Data Analysis and Mining in Ordered Information Tables

Ying Sai,   Y.Y. Yao
Department of Computer Science
University of Regina
Regina, Saskatchewan, Canada S4S 0A2
E-mail: yyao@cs.uregina.ca

Ning Zhong
Department of Information Engineering
Maebashi Institute of Technology
460-1, Kamisadori-Cho, Maebashi 371, Japan
E-mail: zhong@maebashi-it.ac.jp

Abstract

Many real world problems deal with ordering objects instead of classifying objects, although majority of research in machine learning and data mining has been focused on the latter. For modeling ordering problems, we generalize the notion of information tables to ordered information tables by adding order relations on attribute values. The problem of mining ordering rules is formulated as finding association between orderings of attribute values and the overall ordering of objects. An ordering rules may state that "if the value of an object x on an attribute a is ordered ahead of the value of another object y on the same attribute, then x is ordered ahead of y". For mining ordering rules, we first transform an ordered information table into a binary information, and then apply any standard machine learning and data mining algorithms. As an illustration, we analyze in detail Maclean's universities ranking for the year 2000.

1 Introduction

In real world situations, we may be faced with many problems that are not simply classification [1, 7]. One such type of problems is the ordering of objects. Two familiar examples are the ranking of universities and the ranking of consumer products produced by different manufacturers. In both examples, we have a set of attributes that are used to describe the objects under consideration. Consider the example of ranking consumer products. Attributes may be the price, warranty, and other information. The values of a particular attribute, say the price, naturally induce an ordering of objects. The overall ranking of products may be produced by their market shares of different manufacturers. The orderings of objects by attribute values may not necessarily be the same as the overall ordering of objects. In this setting, a number of important issues arise. It would be interesting to know which attributes play more important roles in determining the overall ordering, and which attributes do not contribute at all to the overall ordering. It would also be useful to know which subset of attributes would be sufficient to determine the overall ordering. The dependency information of attributes may also be valuable.

In this paper, we study the problem of data analysis with ordered information in general, and mining ordering rules in particular, based on our earlier results on this topic [12, 14, 15]. The problem of mining ordering rules can be described as follows. There is a set of objects described by a set of attributes. There is an ordering on values of each attribute, and there is also an overall ordering of objects. The overall ordering may be given by experts or obtained from other information, either dependent or independent of the orderings of objects according to their attribute values. We are interested in mining the association between the overall ordering and the individual orderings induced by different attributes. More specifically, we want to derive ordering rules exemplified by the statement that "if the value of an object x on an attribute a is ordered ahead of the value of another object y on the same attribute, then x is ordered ahead of y".

The notion of ordered information tables is introduced as a generalization of information tables. As a starting point, we focus on the dependency of attributes. The notions of reducts and core, which have been studied extensively in the literature of rough sets [6], are derivable from the dependency of attributes and play an important role in mining ordering rules. In mining ordering rules, we first transform an ordered information table into a binary information, and then apply any standard machine learning and data mining algorithms.

Ordered information tables are related to ordinal information systems proposed and studied by Iwinski [4]. Orderings induced by attribute values in information tables were also considered by Greco, Matarazzo and Slowinski [2, 3]. They addressed the problem of finding ordering rules by approximating preference relations through dominance relations. They also discussed approach of transforming an ordered information table into a binary table and mining the

corresponding binary table. However, they suggested that the method is not very suitable in some situations for decision making. Instead, they proposed a method based on the approximation of a preference relation by dominance relations, in which certain parameters must be determined by experiments. Although their method may be more general, it does not provide a pure qualitative framework, which may be important for the analysis of ordered information tables. Furthermore, one may gain more insights by studying a less general model.

2 Ordered Information Tables

In many information processing systems, objects are typically represented by their values on a finite set of attributes. Such information may be conveniently described in a tabular form [6]. The rows of the table correspond to objects of the universe, the columns correspond to a set of attributes, and each cell gives the value of an object with respect to an attribute.

Definition 1 *An information table is a quadruple:*

$$IT = (U, At, \{V_a \mid a \in At\}, \{I_a \mid a \in At\}),$$

where

U *is a finite nonempty set of objects,*

At *is a finite nonempty set of attributes,*

V_a *is a nonempty set of values for $a \in At$,*

$I_a : U \rightarrow V_a$ *is an information function.*

For simplicity, we have considered only information tables characterized by a finite set of objects and a finite set of attributes. Each information function I_a is a total function that maps an object of U to exactly one value in V_a. An information table represents all available information and knowledge about the objects under consideration. Objects are only perceived, observed, or measured by using a finite number of properties.

An information table does not consider any semantic relationships between distinct values of a particular attribute. By incorporating semantics information, we may obtain different generalizations of information tables [12]. Generalized information tables may be viewed as information tables with added semantics. In this paper, we introduce order relations on attribute values.

