Interestingness Measures for Data Mining: A Survey

LIQIANG GENG AND HOWARD J. HAMILTON

University of Regina

Interestingness measures play an important role in data mining, regardless of the kind of patterns being mined. These measures are intended for selecting and ranking patterns according to their potential interest to the user. Good measures also allow the time and space costs of the mining process to be reduced. This survey reviews the interestingness measures for rules and summaries, classifies them from several perspectives, compares their properties, identifies their roles in the data mining process, gives strategies for selecting appropriate measures for applications, and identifies opportunities for future research in this area.

Categories and Subject Descriptors: H.2.8 [**Database Management**]: Database Applications—*Data mining*

General Terms: Algorithms, Measurement

Additional Key Words and Phrases: Knowledge discovery, classification rules, interestingness measures, interest measures, summaries, association rules

1. INTRODUCTION

In this article, we survey measures of interestingness for *data mining*. Data mining can be regarded as an algorithmic process that takes data as input and yields patterns such as *classification rules*, *association rules*, or *summaries* as output. An association rule is an implication of the form $X \rightarrow Y$, where X and Y are nonintersecting sets of items. For example, {milk, eggs} $\rightarrow$ {bread} is an association rule that says that when milk and eggs are purchased, bread is likely to be purchased as well. A classification rule is an implication of the form X_1 op x_1, X_2 op $x_2, \ldots, X_n$ op $x_n \rightarrow Y = y$, where X_i is a conditional attribute, x_i is a value that belongs to the domain of X_i, Y is the class attribute, y is a class value, and op is a relational operator such as $=$ or $>$. For example, $Job = Yes, AnnualIncome > 50{,}000 \rightarrow Credit = Good$, is a classification rule which says that a client who has a job and an annual income of more than \$50,000 is classified as having good credit. A summary is a set of attribute-value pairs and aggregated counts, where the values may be given at a higher level of generality than the values in the input data. For example, the first three columns of Table I form a summary of

---

The authors gratefully acknowledge the National Science and Engineering Research Council of Canada for providing funds to support this research via a Discovery Grant, a Collaborative Research and Development Grant, and a Strategic Project Grant awarded to the H. J. Hamilton.

Authors' address: L. Geng and H. J. Hamilton, Department of Computer Science, University of Regina, Regina, Saskatchewan, Canada; email: {gengl,hamilton}@cs.uregina.ca.

Table I. Summary of Students Majoring in Computer Science

Program	Nationality	# of Students	Uniform Distribution	Expected
Graduate	Canadian	15	75	20
Graduate	Foreign	25	75	30
Undergraduate	Canadian	200	75	180
Undergraduate	Foreign	60	75	70

the students majoring in computer science in terms of two attributes: nationality and program. In this case, the value of "Foreign" for the nationality attribute is given at a higher level of generality in the summary than in the input data, which gives individual nationalities.

Measuring the interestingness of discovered patterns is an active and important area of data mining research. Although much work has been conducted in this area, so far there is no widespread agreement on a formal definition of interestingness in this context. Based on the diversity of definitions presented to-date, interestingness is perhaps best treated as a broad concept that emphasizes *conciseness*, *coverage*, *reliability*, *peculiarity*, *diversity*, *novelty*, *surprisingness*, *utility*, and *actionability*. These nine specific criteria are used to determine whether or not a pattern is interesting. They are described as follows.

Conciseness. A pattern is concise if it contains relatively few attribute-value pairs, while a set of patterns is concise if it contains relatively few patterns. A concise pattern or set of patterns is relatively easy to understand and remember and thus is added more easily to the user's knowledge (set of beliefs). Accordingly, much research has been conducted to find a "minimum set of patterns," using properties such as monotonicity [Padmanabhan and Tuzhilin 2000] and confidence invariance [Bastide et al. 2000].

Generality / Coverage. A pattern is *general* if it covers a relatively large subset of a dataset. Generality (or coverage) measures the comprehensiveness of a pattern, that is, the fraction of all records in the dataset that matches the pattern. If a pattern characterizes more information in the dataset, it tends to be more interesting [Agrawal and Srikant 1994; Webb and Brain 2002]. Frequent itemsets are the most studied general patterns in the data mining literature. An itemset is a set of items, such as some items from a grocery basket. An itemset is *frequent* if its *support*, the fraction of records in the dataset containing the itemset, is above a given threshold [Agrawal and Srikant 1994]. The best known algorithm for finding frequent itemsets is the Apriori algorithm [Agrawal and Srikant 1994]. Some generality measures can form the bases for pruning strategies; for example, the support measure is used in the Apriori algorithm as the basis for pruning itemsets. For classification rules, Webb and Brain [2002] gave an empirical evaluation showing how generality affects classification results. Generality frequently coincides with conciseness because concise patterns tend to have greater coverage.

Reliability. A pattern is reliable if the relationship described by the pattern occurs in a high percentage of applicable cases. For example, a classification rule is reliable if its predictions are highly accurate, and an association rule is reliable if it has high confidence. Many measures from probability, statistics, and information retrieval have been proposed to measure the reliability of association rules [Ohsaki et al. 2004; Tan et al. 2002].

