Table of Contents

1. On Mining Ordering Rules

Y.Y. Yao and Ying Sai

Department of Computer Science, University of Regina
Regina, Saskatchewan, Canada S4S 0A2
e-mail: yyao@cs.uregina.ca

Many real world problems deal with ordering of objects instead of classifying objects, although majority of research in machine learning and data mining has been focused on the latter. In this paper, we formulate the problem of mining ordering rules as finding association between orderings of attribute values and the overall ordering of objects. An example of 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, the notion of information tables is generalized to ordered information tables by adding order relations on attribute values. Such a table can be transformed into a binary information table, on which any standard data mining algorithm can be used.

1.1 Introduction

In real world situations, we may be faced with many problems that are not simply classification [1.1, 1.4]. One such type of problems is the ordering of objects. Two familiar examples of ordering problems are the ranking of universities and the ranking of the consumer products produced by different manufactures. In both examples, we have a set of attributes that are used to describe the objects under consideration, and an overall ranking of objects. Consider the example of ranking consumer products. Attributes may be the price of the products, warranty of the products, 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 the market shares of different manufactures. The orderings of objects by attribute values may not necessarily be the same as the overall ordering of objects.

The problem of mining ordering rules can be stated 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". 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.

For mining ordering rules, we first introduce the notion of ordered information tables as a generalization of information tables. We then transform an ordered information table into a binary information table, on which any standard data mining and machine learning algorithms can be applied. Typically, an ordering rule may not be exact. In order to capture the uncertainty associated with ordering rules, two quantitative measures are used. They are the accuracy and the coverage of the rules [1.5, 1.7]. The former deals with the correctness of the rules, and the latter represents the extent to which the rule covers the positive instances.

Ordered information tables are related to ordinal information systems proposed and studied by Iwinski [1.3]. Mining ordering rules has been studied by Greco, Matarazzo and Slowinski [1.2]. Based on these studies, the main objective of the present paper is to precisely define and formulate the problem of mining ordering rules.

1.2 Ordered Information Tables

Formally, an ordered information table is defined by:

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

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 $a \in At$,

$I_a : U \to V_a$ is an information function,

$\succ_a \subseteq V_a \times V_a$ is an order relation on V_a.

Each information function I_a is a total function that maps an object of U to exactly one value in V_a. An ordered information table can be conveniently given in a tabular form, the rows correspond to objects of the universe, the columns correspond to a set of attributes, and each cell is the value of an object with respect to an attribute. The order relations can be interpreted as additional semantics information about the table.

An order relation should satisfy certain conditions. We consider the following two properties [1.6]:

$$\text{Asymmetry}: \ x \succ y \Longrightarrow \neg(y \succ x),$$
$$\text{Negative transitivity}: \ [\neg(x \succ y), \neg(y \succ z)] \Longrightarrow \neg(x \succ z).$$

An order relation satisfying these properties is called a weak order. An important implication of a weak order is that the following relation,

$$x \sim y \Longleftrightarrow [\neg(x \succ y), \neg(y \succ x)], \tag{1.1}$$

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 on $U/\sim$ can be defined by:

$$[x]_\sim \succ^* [y]_\sim \Longleftrightarrow x \succ y, \tag{1.2}$$

where $[x]_\sim$ is the equivalence class containing x. Moreover, $\succ^*$ is a linear order [1.6]. Any two distinct equivalence classes of $U/\sim$ can be compared. It is therefore possible to arrange the elements 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.

We assume that all order relations are weak orders. An order relation on values of an attribute a naturally induces an ordering of objects:

$$x \succ_{\{a\}} y \Longleftrightarrow I_a(x) \succ_a I_a(y), \tag{1.3}$$

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 simplicity, we also assume that there is a special attribute, called decision attribute. The ordering of objects by the decision attribute is denoted by $\succ_o$ and is called the overall ordering of objects. For a subset of attributes $A \subseteq At$, we define:

$$\begin{aligned} x \succ_A y &\Longleftrightarrow \forall a \in A[I_a(x) \succ_a I_a(y)] \\ &\Longleftrightarrow \bigwedge_{a \in A} I_a(x) \succ_a I_a(y) \Longleftrightarrow \bigcap_{a \in A} \succ_{\{a\}} . \end{aligned} \tag{1.4}$$

That is, x is ranked ahead of y if and only if x is ranked ahead of y according to all attributes in A.