Definition 2 *Let U be a nonempty set and $\succ$ be a binary relation on U. The relation $\succ$ is a weak order if it satisfies the two properties:*

Asymmetry :

$$x \succ y \Longrightarrow \neg(y \succ x),$$

Negative transitivity :

$$(\neg(x \succ y), \neg(y \succ z)) \Longrightarrow \neg(x \succ z).$$

An important implication of a weak order is that the following relation,

$$x \sim y \Longleftrightarrow (\neg(x \succ y), \neg(y \succ x)), \tag{1}$$

is an equivalence relation. For two elements, if $x \sim y$ we say x and y are indiscernible by $\succ$. The equivalence relation $\sim$ induces a partition $U/\sim$ on U, and an order relation $\succ^*$ on $U/\sim$ can be defined by:

$$[x]_\sim \succ^* [y]_\sim \Longleftrightarrow x \succ y, \tag{2}$$

where $[x]_\sim$ is the equivalence class containing x. Moreover, $\succ^*$ is a linear order. Any two distinct equivalence classes of $U/\sim$ can be compared. It is therefore possible to arrange the objects into levels, with each level consisting of indiscernible elements defined by $\succ$. For a weak order, $\neg(x \succ y)$ can be written as $y \succeq x$ or $x \preceq y$, which means $y \succ x$ or $y \sim x$. For any two elements x and y, we have either $x \succ y$ or $y \succeq x$, but not both.

Definition 3 *An ordered information table is a pair:*

$$OIT = (IT, \{\succ_a \mid a \in At\}),$$

where IT is a standard information table and $\succ_a$ is a weak order on V_a.

An ordering of values of a particular attribute a naturally induces an ordering of objects, namely, for $x, y \in U$:

$$x \succ_{\{a\}} y \Longleftrightarrow I_a(x) \succ_a I_a(y), \tag{3}$$

where $\succ_{\{a\}}$ denotes an order relation on U induced by the attribute a. An object x is ranked ahead of another object y if and only if the value of x on the attribute a is ranked ahead of the value of y on a. The relation $\succ_{\{a\}}$ has exactly the same properties as that of $\succ_a$. For a subset of attributes $A \subseteq At$, we define:

$$
\begin{aligned}
x \succ_A y &\Longleftrightarrow \forall a \in A[I_a(x) \succ_a I_a(y)] \\
&\Longleftrightarrow \bigwedge_{a \in A} I_a(x) \succ_a I_a(y) \\
&\Longleftrightarrow \bigcap_{a \in A} \succ_{\{a\}} . \tag{4}
\end{aligned}
$$

That is, x is ranked ahead of y if and only if x is ranked ahead of y according to all attributes in A. The above definition is a straightforward generalization of the standard definition of equivalence relations in rough set theory, where the equality relation $=$ is used [6]. Mining ordering rules based on order relations is a concrete example of applications of our earlier studies on generalizations of rough set model with non-equivalence relations [11, 13].

For simplicity, we also assume that there is a special attribute, called the decision attribute. The ordering of objects by the decision attribute is denoted by $\succ_{\{o\}}$ and is called the overall ordering of objects.

	a	b	c	d	o
p_1	middle	3 years	$200	heavy	1
p_2	large	3 years	$300	very heavy	3
p_3	small	3 years	$300	light	3
p_4	small	3 years	$250	very light	2
p_5	small	2 years	$200	very light	3

$$
\begin{aligned}
\succ_a: \quad & \text{small} \succ_a \text{middle} \succ_a \text{large}, \\
\succ_b: \quad & \text{3 years} \succ_b \text{2 years}, \\
\succ_c: \quad & \$200 \succ_c \$250 \succ_c \$300, \\
\succ_d: \quad & \text{very light} \succ_d \text{light} \succ_d \text{heavy} \succ_d \text{very heavy}, \\
\succ_o: \quad & 1 \succ_o 2 \succ_o 3.
\end{aligned}
$$

Table 1. An ordered information table

Example 1 *Suppose we have an ordered information table of a group of products produced by five manufacturers as shown in Table 1. In this table, a, b, c, d, and o stand for size, warranty, price, weight, and overall ordering on a set of products, respectively. Based on orderings of attribute values, we obtain the following orderings of products:*

$$
\begin{aligned}
\succ^*_{\{a\}} \quad & : \quad [p_3, p_4, p_5] \succ^*_{\{a\}} [p_1] \succ^*_{\{a\}} [p_2], \\
\succ^*_{\{b\}} \quad & : \quad [p_1, p_2, p_3, p_4] \succ^*_{\{b\}} [p_5], \\
\succ^*_{\{c\}} \quad & : \quad [p_1, p_5] \succ^*_{\{c\}} [p_4] \succ^*_{\{c\}} [p_2, p_3], \\
\succ^*_{\{d\}} \quad & : \quad [p_4, p_5] \succ^*_{\{d\}} [p_3] \succ^*_{\{d\}} [p_1] \succ^*_{\{d\}} [p_2], \\
\succ^*_{\{o\}} \quad & : \quad [p_1] \succ^*_{\{o\}} [p_4] \succ^*_{\{o\}} [p_2, p_3, p_5].
\end{aligned}
$$

