systems and their results showed that their method could reduce the number of conditional attributes used in motherboard EMI fault diagnosis and maintain acceptable classification accuracy. The theory has a strong mathematical foundation and is well suited to deal effectively with various decision problems. It can be employed to extract concepts, or decision rules, from a given set of data, and has been used successfully in many application domains [19, 22]. In another article, John W.T. Lee et al. [29] mentioned that, in the article "rough set theory has been successfully applied in selecting attributes to improve the effectiveness in deriving decision trees/rules for decisions and classification problems. When decisions involve ordinal classes, the rough set reduction process should try to preserve the order relation generated by the decision classes". They proposed a new way of evaluating and determining reducts, involving ordinal decision classes, which focused on the order generated by the ordinal decision classes. Zhao et al. [63] present a hormone based nearest neighbor classification algorithm for data stream classification, in which the classifier is updated every time a new record arrives. The records could be seen as locations in the feature space, and each location can accommodate only one endocrine cell.

2.3 Association rules

Associations in complex data objects, such as data items, occur when one set of attributes is likely to co-occur with another set. The prototypical application is the analysis of supermarket transactions where associations like '68 % of all customers who buy fish also buy white wine' may be found in a transaction database. For knowledge discovery—data mining—in databases, an association is a rule to be mined from databases which infers an attribute set from another. As stated by Agrawal et al. discovering association rules is an important data mining problem, and there has been considerable research in using association rules in the field of data mining problems [2]. The association rules algorithm is used mainly to determine the relationships between items or features that occur synchronously in databases. For instance, during a trip to the shopping center, if the people who buy item X also buy item Y as well, there exists a relationship between item X and item Y. Such information is useful for decision makers. Therefore, the main purpose of implementing the association rules algorithm is to find synchronous relationships by analyzing random data and to use these relationships as a reference for decision-making. The association rules are defined as follows [46].

Make $I = \{i_1, i_2, \ldots, i_m\}$ the item set, in which each item represents a specific literal. D stands for a set of transactions in a database in which each transaction T represents an item set such that $T \subseteq I$. That is, each item set T is a non-empty sub-item set of I. The *association rules* are an implication of the form $X \to Y$, where $X \subset I, Y \subset I$ and $X \cap Y = \Phi$. The rule $X \to Y$ holds in the transaction set D according to two measurement standards—*support* and *confidence*. Support (denoted as_$Sup(X, D)$) represents the rate of transactions in D containing the item set X. *Support* is used to evaluate the statistical importance of D, and the higher its value, the more important the transaction set D is. Therefore, the rule $X \to Y$ which has *support* $Sup(X \cup Y, D)$ represents the rate of transactions in D containing $X \cup Y$. Each rule $X \to Y$ also has another measuring standard called confidence (denoted as $Conf(X \to Y)$), representing the rate of transactions in D that contain X and also Y. That is, $Conf(X \to Y) = Sup(X \cap Y)/Sup(X, D)$.

In this case, $Conf(X \to Y)$ denotes that if the transaction includes X, the chance that transaction also contains Y is relatively high. The measure of confidence is then used to evaluate the level of confidence about the association rules $X \to Y$. Given a set of transactions D, the problem of mining association rules is used to generate all transaction rules that have certain user-specified minimum support (called Min sup) and confidence (called Min $conf$). According to Agrawal et al. the problem of mining association rules can be broken down into two steps. The first step is to detect a large item set whose support is greater than Min sup and the second step is to generate association rules, using the large item set. Such rules must satisfy the following two conditions [3]:)

1. $Sup(X \cup Y, D) \geq \text{Min sup}$
2. $Conf(X \to Y) \geq \text{Min}\,conf$

To illustrate the concepts, we use a small example from the supermarket domain. The set of items is $I = \{milk, bread, butter, beer\}$ and a small database containing the items (1 codes presence and 0 absence of an item in a transaction) is shown in the table to the right. An example rule for the supermarket could be $\{butter, bread\} \to \{milk\}$ meaning that if butter and bread are bought, customers also buy milk.

In addition, to explore association rules, many researchers use the Apriori algorithm [2]. In order to reduce the possible biases incurred when using these measurement standards, the simplest way to judge the standard is to use the *lift* judgment. *Lift* is defined as: $Lift = Confidence(X \to Y)/Sup(Y)$.

2.4 Rough set association rules

The primary difficulty associated with the maximal association approach is that the generation of frequent maximal set is based on an underlying assumption—a taxonomy existing for the document collections. However, this assumption may be only feasible for collections of labeled documents with keywords which are mainly for training text classifiers and very expensive to construct, therefore limiting the general applicability of this approach. In addition, Bi et al. investigate the applicability of Rough Set theory to detecting maximal associations [8]. The work reported on some other papers shows that by using Rough Set, rules discovered are similar to maximal association rules, and the rough set approach is much simpler than the maximal association method in discovering association rules for knowledge discovery and reasoning on different data format/scale problem [45, 55].

In view of the prior research, this research does not discuss the rules of order, but rather focuses on decision-makers for consumer product preferences. Thus, this study proposes a new data mining approach, which analyzes ordinal scale data and has the ability to handle uncertainty in the data classification/sorting process. In the domain of knowledge extraction, rough set theory offers the benefits of efficiency, understandability and results that can be interpreted directly. This paper proposes the induction of rough-set rules, to deal with data uncertainty, while creating predictive if–then rules that generalize data values for the beverage industry.

3 Ordinal scale data processing

Traditional association rules ignore the discovery of rules from ordinal data. This study combines association rules with Rough sets, to create an application for ordinal scale data. The processing of ordinal scale data is described in Table 1.