1.3 Mining Ordering Rules

With an ordered information table, we are interested in find ordering rules of the form $\phi \Rightarrow \psi$, where ϕ and ψ are expressions regarding ordering of objects based on certain attributes. For an attribute a, we can construct two atomic expressions $(a, \succ)$ and $(a, \preceq)$. The former indicates that objects are ordered based on $\succ$ and the latter indicates that objects are ordered based on $\preceq$. A set of expressions can be obtained from atomic expressions through the application of logic connectives $\neg$, $\wedge$ and $\vee$. Consider an ordering rule,

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

It can be re-expressed as,

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

and paraphrased as follows. 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.

The meanings of expressions are defined by:

(m1). $m((a, \succ)) = \{(x, y) \in U \times U \mid x \succ_{\{a\}} y\}$,
(m2). $m((a, \preceq)) = \{(x, y) \in U \times U \mid x \preceq_{\{a\}} y\}$,
(m3). $m(\neg\phi) = -m(\phi)$,
(m4). $m(\phi \wedge \psi) = m(\phi) \cap m(\psi)$,
(m5). $m(\phi \vee \psi) = m(\phi) \cup m(\psi)$.

A pair $(x, y) \in m(\phi)$ is said to satisfy the expression ϕ. In terms of the meanings of expressions, we can have many conditional probabilistic interpretations for ordering rules [1.7]. We choose to use two measures called accuracy and coverage, which are defined by [1.5]:

$$accuracy(\phi \Rightarrow \psi) = \frac{|m(\phi \wedge \psi)|}{|m(\phi)|}, \quad coverage(\phi \Rightarrow \psi) = \frac{|m(\phi \wedge \psi)|}{|m(\psi)|}, \tag{1.5}$$

where $|\cdot|$ denotes the cardinality of a set. While the accuracy reflects the correctness of the rule, the coverage reflects the applicability of the rule. If $accuracy(\phi \Rightarrow \psi) = 1$, the orderings by ϕ would determine the orderings by ψ. We thus have a strong association between the two orderings. A smaller value of *accuracy* indicates a weaker association. An ordering rule with higher coverage suggest that ordering of more pairs of objects can be derived from the rule. 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.

From an ordered information table, we can construct a binary information table. We consider all pairs of objects which are the Cartesian product $U \times U$. The information function is defined by:

$$I_a(x, y) = \begin{cases} 1, & x \succ_{\{a\}} y, \\ 0, & x \preceq_{\{a\}} y. \end{cases} \tag{1.6}$$

The value 1 corresponds to the atomic expression $(a, \succ)$ and the value 0 corresponds to the atomic expression $(a, \preceq)$. Statements in an ordered information table can be translated into equivalent statements in the binary information table, and vice versa. For example, a pair (x, y) satisfies the expression $(a, \succ)$ if and only if it satisfies an expression $I_a(x, y) = 1$. In other words, the statement $x \succ_{\{a\}} y$ can be translated into an equivalent statement $I_a(x, y) = 1$. In the translation process, we will not consider object pairs of the form (x, x), as we are not interested in them.

The interpretation of an ordered information table and the translation to a binary information table are crucial for mining ordering rules. Once we obtain the binary information table, any standard machine learning and data mining algorithms can be used to mine ordering rules. One may also 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 $\sim$. It is important to realized that the framework presented in this paper can be easily applied with very simple modification.

1.4 Conclusion

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. 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.

Our main contribution is the formulation of the problem of mining ordering rules, and the translation of the problem to existing data mining problems. Consequently, one can directly apply any existing data mining algorithms for mining ordering rules. Depending on the specific problem, one may use different translation methods.

References

1.1 Cohen, W.W., Schapire, R.E. and Singer, Y. Learning to order things, *Journal of Artificial Intelligence Research*, **10**, 243-270, 1999.
1.2 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.
1.3 Iwinski, T.B. Ordinal information system, I, *Bulletin of the Polish Academy of Sciences, Mathematics*, **36**, 467-475, 1988.
1.4 Pawlak, Z., Slowinski, R. Rough set approach to multi-attribute decision analysis, *European Journal of Operational Research*, **72**, 443-359, 1994.
1.5 Tsumoto, S. Automated discovery of plausible rules based on rough sets and rough inclusion, *Proceedings of PAKDD'99*, 210-219, 1999.
1.6 Yao, Y.Y. Measuring retrieval effectiveness based on user preference of documents, *Journal of the American Society for Information Science*, **46**, 133-145, 1995.
1.7 Yao, Y.Y. and Zhong, N. An analysis of quantitative measures associated with rules, *Proceedings of PAKDD'99*, 479-488, 1999.