For subsets $\{a, b\}$ and $\{c, d\}$, we have:

$$
\begin{aligned}
\succ_{\{a,b\}} \quad & : \quad \emptyset, \\
\succ_{\{c,d\}} \quad & : \quad p_1 \succ_{\{c,d\}} p_2, \quad p_4 \succ_{\{c,d\}} p_2, \\
& \qquad p_5 \succ_{\{c,d\}} p_2, \quad p_4 \succ_{\{c,d\}} p_3, \\
& \qquad p_5 \succ_{\{c,d\}} p_3.
\end{aligned}
$$

By combining attributes a and b, all objects are put into the same class. It is interesting to note that $\succ_{\{c,d\}}$ is not a weak order. That is, the intersection of two weak orders may not produce a weak order. This suggests that rules using simple condition $\bigwedge_{a \in A} I_a(x) \succ_a I_a(y)$ might not be very useful.

3 Analyzing Ordered Information Tables

In an ordered information table OIT, an atomic expression over a single attribute a is defined as either $(a, \succ)$ or $(a, \preceq)$. For a set of attributes $A \subseteq At$, an expression over A in OIT is defined by $\bigwedge_{a \in A} e(a)$, where $e(a)$ is an atomic expression over a. The set of all expressions over A in an ordered information table OIT is denoted by $E(A)$. In general, we may also consider expressions connected by other

logic connectives such as $\neg$ and $\vee$. For simplicity, we restrict our discussion to only $\wedge$.

In an ordered information table, we deals with pairs of objects. We will not consider the object pairs (x, x). The universal set is thus given by $(U \times U)^+ = U \times U - \{(x, x) \mid x \in U\} = \{(x, y) \mid x, y \in U, x \neq y\}$. The meanings of the expressions are given as follows:

$$
\begin{aligned}
m(a, \succ) & = \{(x, y) \in (U \times U)^+ \mid I_a(x) \succ_a I_a(y)\}, \\
m(a, \preceq) & = \{(x, y) \in (U \times U)^+ \mid I_a(x) \preceq_a I_a(y)\}, \\
m(\bigwedge_{a \in A} e(a)) & = \bigcap_{a \in A} m(e(a)).
\end{aligned}
$$

An object pair (x, y) satisfies an expression ϕ, written $(x, y) \models \phi$, if the order as specified by the expression ϕ is (x, y). The set $m(\phi)$ consists of all object pairs that satisfy the expressed ϕ.

For the set of expressions $E(A)$, the family $\{m(\phi) \neq \emptyset \mid \phi \in E(A)\}$ forms a partition of $(U \times U)^+$, denoted by $P(A)$. Each object pair satisfies one and only one expression from $E(A)$.

Definition 4 *An attribute set B depends on another attribute set A, or A determines B, written $A \rightarrow B$, if and only if $P(A)$ is a finer partition of $P(B)$, namely, for each equivalence class $X \in P(A)$, there exist an equivalence class $Y \in P(B)$ such that $X \subseteq Y$.*

Suppose B depends on A. For any expression $\phi \in E(A)$, if an object pair $(x, y) \in (U \times U)^+$ satisfies ϕ, i.e., $(x, y) \models \phi$, then there exists a corresponding expression $\psi \in E(B)$, such that $(x, y) \models \psi$. The ordering of objects by A determines the ordering of objects by B.

Definition 5 *Two attribute sets $A, B \subseteq At$ are equivalent if and only if $P(A) = P(B)$.*

Two sets of attributes A and B are equivalent, if and only if A depends on B and B depends on A. The ordering of objects by A determines the ordering of objects by B, and vice versa. Regarding the ordering of objects, A and B contain exactly the same information.

Definition 6 *An attribute $a \in A$ is dispensable in A if and only if A and $A - \{a\}$ are equivalent; otherwise a is indispensable in A.*

Definition 7 *An attribute set A is an independent set if and only if every $a \in A$ is indispensable; otherwise A is a dependent set.*

Each attribute in an independent attribute set contributes to the ordering of objects.

Definition 8 *A subset $B \subseteq A$ is a reduct of A if and only if B is an independent set and is equivalent to A.*

A reduct $B \subseteq A$ is a minimal subset of A, in the sense that it does not contain any dispensable attributes, and produces the ordering of objects as given by A. Typically, a set of attributes may have many reducts. The set of all reducts of an attribute set A is denoted by $RED(A)$.

Definition 9 *The set of all indispensable elements of A is called the core of A and is denoted by $CORE(A)$.*

An attribute in the core must be in every reduct. Thus, we have:

$$CORE(A) = \bigcap RED(A). \tag{5}$$

The set of attributes $CORE(A)$ are so essential that they can not be eliminated without losing the ordering information provided by A. The core may be empty.