Table 1 Example database with 4 items and 5 transactions of association rules

Transaction ID	Milk	Bread	Butter	Beer
1	1	1	0	0
2	0	0	1	0
3	0	0	0	1
4	1	1	1	0
5	0	1	0	0

Table 2 Information system: ordinal scale data sets

U	A							
	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
x_1	1	5	7	4	3	2	8	6
x_2	1	6	8	2	4	5	3	7
x_3	1	7	2	4	6	5	3	8
x_4	1	2	3	5	7	6	4	8
x_5	1	3	6	2	4	5	8	7

Definition 1 Transform the questionnaire answers into information system $IS = (U, A)$, where $U = \{x_1, x_2, \ldots, x_i\}$ is a finite set of objects and $i = 1, \ldots, n$, $A = \{a_1, a_2, \ldots, a_j\}$ is a finite set of general attributes/criteria and $j = 1, \ldots, m$. $f_a = U \times A \to V_a$ called the information function, V_a is the domain of the attribute/criterion a, and f_a is a ordinal function set such that $f(x, a) \in V_a$ for each $x_i \in U$.

Example Table 2 shows the ranking of non-alcoholic beverages, from the first to eighth, by x_1, named Tea, Packaged-waters, Sports, Juice, Soda, Others, Coffee and Energy.

Then:

$$f_{a_1} = \{1\}, \qquad f_{a_2} = \{2, 3, 5, 6, 7\},$$
$$f_{a_3} = \{2, 3, 6, 7, 8\}, \qquad f_{a_4} = \{2, 4, 5\}$$
$$f_{a_5} = \{3, 4, 6, 7\}, \qquad f_{a_6} = \{2, 5, 6\},$$
$$f_{a_7} = \{3, 4, 8\}, \qquad f_{a_8} = \{6, 7, 8\}$$
$$V_{a_1}^{x_1} = 1, \quad V_{a_2}^{x_1} = 5, \quad V_{a_3}^{x_1} = 7, \quad V_{a_4}^{x_1} = 4$$
$$V_{a_5}^{x_1} = 3, \quad V_{a_6}^{x_1} = 2, \quad V_{a_7}^{x_1} = 8, \quad V_{a_8}^{x_1} = 6$$

Definition 2 According to specific universe of discourse classification, a similarity relation of the general attributes $a \in A$, denoted by $\frac{U}{A}$. All of the similarity relation, denoted by $R(a_j)$.

$$\frac{U}{A} = \{[x_i]_A | x_i \in U\}$$

« APIN 10489 layout: Large v.1.3.2 file: apin465.tex (Judita) class: spr-twocol-v1.4 v.2013/04/17 Prn:2013/07/19; 12:26 p. 6/15»
« doctopic: OriginalPaper numbering style: ContentOnly reference style: basic»

Table 3 The core attribute values of the ordinal scale data for non-alcoholic beverages

$R(a_j)$	f_{a_5}	f_{a_6}	D_a
$\{x_1\}$	3	2	D_a^+
$\{x_2, x_5\}$	4	5	D_a^-
$\{x_3\}$	6	5	D_a^+
$\{x_4\}$	7	6	D_a^+
$R(a_j)$	f_{a_5}	f_{a_6}	D_a

Example

$$R(a_3) = \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5\}\},$$
$$R(a_5) = \{\{x_1\}, \{x_2, x_5\}, \{x_3\}, \{x_4\}\}$$
$$R(a_6) = \{\{x_1\}, \{x_2, x_3, x_5\}, \{x_4\}\},$$
$$R(a_7) = \{\{x_1, x_5\}, \{x_2, x_3\}, \{x_4\}\}$$

Definition 3 The Information system is an ordinal scale data, therefore between the two attributes will have the ordinal response, where D_a is the pair wise comparison results of ordinal scale data, which are defined as follows,

$$D_a^+ = \left\{x_i \middle| \frac{U}{a}, V_{a_1} > V_{a_j}\right\}, \qquad D_a^- = \left\{x_i \middle| \frac{U}{a}, V_{a_i} < V_{a_j}\right\}$$

Then, using the concept of similarity relation in rough set theory foundation, and finding the value of ordinal scale data between a_i and a_j, where $ind(B)$ is the core attribute value of ordinal scale data in the first step, and B is the subset of A.

$$ind(B) = [f_a]_{B\subseteq A} = \bigcap_{B\subseteq A}\left[\frac{U}{a}\right]$$

Example According to the similarity relation and the fact that $R(a_5) = \{\{x_1\}, \{x_2, x_5\}, \{x_3\}, \{x_4\}\}$ and $R(a_6) = \{\{x_1\}, \{x_2, x_3, x_5\}, \{x_4\}\}$ both belong to the same fundamental set, the ordinal function set is $f_{a_5} = \{3, 4, 6, 7\}$ and $f_{a_6} = \{2, 5, 6\}$. Therefore, a_5 and a_6 are both core attribute values of the ordinal scale data for non-alcoholic beverages and for customer x_1, x_3 and x_4, a_5 always places after a_6, denoted by D_a^+. The pair wise comparison of a_5 and a_6, as shown in Table 3.

$$ind(B) = [\mathrm{a}_5, \mathrm{a}_6]$$

4 Rough set method for the generation of associational rules

Definition 4 As a first step, this study identifies the core attribute values of ordinal scale data. In this step, the object generates the rough associational rule. The consideration of other attributes and the core attributes of ordinal scale data as the highest decision-making attributes is used to establish the decision table and to generate rules, as shown in Table 4.

Table 4 Decision-making table showing drinking habits for "non-alcoholic beverages"

U	General attributes, Q				Decision attributes	
	g_1	g_2	g_3	g_4	Product ranking	
x_1	g_{1_1}	g_{2_1}	g_{3_1}	g_{4_1}	3	a_5
x_2	g_{1_1}	g_{2_2}	g_{3_1}	g_{4_2}	6	a_5
x_3	g_{1_2}	g_{2_1}	g_{3_2}	g_{4_1}	6	a_5
x_4	g_{1_2}	g_{2_1}	g_{3_2}	g_{4_1}	7	a_5
x_5	g_{1_1}	g_{2_2}	g_{3_1}	g_{4_2}	5	a_5

$DT = (U, Q)$, where $U = \{x_1, x_2, \ldots, x_i\}$ is a finite set of objects and $i = 1, \ldots, n$, Q is usually divided into two parts. $G = \{g_1, g_2, \ldots, g_j\}$ is a finite set of general attributes/criteria and $j = 1, \ldots, m$, $D = \{d_1, d_2, \ldots, d_l\}$ is a set of decision attributes and $k = 1, \ldots, p$. $f_g = U \times G \to V_g$ is called the information function, V_g is the domain of the attribute/criterion, g, and f_g is a total function, such that $f(x, g) \in V_g$, for each $g \in Q$; $x \in U$. $f_d = U \times D \to V_d$ is called the sorting decision-making information function, V_d is the domain of the decision attributes/criterion, d, and f_d is a total function, such that $f(x, d) \in V_d$, for each $d \in Q$; $x \in U$.