Peculiarity. A pattern is peculiar if it is far away from other discovered patterns according to some distance measure. Peculiar patterns are generated from peculiar data (or outliers), which are relatively few in number and significantly different from the rest of the data [Knorr et al. 2000; Zhong et al. 2003]. Peculiar patterns may be unknown to the user, hence interesting.

Diversity. A pattern is diverse if its elements differ significantly from each other, while a set of patterns is diverse if the patterns in the set differ significantly from each other. Diversity is a common factor for measuring the interestingness of summaries [Hilderman and Hamilton 2001]. According to a simple point of view, a summary can be considered diverse if its probability distribution is far from the uniform distribution. A diverse summary may be interesting because in the absence of any relevant knowledge, a user commonly assumes that the uniform distribution will hold in a summary. According to this reasoning, the more diverse the summary is, the more interesting it is. We are unaware of any existing research on using diversity to measure the interestingness of classification or association rules.

Novelty. A pattern is novel to a person if he or she did not know it before and is not able to infer it from other known patterns. No known data mining system represents everything that a user knows, and thus, novelty cannot be measured explicitly with reference to the user's knowledge. Similarly, no known data mining system represents what the user does not know, and therefore, novelty cannot be measured explicitly with reference to the user's ignorance. Instead, novelty is detected by having the user either explicitly identify a pattern as novel [Sahar 1999] or notice that a pattern cannot be deduced from and does not contradict previously discovered patterns. In the latter case, the discovered patterns are being used as an approximation to the user's knowledge.

Surprisingness. A pattern is surprising (or *unexpected*) if it contradicts a person's existing knowledge or expectations [Liu et al. 1997, 1999; Silberschatz and Tuzhilin 1995, 1996]. A pattern that is an exception to a more general pattern which has already been discovered can also be considered surprising [Bay and Pazzani 1999; Carvalho and Freitas 2000]. Surprising patterns are interesting because they identify failings in previous knowledge and may suggest an aspect of the data that needs further study. The difference between surprisingness and novelty is that a novel pattern is new and not contradicted by any pattern already known to the user, while a surprising pattern contradicts the user's previous knowledge or expectations.

Utility. A pattern is of utility if its use by a person contributes to reaching a goal. Different people may have divergent goals concerning the knowledge that can be extracted from a dataset. For example, one person may be interested in finding all sales with high profit in a transaction dataset, while another may be interested in finding all transactions with large increases in gross sales. This kind of interestingness is based on user-defined utility functions in addition to the raw data [Chan et al. 2003; Lu et al. 2001; Yao et al. 2004; Yao and Hamilton 2006].

Actionability/Applicability. A pattern is actionable (or *applicable*) in some domain if it enables decision making about future actions in this domain [Ling et al. 2002; Wang et al. 2002]. Actionability is sometimes associated with a pattern selection strategy. So far, no general method for measuring actionability has been devised. Existing measures depend on the applications. For example, Ling et al. [2002], measured accountability as the cost of changing the customer's current condition to match the objectives, whereas Wang et al. [2002], measured accountability as the profit that an association rule can bring.

The aforementioned interestingness criteria are sometimes correlated with, rather than independent of, one another. For example, Silberschatz and Tuzhilin [1996] argue that actionability may be a good approximation for surprisingness, and vice versa. As previously described, conciseness often coincides with generality, and generality often coincides with reduced sensitivity to noise, which is a form of reliability. Also, generality conflicts with peculiarity, while the latter may coincide with novelty.

These nine criteria can be further categorized into three classifications: *objective, subjective,* and *semantics-based*. An objective measure is based only on the raw data. No knowledge about the user or application is required. Most objective measures are based on theories in probability, statistics, or information theory. Conciseness, generality, reliability, peculiarity, and diversity depend only on the data and patterns, and thus can be considered objective.

A subjective measure takes into account both the data and the user of these data. To define a subjective measure, access to the user's domain or background knowledge about the data is required. This access can be obtained by interacting with the user during the data mining process or by explicitly representing the user's knowledge or expectations. In the latter case, the key issue is the representation of the user's knowledge, which has been addressed by various frameworks and procedures for data mining [Liu et al. 1997, 1999; Silberschatz and Tuzhilin 1995, 1996; Sahar 1999]. Novelty and surprisingness depend on the user of the patterns, as well as the data and patterns themselves, and hence can be considered subjective.

A semantic measure considers the semantics and explanations of the patterns. Because semantic measures involve domain knowledge from the user, some researchers consider them a special type of subjective measure [Yao et al. 2006]. Utility and actionability depend on the semantics of the data, and thus can be considered semantic. Utility-based measures, where the relevant semantics are the utilities of the patterns in the domain, are the most common type of semantic measure. To use a utility-based approach, the user must specify additional knowledge about the domain. Unlike subjective measures, where the domain knowledge is about the data itself and is usually represented in a format similar to that of the discovered pattern, the domain knowledge required for semantic measures does not relate to the user's knowledge or expectations concerning the data. Instead, it represents a utility function that reflects the user's goals. This function should be optimized in the mined results. For example, a store manager might prefer association rules that relate to high-profit items over those with higher statistical significance.