Definition 10 *An attribute $a \in A$ is absolutely dispensable in A if it does not belong to any reduct, namely, $a \in A - \bigcup RED(A)$.*

An absolutely dispensable attribute does not contribute at all to ordering of objects based on A.

Example 2 *Let $U = \{x, y, z\}$ and $A = \{a, b, c\}$. Consider the order relations induced by the three attributes:*

$$\begin{aligned}
\succ_{\{a\}}^*: \quad & [x] \succ_{\{a\}}^* [y] \succ_{\{a\}}^* [z], \\
\succ_{\{b\}}^*: \quad & [z] \succ_{\{b\}}^* [x, y], \\
\succ_{\{c\}}^*: \quad & [z] \succ_{\{c\}}^* [y] \succ_{\{c\}}^* [x].
\end{aligned}$$

For the entire set of attributes, we have:

$$\begin{aligned}
E(\{a, b, c\}) = \{ & (a, \succ) \wedge (b, \preceq) \wedge (c, \preceq), \\
& (a, \preceq) \wedge (b, \preceq) \wedge (c, \succ), \\
& (a, \preceq) \wedge (b, \succ) \wedge (c, \succ), \ldots \}.
\end{aligned}$$

For clarity, we only explicitly give the expressions whose meanings are nonempty sets of object pairs. The corresponding partition, $P(\{a, b, c\})$, is given by:

$$\begin{aligned}
m((a, \succ) \wedge (b, \preceq) \wedge (c, \preceq)) &= \{(x, y), (x, z), (y, z)\}, \\
m((a, \preceq) \wedge (b, \preceq) \wedge (c, \succ)) &= \{(y, x)\}, \\
m((a, \preceq) \wedge (b, \succ) \wedge (c, \succ)) &= \{(z, x), (z, y)\}.
\end{aligned}$$

For the subset of attributes $\{a, b\}$, we have:

$$\begin{aligned}
E(\{a, b\}) = \{ & (a, \succ) \wedge (b, \preceq), (a, \preceq) \wedge (b, \preceq), \\
& (a, \preceq) \wedge (b, \succ), \ldots \},
\end{aligned}$$

which induces the partition $P(\{a, b\})$:

$$\begin{aligned}
m((a, \succ) \wedge (b, \preceq)) &= \{(x, y), (x, z), (y, z)\}, \\
m((a, \preceq) \wedge (b, \preceq)) &= \{(y, x)\}, \\
m((a, \preceq) \wedge (b, \succ)) &= \{(z, x), (z, y)\}.
\end{aligned}$$

Since $P(\{a, b, c\}) = P(\{a, b\})$, the attribute sets $\{a, b, c\}$ and $\{a, b\}$ are equivalent. It can be easily verified that $P(\{a, b\}) \neq P(\{a\})$ and $P(\{a, b\}) \neq P(\{b\})$. Thus, the attribute set $\{a, b\}$ is an independent set and is a reduct of A. Another reduct is given by $\{b, c\}$. We have $RED(A) = \{\{a, b\}, \{b, c\}\}$ and $CORE(A) = \{b\}$. There is no absolute dispensable attribute.

The concepts of dependency, dispensable attributes and reduct capture the connections between attributes with respect to ordering produced by a set A. They provide us with a basis for analyzing and simplifying an ordered information table by removing certain attribute without loss of information. In some situations, it is necessary to study relationship between attributes relative to the overall ordering of objects. In what follows, we consider an ordered information table with a decision attribute o, i.e., the overall ordering of objects. All the concepts developed so far can be easily extended.

Definition 11 *Let $A \subseteq At$ be an attribute set. For an expression $\phi \in E(A)$, we say o ϕ-depends on A, denoted by $A \rightarrow_\phi o$, if and only if $m(\phi) \subseteq m(o, \succ)$ or $m(\phi) \subseteq m(o, \preceq)$.*

While dependency of attributes show global connection between attributes, ϕ-dependency shows local connection. Assume o ϕ-depends on A. If any two objects satisfy the particular orderings specified by ϕ on A, one can tell their ordering by o.

Definition 12 *Let $A, B \subseteq At$. Two attributes sets A and B are equivalent with respect to o, if and only if*

$$\bigcup_{\phi \in E(A)} M_o(\phi) = \bigcup_{\phi' \in E(B)} M_o(\phi'), \tag{6}$$

where $M_o(\phi)$ is defined by:

$$M_o(\phi) = \begin{cases} m(\phi), & A \rightarrow_\phi o, \\ \emptyset, & otherwise. \end{cases} \tag{7}$$

If A and B are equivalent with respect to o, any two objects that can be properly ordered on o according to A must be properly ordered according to B, and vice versa.