Then:

$$f_{g_1} = \{g_{1_1}, g_{1_2}\}, \qquad f_{g_2} = \{g_{2_1}, g_{2_2}\}$$
$$f_{g_3} = \{g_{3_1}, g_{3_2}\}, \qquad f_{g_4} = \{g_{4_1}, g_{4_2}\}$$

Definition 5 According to the specific universe of discourse classification, a similarity relation for the general attributes is denoted by $\frac{U}{G}$. All of the similarity relations are denoted by $R(g_t)$ and t is the combination of all the general attributes.

$$R(g_t) = \frac{U}{G} = \{[x_i]_G | x_i \in U\}$$

Example

$$R_1 = \frac{U}{g_1} = \{\{x_1, x_2, x_5\}, \{x_3, x_4\}\}$$

$$R_2 = \frac{U}{g_2} = \{\{x_1, x_3, x_4\}, \{x_2, x_5\}\}$$

$$\vdots$$

Journal ID: 10489, Article ID: 465, Date: 2013-07-19, Proof No: 1, UNCORRECTED PROOF

Table 5 Similarity relations and relational attribute values

ind(*B*)	*R*	Product features g_1	Product information source g_2	Consumer behavior g_3	Channels g_4	Decision attributes *D* (sports)
$\frac{U}{g_1g_2g_3g_4}$	$\{x_1\}$	Price	Seen on shelves	Purchased due to promotions	Convenience stores	Third $d^1_{a_5}=3$
	$\{x_2,x_5\}$	Price	Advertising	Purchased due to promotions	Hypermarkets	Sixth $d^2_{a_5}=6$
						Fifth $d^5_{a_5}=5$
	$\{x_3,x_4\}$	Brand	Seen on shelves	Not purchased due to promotions	Convenience stores	Sixth $d^3_{a_5}=6$
						Seventh $d^4_{a_5}=7$
$\frac{U}{g_2g_4}$	$\{x_2,x_5\}$	Price	Advertising	Purchased due to promotions	Hypermarkets	Fourth $d^2_{a_5}=4$
						Fourth $d^2_{a_5}=4$
	$\{x_1,x_3,x_4\}$	Price	Seen on shelves	Purchased due to promotions		Third $d^1_{a_5}=3$
		Brand		Not purchased due to promotions	Convenience stores	Sixth $d^3_{a_5}=6$
		Brand		Not purchased due to promotions		Seventh $d^4_{a_5}=7$

$$R_5=\frac{U}{g_2g_4}=\bigl\{\{x_1,x_3,x_4\},\{x_2,x_5\}\bigr\}$$

$$\vdots$$

$$R_t=\frac{U}{g_1g_2g_3g_4}=\bigl\{\{x_1\},\{x_2,x_5\},\{x_3,x_4\}\bigr\}$$

Definition 6 By the similarity relation, and determination of the reduct and core, the attribute, g, of G and the set G, which was ignored, has no effect, so g is an unnecessary attribute and can be reducted. $R\subseteq G$ and $\forall_g\in R$. A similarity relation for the general attributes of the decision table is denoted by $ind(G)$. If $ind(G)=ind(G-g_1)$, then g_1 is the reduct attribute and if $ind(G)\neq ind(G-g_1)$, then g_1 is the core attribute.

Example

$$\frac{U}{ind(G)}=\bigl\{\{x_1\},\{x_2,x_5\},\{x_3,x_4\}\bigr\}$$
$$\frac{U}{ind(G-g_1)}=\frac{U}{g_2g_3g_4}=\bigl\{\{x_1\},\{x_2,x_5\},\{x_3,x_4\}\bigr\}$$
$$=\frac{U}{ind(G)}=\frac{U}{g_1g_2g_3g_4}$$
$$\frac{U}{ind(G-g_1g_3)}=\frac{U}{g_2g_4}=\bigl\{\{x_1,x_3,x_4\},\{x_2,x_5\}\bigr\}$$
$$\neq\frac{U}{ind(G)}=\frac{U}{g_1g_2g_3g_4}$$

When considering g_1, alone, g_1 is the reduct attribute, but when considering g_1 and g_3, simultaneously, g_1 and g_3 are the core attributes. A similarity relation and the relational attribute value are shown in Table 5.

Definition 7 The lower approximation, denoted as $\underline{G}(X)$, is defined as the union of all of the elementary sets that are contained in $[x_i]_G$. More formally:

$$\underline{G}(X)=\bigcup\left\{[x_i]_G\in\frac{U}{G}\middle|[x_i]_G\subseteq X\right\}$$

The upper approximation, denoted as $\overline{G}(X)$, is the union of those elementary sets that have a non-empty intersection with $[x_i]_G$. More formally:

$$\overline{G}(X)=\bigcup\left\{[x_i]_G\subseteq\frac{U}{G}\middle|[x_i]_G\cap X\neq\phi\right\}$$

The difference: $Bn_G(X)=\overline{G}(X)-\underline{G}(X)$ is called a boundary of $[x_i]_G$.

Example $\{x_1,x_2,x_4\}$ are the customers of interest, so $\underline{G}(X)=\{x_1\}$, $\overline{G}(X)=\{x_1,x_2,x_3,x_4,x_5\}$ and $Bn_G(X)=\{x_2,x_3,x_4,x_5\}$.