Definition 13 *An attribute $a \in A$ is dispensable with respect to o, if and only if A and $A - \{a\}$ are equivalent with respect to o; otherwise a is indispensable with respect to o.*

Definition 14 *An attribute set A is an independent set with respect to o, if and only if there is no dispensable attribute in A with respect to o.*

Definition 15 *A subset $B \subseteq A$ is a reduct of A with respect to o, if and only if B is an independent set and equivalent to A with respect to o.*

4

A reduct $B \subseteq A$ with respect to o is sometimes called a relative reduct of A with respect to o. It consists of a minimal subset of attributes of A that provides the same ordering of objects on o as A itself. A set of attributes may have more than one relative reduct. The set of all relative reducts of an attribute set A is denoted by $RED_o(A)$.

Definition 16 *The set of all indispensable elements of A with respect to o is called the core of A with respect to o and is denoted by $CORE_o(A)$.*

Similarly, the core of A with respect to o is called a relative core. The relative core of A is the intersection of all its relative reducts:

$$CORE_o(A) = \bigcap RED_o(A). \qquad (8)$$

A relative core may be empty. Attributes in a nonempty relative core play essential roles in describing the overall ordering induced by o.

Example 3 *Let $U = \{x, y, z\}$, $A = \{a, b, c\}$, and $At = A \cup \{o\}$. Consider order relations induced by attributes given by:*

$$\begin{aligned}
\succ_{\{a\}}^*: & \quad [x] \succ_{\{a\}}^* [y, z], \\
\succ_{\{b\}}^*: & \quad [x, y] \succ_{\{b\}}^* [z], \\
\succ_{\{c\}}^*: & \quad [z] \succ_{\{c\}}^* [y] \succ_{\{c\}}^* [x], \\
\succ_{\{o\}}^*: & \quad [x] \succ_{\{o\}}^* [y] \succ_{\{o\}}^* [z].
\end{aligned}$$

For the set $\{a, b, c\}$, we have:

$$\begin{aligned}
E(\{a, b, c\}) = & \ \{(a, \succ) \wedge (b, \preceq) \wedge (c, \preceq), \\
& \ (a, \succ) \wedge (b, \succ) \wedge (c, \preceq), \\
& \ (a, \preceq) \wedge (b, \preceq) \wedge (c, \succ), \\
& \ (a, \preceq) \wedge (b, \succ) \wedge (c, \preceq), \ldots\} \\
= & \ \{\phi_1, \phi_2, \phi_3, \phi_4, \ldots\}.
\end{aligned}$$

The relationships between orderings induced by a, b, c and o can be seen as follows:

$$\begin{aligned}
m(o, \succ) & = \{(x, y), (x, z), (y, z)\}, \\
m(o, \preceq) & = \{(y, x), (z, x), (z, y)\}, \\
m(\phi_1) & = \{(x, y)\} \subseteq m(o, \succ), \\
m(\phi_2) & = \{(x, z)\} \subseteq m(o, \succ), \\
m(\phi_3) & = \{(y, x), (z, x), (z, y)\} \subseteq m(o, \preceq), \\
m(\phi_4) & = \{(y, z)\} \subseteq m(o, \succ).
\end{aligned}$$

We have:

$$\bigcup_{\phi \in E(A)} M_o(\phi) = (U \times U)^+.$$

For the subset $\{a, b\}$, we obtain:

$$\begin{aligned}
E(\{a, b\}) & = \ \{(a, \succ) \wedge (b, \preceq), (a, \succ) \wedge (b, \succ), \\
& \quad (a, \preceq) \wedge (b, \preceq), (a, \preceq) \wedge (b, \succ)\} \\
& = \ \{\phi_1', \phi_2', \phi_3', \phi_4'\}.
\end{aligned}$$

We have:

$$\begin{aligned}
m(\phi_1') & = \{(x, y)\} \ \subseteq m(o, \succ), \\
m(\phi_2') & = \{(x, z)\} \ \subseteq m(o, \succ), \\
m(\phi_3') & = \{(y, x), (z, x), (z, y)\} \ \subseteq m(o, \preceq), \\
m(\phi_4') & = \{(y, z)\} \ \subseteq m(o, \succ),
\end{aligned}$$

and

$$\bigcup_{\phi' \in E(\{a, b\})} M_o(\phi') = (U \times U)^+.$$

Since $\bigcup_{\phi \in E(A)} M_o(\phi) = \bigcup_{\phi' \in E(\{a, b\})} M_o(\phi')$, attribute sets $\{a, b, c\}$ and $\{a, b\}$ are equivalent with respect to o. It is easy to check that $\{a\}$ and $\{b\}$ are not equivalent to A, i.e., the attribute set $\{a, b\}$ is an independent set, and is is a relative reduct. One can also show that $\{c\}$ is the other relative reduct of A. Hence we have $RED_o(A) = \{\{a, b\}, \{c\}\}$ and $CORE_o(A) = \emptyset$.

4 Mining Ordering Rules

Data mining in an ordered information table may be formulated as finding association between orderings induced by attributes. One is interested in finding associations between two arbitrary subsets of attributes A and B.

Definition 17 *Consider two subsets of attributes $A, B \subseteq At$. For two expressions $\phi \in E(A)$ and $\psi \in E(B)$, an ordering rule is read "if ϕ then ψ" and denoted by $\phi \Rightarrow \psi$. The expression ϕ is called the rule's antecedent, while the expression ψ is called the rule's consequent.*

An ordering rule states how orderings of objects by attributes in A determines orderings of objects by attributes in B. For example, an ordering rule,

$$(a, \succ) \wedge (b, \preceq) \Rightarrow (c, \succ),$$

can be re-expressed as,

$$x \succ_{\{a\}} y \wedge x \preceq_{\{b\}} y \Rightarrow x \succ_{\{c\}} y.$$

That is, for two arbitrary objects x and y, if x is ranked ahead of y by attribute a, and at the same time, x is not ranked ahead of y by attribute b, then x is ranked ahead of y by attribute c. We adopt conditional probabilistic interpretations for ordering rules. A systematic analysis of probability related quantities associated with rules was given by Yao and Zhong [16]. We choose to use two measures called accuracy and coverage [10].

Definition 18 *The accuracy and coverage of an ordering rule, $\phi \Rightarrow \psi$, are defined, respectively, by:*

$$accuracy(\phi \Rightarrow \psi) = \frac{|m(\phi \wedge \psi)|}{|m(\phi)|},$$

$$coverage(\phi \Rightarrow \psi) = \frac{|m(\phi \wedge \psi)|}{|m(\psi)|}, \qquad (9)$$

where $|\cdot|$ denotes the cardinality of a set.

The quantity $accuracy(\phi \Rightarrow \psi)$ gives a measure of the correctness of the rule, and $coverage(\phi \Rightarrow \psi)$ gives a measure of the applicability of the rule. In the extreme case, if $accuracy = 1$, the ordering by ϕ would determine the ordering by ψ. If $coverage = 1$, the rule states something about the entire ordering by ψ. The accuracy and coverage are not independent of each other, as both are related to the quantity $|m(\phi \wedge \psi)|$. It is desirable for a rule to be accurate as well as to have a high degree of coverage. In general, one may observe a trade-off between accuracy and coverage. A rule with higher coverage may have a lower accuracy, while a rule with higher accuracy may have a lower coverage.

Example 4 *From the data in Example 1, we can get, for example, two ordering rules:*

$$(b, \preceq) \wedge (c, \preceq) \Rightarrow (o, \preceq),$$
$$accuracy = 1.0, \ \ coverage = 0.615,$$
$$(c, \succ) \Rightarrow (o, \succ),$$
$$accuracy = 0.625, \ \ coverage = 0.714.$$

To mine ordering rules from an ordered information table, we transform an ordered information table into a binary information. Any standard data mining algorithms can be immediately applied. In the binary information table, we consider object pairs $(x, y) \in (U \times U)^+$. The information function is defined by:

$$I_a(x, y) = \begin{cases} 1, & x \succ_{\{a\}} y, \\ 0, & x \preceq_{\{a\}} y. \end{cases} \qquad (10)$$

Statements in an ordered information table can be translated into equivalent statements in the binary information table. For example, $x \succ_{\{a\}} y$ can be translated into $I_a(x, y) = 1$. In the translation process, we will not consider object pairs of the form (x, x), as we are not interested in them.

In our approach, the interpretation of an ordered information table and the translation to a binary information table are crucial. Once we obtain the binary information table, any existing data mining algorithm can be used to mine ordering rules. For example, using Rosetta, a rough set toolkit for analyzing data, one can obtain a set of minimal ordering rules from reducts with certain properties [9]. A well known machine learning system C4.5 can also be used to learn ordering rules [8].

Object	a	b	c	d	o
$(1,2)$	1	0	1	1	1
$(1,3)$	0	0	1	0	1
$(1,4)$	0	0	1	0	1
$(1,5)$	0	1	0	0	1
$(2,1)$	0	0	0	0	0
$(2,3)$	0	0	0	0	0
$(2,4)$	0	0	0	0	0
$(2,5)$	0	1	0	0	0
$(3,1)$	1	0	0	1	0
$(3,2)$	1	0	0	1	0
$(3,4)$	0	0	0	0	0
$(3,5)$	0	1	0	0	0
$(4,1)$	1	0	0	1	0
$(4,2)$	1	0	1	1	1
$(4,3)$	0	0	1	1	1
$(4,5)$	0	1	0	0	1
$(5,1)$	1	0	0	1	0
$(5,2)$	1	0	1	1	0
$(5,3)$	0	0	1	1	0
$(5,4)$	0	0	1	0	0