Definition 8 Using the traditional association rule to calculate the value of Support and Confidence, the formula is shown as follows:

$$Sup\bigl(ind(B)\bigr)=\bigl|\bigl\{ind(B)\big|\underline{G}(X)\subseteq\overline{G}(X)\bigr\}\bigr|$$
$$=\left|\frac{ind(B)|\underline{G}(X)}{\overline{G}(X)}\right|$$
$$Conf\bigl(ind(B)\to d_{g_m}\bigr)=\bigl|\bigl\{ind(B)\cap d_{g_m}\big|Sup\bigl(ind(B)\bigr)\bigr\}\bigr|$$
$$=\left|\frac{Sup(ind(B)\cap d_{g_m})}{Sup(ind(B))}\right|$$

Definition 9 Rough set-based association rules.

$$\frac{\{x_1\}}{g_1g_3}:g_{1_1}\cap g_{3_1}\quad\Rightarrow\quad d^1_{d_1}=4$$
$$\frac{\{x_1\}}{g_1g_2g_3g_4}:g_{1_1}\cap g_{2_1}\cap g_{3_1}\cap g_{4_1}\quad\Rightarrow\quad d^1_{d_1}=4$$

The algorithm:

Algorithm-Step 1

Input:

Information System (IS);

Output:

{Core Attributes};

Method:

1. Begin
2. $IS = (U, A)$;
3. $x_i \in U$; /* where $x_1, x_2, \ldots, x_n$ are the objects of set U */
4. $a_1, a_2, \ldots, a_m \in A$; /* where $a_1, a_2, \ldots, a_m$ are the elements of set A */ $a_j \in A$; /* where $a_1, a_2, \ldots, a_m$ are the elements of set A */
5. For each a_m do;
6. compute $R(a_j)$; /* where $R(a_j)$ are the similarity relation in IS as described in Definition 5 */
7. generate D_a; /* where D_a are the result that compute the V_a as condition attributes in $R(a_j)$ as described in Definition 6 */
8. Endfor;
9. Output {Core Attributes};
10. End;

Algorithm-Step 2

Input:

Decision Table (DT);

Output:

{Classification Rules};

Method:

1. Begin
2. $DT = (U, Q)$;
3. $x_i \in U$; /* where $x_1, x_2, \ldots, x_n$ are the objects of set U */
4. $Q = (G, D)$;
5. $g_j \in G$; /* where $g_1, g_2, \ldots, g_m$ are the elements of set G */
6. $d_k \in D$; /* where $d_1, d_2, \ldots, d_p$ are the "core attributes" generated in Step 1 */
7. For each g_j do;
8. compute $R(g_t)$; /* where $R(g_t)$ are the similarity relation in DT as described in Definition 5 */
9. compute $ind(G - g_j)$; /$*$ compute the relative reduct of the elements for element m as described in Definition 6 */
10. generate $ind(B)$; /* where $ind(B)$ are the indiscernibility relation of DT as described in Definition 6 */
11. compute $\underline{\underline{G}}(X)$; /* where $\underline{\underline{G}}(X)$ are the lower-approximation of DT as described in Definition 7 */
12. compute $\overline{\overline{G}}(X)$; /* where $\overline{\overline{G}}(X)$ are the upper-approximation of DT as described in Definition 7 */
13. compute $Bn_G(X)$; /* where $Bn_G(X)$ are the bound of DT as described in Definition 7 */
14. compute $Sup(ind(B))$; /* where $Sup(ind(B))$ are the core attribute support as described in Definition 8 */
15. compute $Conf(ind(B) \rightarrow d_{g_j})$; /* where $Conf(ind(B) \rightarrow d_{g_j})$ are the core confidence as described in Definition 8 */
16. Endfor;
17. Output {Classification Rules};
18. End;

5 Computational experiments

5.1 Ordinal scale data on consumer behavior

Intuitively, consumers who bought beer after buying diapers and those who bought diapers after buying beer, represent groups of decision-maker. Using traditional association rules, it is only possible to see the relationship between diapers and beer, but no deeper understanding of the information is possible. Furthermore, decision-makers want to know the customer's product preferences, in order, for ex-

Table 6 Description of attributes in the primary survey

Attribute name	Attribute value	Attribute name set
Basic information		
Gender	Male; Female	{1,2}
Age	Below 18 years old; 18–25; 26–30; 31–35; 36–40; 41–45; above 51 years old	{1,2,3,4,5,6,7}
Income	Below NT$5,000; BetweenNT$5,001 and NT$15,0000; Between NT$15,001 and NT$20,000; Between NT$20,001 and NT$25,000; Between NT$25,001 and NT$30,000; Between NT$30,001 and NT$35,000; Above NT$35,001	{1,2,3,4,5,6,7}
Consumer behaviors		
Non-alcoholic beverages	Tea; Soda; Coffee; Juice; Sports; Packaged-waters; Energy; Others	{1,2,3,4,5,6,7,8}
Medium	Advertising; Seen on shelves; Internet; Magazine; Newspaper; Broadcasting; Billboard; Belongings	{1,2,3,4,5,6,7,8}
Channel	Hypermarkets; Supermarkets; Convenience Stores	{1,2,3}
Product Features	Price; Brand; Flavor; Quality	{1,2,3,4}
Consumer Behavior	Purchased due to promotions; Not purchased due to promotions	{1,2}

ample, the favorite brand, next favorite brands and overall brand ranking. It can be seen that the sequence of information for decision makers is very important. Therefore, the non-alcoholic beverages sold in the drink market are consolidated and then divided into eight items, which are listed in the questionnaire, for consumers to rank. The questionnaire is shown below:

---

(1) Tea (Oolong Tea, Red Tea, Green Tea, Fruit Tea...)
(2) Soda (Cola, soft drinks, ...)
(3) Cofee (Latte, Mocha, ...)
(4) Juice (Grape juice, Apple juice, ...)
(5) Sports (Shupao, Pocari, ...)
(6) Packaged-waters (Pure water, mineral water, Deep-sea water)
(7) Energy (Comebest, ...)
(8) Others (Milk, Rice Milk, Other vinegar, ...)
Please indicate your preferred product choices, in the following space:
____/____/____/____/____/____/____/____

By means of this open-ended questionnaire, consumers rank each product category. In order to demonstrate the superiority of the proposed approach over traditional association rules, an empirical study was undertaken and is described in this section. A questionnaire, with single and multiple-choice answers, was produced, to determine customer behavior. The questionnaire comprised two parts; the first to collect basic information and the second to determine the consumer behaviors that are involved in the decision process. The results provide the retailer with useful information about the beverage product spectrum, to allow the development of effective marketing strategies.