Table 2. A binary information table derived from an ordered information table

Example 5 *Consider the ordered information table in Example 1. It can be transformed into the binary information table, given by Table 2. For this table, we obtain a reduct consisting of $\{b, c\}$. Examples of rules are given by:*

$R_1:$ $(b, \preceq) \wedge (c, \preceq) \Rightarrow (o, \preceq),$
       $accuracy = 1, \ \ coverage = 0.615;$

$R_2:$ $(c, \succ) \Rightarrow (o, \succ),$
       $accuracy = 0.625, \ \ coverage = 0.714;$

$R_3:$ $(b, \succ) \Rightarrow (o, \succ),$
       $accuracy = 0.5, \ \ coverage = 0.286.$

For R_2, it can be paraphrased as saying that if $x \succ_{\{c\}} y$ then $x \succ_{\{o\}} y$ with accuracy 0.625 and coverage 0.714.

One may use other types of translation methods. For example, we may consider two strict order relations $\succ$ and $\prec$, instead of $\succ$ and $\preceq$. Alternatively, one may translate an ordered information table into a three-valued information table, corresponding to $\succ$, $\prec$, and $\sim$. Greco, Matarazzo and Slowinski used graded preference relations which produce a multi-valued information table [3]. By translating an ordered information table into a multi-valued information table, one may consider magnitudes of attribute values. Unfortunately, it may be difficult to interpret these magnitudes. They must be obtained from either experts or experiments.

For this reason, we focus on the binary information table approach. It is important to realize that the framework presented in this paper can be easily applied with very simple modification.

5   A Real World Example

To demonstrate the usefulness and effectiveness of the proposed model, we analyze the data from the Maclean's universities 2000 ranking [5]. The Maclean's ranking compared universities with similar structures and mandates. All universities in Canada are classified into one of three categories: Medical/Doctoral universities, Comprehensive universities, and Primarily Undergraduate universities. Maclean's ranked 15 Medical/Doctoral universities based on 22 performance measure that are divided into six groupings, namely, Student Body (22%), Classes (17%), Faculty (17%), Finances (12%), Library (12 %), and Reputation (20%), where the weights in the parentheses indicate the contributions of the measures to the overall ranking. Table 3 summarizes the ranking measures.

The 22 performance measures produce 22 individual rankings and an overall ranking suggested by Maclean's. The details can be found in the November issue of Maclean's [5]. Using data analysis and data mining method introduced earlier, we can find reduct(s) and the core of the ordered information table. We can also induce ordering rules which state association between the overall ranking and the rankings by 22 performance measures.

For our experiment, we used the Rosetta, a rough set toolkit for analyzing data [9]. We found that there are several reducts for this information table. One reduct with the minimal set length for explanation is $\{c1, f3, fi3, l4, r1, s2, s5\}$. This suggests that one can use seven measures to provides the same overall ranking of universities as given by Maclean's. The total weight of the seven measures is only 30.8%. In addition, measures in this reduct do not necessarily have the highest weights among performance measures. The number of measures in the reduct and the total weight are less than 1/3 of the ones used in the entire table. We also find that the core is empty, which means that there does not exist a measure that is absolutely essential for producing the overall ranking. These observations are interesting, as they indicate that some measures used by Maclean's may not be necessary, and moreover the weights may not necessarily reflect the importance of measures in producing the overall ranking. Another interesting observation of the the reduct $\{c1, f3, fi3, l4, r1, s2, s5\}$ is that it contains at least one measure from each of the six categories.

Many ordering rules were found by using Rosetta. As examples, we list some rules with accuracy 1 and the highest degrees of coverage:

$$R_1 : (f4, \succ) \land (fi3, \succ) \land (s2, \succ) \Rightarrow (o, \succ),$$

$$coverage = 0.6,$$
$$R_2 : (f2, \succ) \land (fi3, \succ) \land (r1, \succ) \Rightarrow (o, \succ),$$
$$coverage = 0.590476,$$
$$R_3 : (fi3, \succ) \land (r1, \succ) \land (s2, \succ) \Rightarrow (o, \succ),$$
$$coverage = 0.590476,$$
$$R_4 : (f1, \succ) \land (fi3, \succ) \land (s1, \succ) \Rightarrow (o, \succ),$$
$$coverage = 0.580952,$$
$$R_5 : (f4, \succ) \land (r2, \succ) \land (s2, \succ) \Rightarrow (o, \succ),$$
$$coverage = 0.580952,$$
$$R_6 : (f4, \succ) \land (fi3, \succ) \land (s1, \succ) \Rightarrow (o, \succ),$$
$$coverage = 0.571429.$$

All the rules use only three performance measures. Even using such a small number of measures, one can correctly state close to 60% of the overall ranking.

Although the preliminary results obtained from the Maclean's university ranking may need further study, indepth analysis, and careful interpretation, they at least present another point of view to look at the important issue of university ranking. An added advantage is that our method is of qualitative nature, without considering any quantitative information.

6   Conclusions

Ordering of objects is a fundamental issue in human decision making and may play a significant role in the design of intelligent information systems. This problem is considered from the perspective of data mining. The commonly used attribute value approaches are extended by introducing order relations on attribute values. A data analysis model is thus proposed to describe properties of ordered information tables. Mining ordering rules is formulated as the process of finding associations between orderings on attribute values and the overall ordering of objects. These ordering rules tell us, or explain, how objects should be ranked according to orderings on their attribute values.

The proposed solution for mining ordering rules is simple. Our main contribution is the formulation of the problem, and the translation of the problem to existing data mining problem. Consequently, one can directly apply any existing data mining algorithms for mining ordering rules. The application of the proposed model to a real world example demonstrates its usefulness and effectiveness. Depending on the specific problems, one may use different translation methods.

References

[1] Cohen, W.W., Schapire, R.E., and Singer, Y., Learning to order things, *Advances in Neural Information*

measure	meaning	weight
s1	incoming students' average high-school grades	12%
s2	proportion of those with averages of 75 per cent or more	3%
s3	proportion of out-of-province first-year undergraduate students	1%
s4	percentage of international students at the graduate level	1%
s5	graduation rates	2%
s6	proportion of winning national academic awards	3%
c1	class sizes at the first- and second-year levels	7%
c2	class sizes at the third- and fourth-year levels	7%
c3	percentage of first-year classes taught by tenured/tenure-track professors	3%
f1	percentage of those with PhDs	3%
f2	the number who win national awards	3%
f3	social sciences and humanities grants, Canada Council grants last year	5.5%
f4	medical/science grants last year	5.5%
fi1	the amount of money for current expenses per weighted full-time-equivalent student	3.3%
fi2	percentage of the budget spent on student services	4.3%
fi3	scholarships and bursaries	4.3%
l1	the number of volumes and volume equivalents per total number of students	3%
l2	total holdings	1%
l3	percentage of a university's operating budget that was allocated to library services	4%
l4	percentage of the actual library budget that was spent on updating the collection	4%
r1	alumni support	5%
r2	reputational survey	15%

Table 3. Maclean's performance measures for universities ranking, 2000

Processing Systems, **10**, 1998.

[2] Greco, S., Matarazzo, B., and Slowinski, R., The use of rough sets and fuzzy sets in MCDM, In: *Advances in Multiple Criteria Decision Making*, Gal, T., Hanne, T., and Stewart, T. (Eds.) Kluwer Academic Publishers, Boston, pp. 14.1-14.59, 1999.

[3] Greco, S., Matarazzo, B., and Slowinski, R., Rough approximation of a preference relation by dominance relations, *European Journal of Operational Research* **117**, 63-83, 1999.

[4] Iwinski, T.B., Ordinal information system, I, *Bulletin of the Polish Academy of Sciences, Mathematics*, **36**, 467-475, 1988.

[5] Maclean's Universities 2000, *Macleans*, November 20, 2000.

[6] Pawlak, Z., *Rough Sets, Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Dordrecht, 1991.

[7] Pawlak, Z. and Slowinski, R., Rough set approach to multi-attribute decision analysis. *European Journal of Operational Research*, **72**, 443-359, 1994.

[8] Quinlan, J.R., *C4.5: Program for Machine Learning*, Morgan Kaufmann Publishers, San Marteo, 1993.

[9] Rosetta, a rough set toolkit for analyzing data, *http://www.idi.ntnu.no/ aleks/rosetta/*.

[10] Tsumoto, S., Modelling medical diagnostic rules based on rough sets, *Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence, 1424*, Springer-Verlag, Berlin, pp. 475-482, 1998.

[11] Yao, Y.Y., Generalized rough set models, in: *Rough Sets in Knowledge Discovery*, Polkowski, L. and Skowron, A. (Eds.), Physica-Verlag, Heidelberg, pp. 286-318, 1998.

[12] Yao, Y.Y., Information tables with neighborhood semantics, in: *Data Mining and Knowledge Discovery: Theory, Tools, and Technology II*, Dasarathy, B.V. (Ed.), Society for Optical Engineering, Bellingham, Washington, pp. 108-116, 2000.

[13] Yao, Y.Y. and Lin, T.Y., Generalization of rough sets using modal logic, *Intelligent Automation and Soft Computing, An International Journal*, **2**, 103-120, 1996.

[14] Yao, Y.Y. and Sai, Y., Mining ordering rules using rough set theory, *Bulletin of International Rough Set Society*, **5**, 99-106, 2001.

[15] Yao, Y.Y. and Sai, Y., On mining ordering rules, manuscript, 2001.

[16] Yao, Y.Y. and Zhong, N., An analysis of quantitative measures associated with rules, *Proceedings of PAKDD'99*, 479-488, 1999.