5.2 Construction of the information table for customer behavior in the retail market

The research sample comprised mainly members of the public who had purchased non-alcoholic beverage products in retail chain stores, within the last three months. One thousand questionnaires were distributed and 772 were returned, of which 172 were disqualified, as incomplete, or invalid. This left a total of 600 valid questionnaires, yielding a valid completion rate of 60 %. The domain values of the personal attributes for the primary survey are shown in Table 6. The profiles are shown in Table 7.

5.3 Results using reducts and core

According to Sai et al. [48], if an ordered information table has one or more reducts, then attributes that are not part of any reduct are dispensable. These dispensable attributes can be removed from the data table, without affecting the ordering of the objects. Table 8 shows that three non-alcoholic beverages are related (No 1). In other words, when these three non-alcoholic beverage products are combined, it is found that consumers prefer Tea, to Juice, and like Juice more than others. The Strength is the total number of consumers in the sample that make such a choice. Using data processing algorithms, the eight non-alcoholic beverages can be classified according to the results of Table 8. Marketing decision-makers can assign the preferences of consumers to a promotion or the development of a new product. For example, the first class of 73 consumers ranked Tea as first, Juice as second and others as fourth. The second class of 65 consumers ranked Tea as second, Juice as third and others as fifth. The first

« APIN 10489 layout: Large v.1.3.2 file: apin465.tex (Judita) class: spr-twocol-v1.4 v.2013/04/17 Prn:2013/07/19; 12:26 p. 10/15»
« doctopic: OriginalPaper numbering style: ContentOnly reference style: basic»

Table 7 Profile of respondents

Distribution	Distribution sample size	Frequency (%)
Gender		
Male	367	61.2
Female	233	38.8
Age		
Below 18 years old	21	3.5
18–25	237	39.5
26–30	174	29.0
31–35	78	13.0
36–40	29	4.8
41–45	43	7.2
Above 51 years old	18	3.0
Income		
Below NT$5,000	127	21.2
Between NT$5,001 and NT$15,0000	90	15.0
Between NT$15,001 and NT$20,000	66	11.0
Between NT$20,001 and NT$25,000	61	10.2
Between NT$25,001 and NT$30,000	74	12.3
Between NT$30,001 and NT$35,000	81	13.5
Above NT$35,001	101	16.8
Non-alcoholic beverages		
Tea	Ranking of non-alcoholic beverages	See the results
Soda		
Coffee		
Juice		
Sports		
Packaged-waters		
Energy		
Others		
Medium (*multiple-choice answers*)		
Advertising	489	81.5
Seen on shelves	105	17.5
Internet; Magazine	47	7.8
Newspaper	30	5.0
Broadcasting	53	8.8
Billboard	122	20.3
Belongings	349	58.2
Channel (*multiple-choice answers*)		
Hypermarkets	202	33.7
Supermarkets	255	42.5
Convenience Stores	561	93.5
Product features (*multiple-choice answers*)		
Price	418	69.7
Brand	175	29.2
Flavor	367	61.2
Quality	183	30.5
Consumer behavior		
Purchased due to promotions	402	67.0
Not purchased due to promotions	198	33.0

Journal ID: 10489, Article ID: 465, Date: 2013-07-19, Proof No: 1, UNCORRECTED PROOF

Table 8 The significance of condition attributes/criteria

No.	$ind(B)$	f_a	D_a	Strength (U/a)
1	{Tea, Juice, Others}	{{2,3,5},{1,2,4},{1,4,7}}	D_a^+	{65,73,52}
2	{Tea, Juice, Soda}	{{1,2,3},{1,3,7},{2,3,6}}	D_a^+	{73,64,52}
3	{Tea, Juice, Others, Sports}	{{1,2,4,6},{2,3,5,7}}	D_a^+	{60,52}
4	{Tea, Juice, Packaged, Energy}	{{1,2,5,8},{2,3,4,8},{1,4,6,8}}	D_a^+	{73,52,52}
5	{Tea, Juice, Others, Energy}	{{1,2,4,8},{2,3,5,8},{1,4,7,8}}	D_a^+	{60,52,52}
6	{Tea, Juice, Packaged, Sports}	{{1,2,5,6},{2,3,4,7}}	D_a^+	{60,52}
⋮	⋮	⋮	⋮	⋮

Table 9 Possible rules for non-alcoholic beverages by rough set association rule

Condition	Preference ranking of non-alcoholic	$Sup(ind(B))$	$Conf(ind(B) \rightarrow d_{g_j})$
(Channels = 3) & (Product Features = 1) & (Medium = 1)	{Tea = 1}	44.44 %	60.00 %

and second categories of consumers like the product category equally, but the preferences for non-alcoholic beverages are different, so marketing decision-makers can change the consumer behavior of the second class, so that the first and second categories of consumer behaviors are the same. This not only enhances the ordering of the product, but also increases the market share for Tea, juice and others.

In this case, the significance of the condition attributes/criteria associated with dispensable attributes can be used to help retail decision-makers understand the spectrum of beverage products.

5.4 Rules using core criteria and personal attributes

A database always contains a lot of attributes that are redundant and not necessary for rule discovery. If these redundant attributes cannot be removed, the time complexity of the rule discovery process increases and the quality of the discovered rules may be much degraded. Decisions whether to delete attributes are very difficult for non-experts and even for experts. Clearly, it is necessary to develop methods for the selection of feature (attribute) subsets. An optimal feature subset should contain all of the indispensable features, because removing any of these features causes inconsistency, in a decision table. The discernibility matrix [5, 62, 64, 65] can be used for *CORE* searching. *CORE* searching searches such a subset of features, each of which uniquely discerns some instances. If *CORE* is not a reduct, some of the dispensable features must be selected and added to it, to make a reduct.

Using the reduced core criteria, shown in Table 8, a set of rule was established. These consider the personal profile attributes, including purchasing medium, channel, product features and consumer behavior.

The calculus of the research process generated by the rough set association rules, and the consumers ranked Tea as first as an example; the interesting target group is those customers who ranked Tea as first, and B is all ranked sets included tea. Thus, according to this calculation process produced 20 sets (all sets in the study are 45), therefore the rough associated support ($Sup(ind(B))$) is 44.44 %. Furthermore, the ranked sets included tea and ranked Tea as first are 12, therefore the rough associated confidence ($Conf(ind(B) \rightarrow d_{g_j})$) is 60 %. The rough set association rule for non-alcoholic beverages ranked Tea as first as an example is shown in Table 9.

In addition, calculated by use of the traditional association rules, thereby creating a rough association rules, under the conditions of the support is greater than 10 %, and confidence greater than 20 %, take the life is greater than 1, and the decision rules generated by the non-alcoholic beverage preferences are 201. A part of rules set of non-alcoholic beverage preferences, as shown in Table 10.

6 Conclusion

This study explores the association between sequences, for non-alcoholic beverages, according to the condition attributes set for non-alcoholic beverages, consumer products channel, advertising media sources, purchaser's consideration of product characteristics and consumer behavior, considered in conjunction with the purchase order. The core criteria of the non-alcoholic beverages product spectrum are shown in Fig. 1.

The study finds that most consumers buy non-alcoholic beverages because of the price and that advertising is the source of most product information. Before, or after the

Table 10 Possible rules for non-alcoholic beverages by traditional association rule

No.	Condition	Preference ranking of non-alcoholic	*Sup.* (%)	*Conf.* (%)	Lift
	(Channels = Convenience Stores)&	{Tea = 1}	23.00	52.17	1.06
1	(Product Features = Price)&	{Tea = 2}	23.00	31.16	1.10
	(Medium = Broadcasting)				
	(Channels = Hypermarkets)&	{Tea = 1}	17.50	53.33	1.09
2	(Product Features = Price)&				
	(Medium = Broadcasting)&				
	(Consumer Behavior = Purchase by promotions)				
	(Channels = Hypermarkets)&	{Juice = 2}	24.00	25.69	1.09
3	(Product Features = Price)&	{Juice = 3}	24.00	38.89	1.13
	(Medium = Broadcasting)				
	(Consumer Behavior = Purchase by promotions)&	{Juice = 3}	15.33	36.96	1.07
4	(Product Features = Convenience Stores)&				
	(Product Features = Price)&				
	(Medium = Broadcasting)				
	(Consumer Behavior = Purchase by promotions)&	{Sports = 6}	40.17	31.95	1.14
5	(Product Features = Price)&	{Sports = 7}	40.17	26.14	1.05
	(Medium = Broadcasting)				
6	(Product Features = Hypermarkets)&	{Sports = 6}	24.00	31.94	1.14
	(Product Features = Price)&				
	(Medium = Broadcasting)				

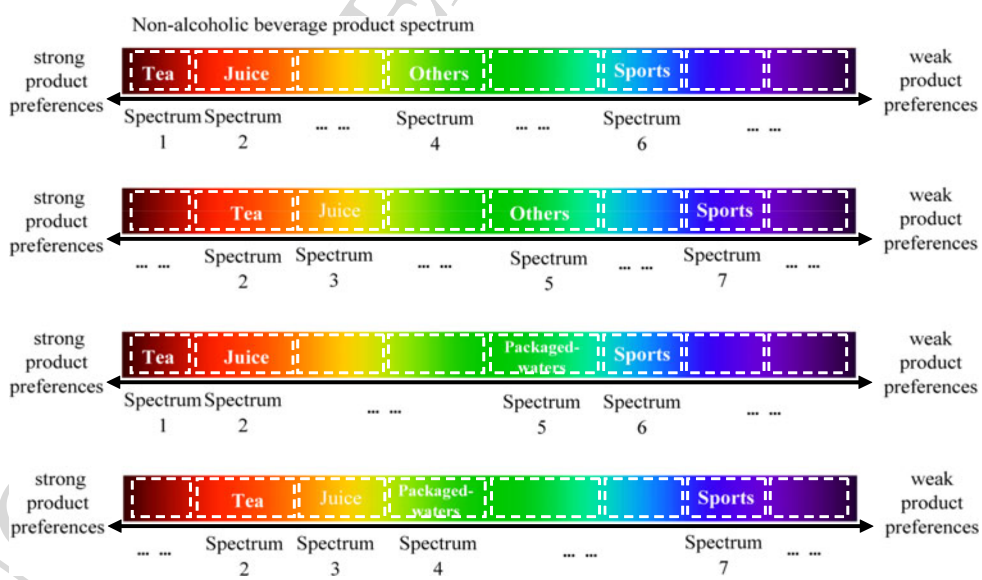

Fig. 1 The core criteria of the non-alcoholic beverages product spectrum

product ranking, consumers buy non-alcoholic beverages in Hypermarkets and convenience stores, but those who buy non-alcoholic beverages in convenience stores are more significantly affected by promotions. The integrated rules for the non-alcoholic beverages product spectrum are shown in Table 11.

Although rough set theory has found uses in a variety of areas, it is still not often applied in the study of customer behavior [41]. Traditional association rules should be fixed, in order to avoid both the retention of only trivial rules that the discarding of interesting rules. In fact, the use of relative comparison, to express preferences, yields better results than absolute comparison. This paper presents a new method for the determination of association rules, which has the ability to handle uncertainty in the classification process and is suitable for ratio scale data. In contrast with other research this study's data processing included that of quantity attribute data and quality attribute data. In the second step, the gen-

Table 11 The integrated rules for the non-alcoholic beverage products spectrum

Product	Medium	Channel		Product features	Consumer behavior
	Advertising	Hypermarkets	Convenience stores	Price	Purchased due to promotions
Strong product spectrum					
Tea	V		V	V	
	V	V		V	V
Juice	V	V		V	
	V		V	V	V
Middle product spectrum					
Others	V		V	V	V
Packaged waters	V	V		V	V
	V		V	V	V
Weak product spectrum					
Sports	V			V	V
	V	V		V	
⋮				⋮	

eration of rough association rules, the decision variable is generated from the core data of the first step, which provides a scientific method of addressing the problem. This study proposes a new data mining method, for ordinal scale data, which has the ability to handle uncertainty in the data classification/sorting process.

The products at the front end of the product spectrum were more popular with consumers; these products were favored by consumers, had large sales volumes and made good profits. The products at the back end of the product spectrum were less favored by consumers and had relatively lower sales volumes and profits. It is suggested that manufacturers could create marketing strategies that move their products toward the front end of the spectrum, in order to increase sales volumes and profits.

Finally, this study suggests that customer market segmentation allows a greater understanding of consumers' demands and preferences. In addition, the characteristics of the product spectrum can be used to determine whether brands are ideal, from the perspective of customers'. The product spectrum analysis diagram can be used to understand the product and to construct marketing strategies that allow greater penetration of the market.

References

1. Abdellaoui M, Barrios C, Wakker P (2002) Reconciling introspective utility with revealed preference; experimental arguments based on prospect theory. Technical report, CREED, Department of Economics, Amsterdam, The Netherlands
2. Agrawal R, Imilienski T, Swami A (1993) Mining association rules between sets of items in large databases. In: Proceedings of the 1993 ACM SIGMOD international conference on management of data, pp 207–216
3. Agrawal R, Shafer J (1996) Parallel mining of association rules. IEEE Trans Knowl Data Eng 8(6):962–969
4. Akter S, Bennett J (2013) Preference uncertainty in stated preference studies: facts and artifacts. Appl Econ 45(15):2107–2115
5. Alisantoso D, Khoo LP, Lee BH, Fok SC (2005) A rough set approach to design concept analysis in a design chain. Int J Adv Manuf Technol 26:427–435
6. Baker M (1996) Marketing—an introductory text. Macmillan Co., London
7. Batyrshin IZ, Rudas IR, Villa LA, Antonio PC (2013) On the monotone sum of basic t-norms in the construction of parametric families of digital conjunctors for fuzzy systems with reconfigurable logic. Knowl-Based Syst 38(1):27–36
8. Bi Y, Anderson T, McClean S (2003) A rough set model with ontologies for discovering maximal association rules in document collections. Knowl-Based Syst 16(5–6):243–251
9. Czerniecki JM, Turner AP, Williams RM, Hakimi KN, Norvell DC (2012) The effect of rehabilitation in a comprehensive inpatient rehabilitation unit on mobility outcome after dysvascular lower extremity amputation. Arch Phys Med Rehabil 93(8):1384–1391
10. Denneberg D, Grabisch M (2004) Measure and integral with purely ordinal scales. J Math Psychol 48(1):15–27
11. Díaz S, García-Lapresta JL, Montes S (2008) Consistent models of transitivity for reciprocal preferences on a finite ordinal scale. Inf Sci 178(13):2832–2848
12. Dolnicar S, Grün B (2007) How constrained a response: a comparison of binary, ordinal and metric answer formats. J Retail Consum Serv 14(2):108–122
13. Castillo PA, Arenas MG, Rico N, Mora AM, García-Sánchez P, Laredo JLJ, Merelo JJ (2012) Determining the significance and relative importance of parameters of a simulated quenching algorithm using statistical tools. Appl Intell 37(2):239–254
14. Chen YL, Weng CH (2008) Mining association rules from imprecise ordinal data. Fuzzy Sets Syst 159(4):460–474

15. Darshit Parmar TW, Blackhurst J (2007) MMR: an algorithm for clustering categorical data using rough set theory. Data Knowl Eng 63(6):879–893
16. Dibb S, Simkin L, Pride W, Ferrell OC (1997) In: Marketing concepts and strategies. Houghton Mifflin, New York
17. Fan B, Zhang P (2009) Spatially enabled customer segmentation using a data classification method with uncertain predicates. Decis Support Syst 47:343–350
18. Felix R, Ushio T (1999) Rules induction from inconsistent and incomplete data using rough sets. In: Proceedings of IEEE international conference on systems, man, and cybernetics, pp 154–158
19. Felix R, Ushio T (1999) Rough sets based machine learning using a binary discernibility matrix. In: Proceedings of the second international conference on intelligent processing and manufacturing of materials (IPMM '99). IEEE Operation Center Press, Hawaii, pp 299–305
20. Godo L, Zapico A (2001) On the possibilistic-based decision model: characterization of preference relations under partial inconsistency. Appl Intell 14(3):319–333
21. Griffith A Jr., Ryans KJ (1997) Organizing global communications to minimize private spill-over damage to brand equity. J World Bus 38:35–45
22. Griffin G, Chen Z (1998) Rough set extension of Tcl for data mining. Knowl-Based Syst 11(3–4):249–253
23. Grsbisch M (2008) How to score alternatives when criteria are scored on an ordinal scale. J Multi-Criteria Decis Anal 15(1):31–44
24. Guo QL, Zhang M (2009) Implement web learning environment based on data mining. Knowl-Based Syst 22(6):439–442
25. Han SW, Kim JY (2008) Rough set-based decision tree using a core attribute. Int J Inf Technol Decis Mak 7(2):275–290
26. Hassan YF (2011) Rough sets for adapting wavelet neural networks as a new classifier system. Appl Intell 35(2):260–268
27. Hicks J, Allen A (1934) A reconsideration of the theory of value: I, II. Economica 1(52–75):196–219
28. Huang CL, Li TS, Peng TK (2005) A hybrid approach of rough set theory and genetic algorithm for fault diagnosis. Int J Adv Manuf Technol 27:119–127
29. Lee JWT, Yeung DS, Tsang CC (2006) Rough sets and ordinal reducts. Soft Comput 10(1):27–33
30. Jullien-Ramasso S, Mauris G, Valet L, Bolon P (2012) A decision support system for animated film selection based on a multi-criteria aggregation of referees' ordinal preferences. Expert Syst Appl 39(4):4250–4257
31. Khoo LP, Tor SB, Zhai LY (1999) A rough-set based approach for classification and rule induction. Int J Adv Manuf Technol 15:438–444
32. Khoo LP, Tor SB, Li JR (2001) A rough set approach to the ordering of basic events in a fault tree for fault diagnosis. Int J Adv Manuf Technol 17:769–794
33. Khoo LP, Zhai LY (2001) Multiconcept classification of diagnostic knowledge to manufacturing systems: analysis of incomplete data with continuous-valued attributes. Int J Prod Res 39(17):3941–3957
34. Kotler P (1997) Marketing management: analysis. Houghton Mifflin, New York
35. Kryszkiewicz M (1998) Rough set approach to incomplete information systems. Inf Sci 112:39–49
36. Kryszkiewicz M (1999) Rules in incomplete information systems. Inf Sci 113:71–292
37. Lalla M, Facchinetti G, Mastroleo G (2005) Ordinal scales and fuzzy set systems to measure agreement: an application to the evaluation of teaching activity. Qual Quant 38(5):577–601
38. Li WH, Chen SB, Wang B (2008) A variable precision rough set based modeling method for pulsed GTAW. Int J Adv Manuf Technol 36:1072–1079
39. Liao SH, Ho HH, Yang FC (2010) Ontology-based data mining approach implemented on exploring product and brand spectrum. Expert Syst Appl 36(9):11730–11744
40. Lingras P (2001) Unsupervised rough set classification using gas. J Intell Inf Syst 16:215–228
41. Liou JH, Tzen GH (2010) A dominance-based rough set approach for customer behavior in the airline market. Inf Sci 180(11):2230–2238
42. Mi JS, Wu WZ, Zhang WX (2004) Approaches to knowledge reduction based on variable precision rough set model. Inf Sci 159(3–4):255–272
43. Pal SK, Dasgupta B, Mitra P (2004) Rough self organizing map. Appl Intell 21(3):289–299
44. Pawlak Z (1982) Rough Sets Int J Comput Inf Sci 11:341–356
45. Pawlak Z, Skowron A (2007) A rough sets and boolean reasoning. Inf Sci 177(1):41–73
46. Rauch J (2005) Logic of association rules. Appl Intell 22(1):9–28
47. Rebai A, Aouni B, Martel JM (2006) A multi-attribute method for choosing among potential alternatives with ordinal evaluation. Eur J Oper Res 174(1):360–373
48. Sai Y, Yao Y, Zhong N (2001) Data analysis and mining in ordered information tables. In: IEEE international conference on data mining, pp 497–504
49. Santos E, Santos EE, Nguyen H, Pan L, Korah J (2011) A large-scale distributed framework for information retrieval in large dynamic search spaces. Appl Intell 35(3):375–398
50. Schwer CH (2002) Formalizing calendars with the category of ordinals. Appl Intell 17(3):275–295
51. Shie BE, Yu PS, Tseng VS (2013) Mining interesting user behavior patterns in mobile commerce environments. Appl Intell 38(3):418–435
52. Skowron A, Rauszer C 1992) The discernibility matrices and functions in information systems, in intelligent decision support—handbook of advances and applications of the rough set theory. Kluwer Academic, Dordrecht, pp 331–362
53. Swiniarski RW, Skowron A (2003) Rough set methods in feature selection and recognition. Pattern Recognit Lett 24:833–849
54. Tastle WJ, Wierman MJ (2006) An information theoretic measure for the evaluation of ordinal scale data. Behav Res Methods 38(3):487–494
55. Wang H, Wang S (2009) Discovering patterns of missing data in survey databases: an application of rough sets. Expert Syst Appl 36(3):6256–6260
56. Wang YT, Cheng JT (2011) Mining periodic movement patterns of mobile phone users based on an efficient sampling approach. Appl Intell 35(1):32–40
57. Wakker P (1989) Additive representations of preferences. Kluwer Academic, Dordrecht
58. Walczak B, Massart DL (1999) Rough sets theory. Chenom Intell Lab Syst 47(1):1–16
59. Wu CD, Zhang Y, Li MX, Yue Y (2006) A rough set GA-based hybrid method for robot path planning. Int J Autom Comput 3:29–34
60. Yue K, Yao Y, Li J, Liu WY (2010) Qualitative probabilistic networks with reduced ambiguities. Appl Intell 33(2):159–178
61. Zhai LY, Khoo LP, Fok SC (2001) Derivation of decision rules for the evaluation of product performance using genetic algorithms and rough set theory. In: Brasha D (ed) Data mining for design and manufacturing, methodology and applications. Kluwer Academic, Dordrecht
62. Zhang WX, Qiu GF, Wu WZ (2007) A general approach to attribute reduction in rough set theory. Sci China, Ser F 50(2):188–197

« **APIN 10489** layout: Large v.1.3.2 file: apin465.tex (Judita) class: spr-twocol-v1.4 v.2013/04/17 Prn:2013/07/19; 12:26 p. 15/15»
« doctopic: OriginalPaper numbering style: ContentOnly reference style: basic»

63. Zhao L, Wang L, Xu Q (2012) Data stream classification with artificial endocrine system. Appl Intell 37(3):390–404
64. Ziarko W (1993) Variable precision rough set model. J Comput Syst Sci 46:39–59
65. Ziarko W (1994) Rough sets and knowledge discovery: an overview. In: Ziarko W (ed) Rough sets, fuzzy sets and knowledge discovery—proceedings of the international workshop on rough sets and knowledge discovery (RSKD '93). Springer, London, pp 11–15

Authors' biographies with photos are desired!

Journal ID: 10489, Article ID: 465, Date: 2013-07-19, Proof No: 1, UNCORRECTED PROOF