Having considered nine criteria for determining whether a pattern is interesting, let us now consider three methods for performing this determination, which we call *interestingness determination*. First, we can classify each pattern as either interesting or uninteresting. For example, we use the chi-square test to distinguish between interesting and uninteresting patterns. Secondly, we can determine a preference relation to represent that one pattern is more interesting than another. This method produces a partial ordering. Thirdly, we can rank the patterns. For the first or third approach, we can define an interestingness measure based on the aforementioned nine criteria and use this measure to distinguish between interesting and uninteresting patterns in the first approach or to rank patterns in the third approach.

Thus, using interestingness measures facilitates a general and practical approach to automatically identifying interesting patterns. In the remainder of this survey, we concentrate on this approach. The attempt to compare patterns classified as interesting by the interestingness measures to those classified as interesting by human subjects has rarely been tackled. Two recent studies have compared the ranking of rules by human experts to the ranking of rules by various interestingness measures, and suggested choosing the measure that produces the ranking which most resembles the ranking of experts [Ohsaki et al. 2004; Tan et al. 2002]. These studies were based on specific datasets and experts, and their results cannot be taken as general conclusions.

During the data mining process, interestingness measures can be used in three ways, which we call the *roles* of interestingness measures. Figure 1 shows these three roles. First, measures can be used to prune uninteresting patterns during the mining process so as to narrow the search space and thus improve mining efficiency. For example, a threshold for support can be used to filter out patterns with low support during the mining process and thus improve efficiency [Agrawal and Srikant 1994]. Similarly, for some utility-based measures, a utility threshold can be defined and used for pruning

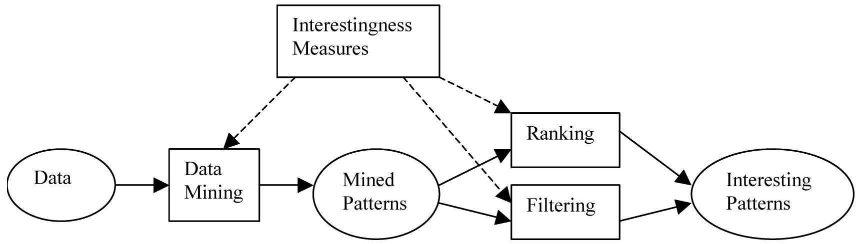

Fig. 1. Roles of interestingness measures in the data mining process.

patterns with low utility values [Yao et al. 2004]. Second, measures can be used to rank patterns according to the order of their interestingness scores. Third, measures can be used during postprocessing to select interesting patterns. For example, we can use the chi-square test to select all rules that have significant correlations after the data mining process [Bay and Pazzani 1999].

Researchers have proposed interestingness measures for various kinds of patterns, analyzed their theoretical properties, evaluated them empirically, and suggested strategies to select appropriate measures for particular domains and requirements. The most common patterns that can be evaluated by interestingness measures include association rules, classification rules, and summaries.

For the purpose of this survey, we categorize the measures as follows.

ASSOCIATION RULES/CLASSIFICATION RULES:

- *Objective Measures*
  - Based on probability (generality and reliability)
  - Based on the form of the rules
    - Peculiarity
    - Surprisingness
    - Conciseness
      - Nonredundant rules
      - Minimum description length
- *Subjective Measures*
  - Surprisingness
  - Novelty
- *Semantic Measures*
  - Utility
  - Actionability

SUMMARIES:

- *Objective Measures*
  - Diversity of summaries
  - Conciseness of summaries
  - Peculiarity of cells in summaries
- *Subjective Measures*
  - Surprisingness of summaries

McGarry [2005] recently made a comprehensive survey of interestingness measures for data mining. He described the measures in the context of the data mining process. Our article concentrates on both the interestingness measures themselves and the

Table II. Example Transaction Dataset

Milk	Bread	Eggs
1	0	1
1	1	0
1	1	1
1	1	1
0	0	1

analysis of their properties. We also provide a more comprehensive categorization of the measures. Furthermore, we analyze utility-based measures, as well as the measures for summaries, which were not covered by McGarry. Since the two articles survey research on interestingness from different perspectives, they can be considered complementary.

2. MEASURES FOR ASSOCIATION AND CLASSIFICATION RULES

In data mining research, most interestingness measures have been proposed for evaluating association and classification rules.

An association rule is defined in the following way [Agrawal and Srikant 1994]: Let $I = \{i_1, i_2, \ldots, i_m\}$ be a set of items. Let D be a set of transactions, where each transaction T is a set of items such that $T \subseteq I$. An association rule is an implication of the form $X \rightarrow Y$, where $X \subset I$, $Y \subset I$, and $X \cap Y = \phi$. The rule $X \rightarrow Y$ holds for the dataset D with support s and confidence c if $s\%$ of transactions in D contain $X \cup Y$ and $c\%$ of transactions in D that contain X also contain Y. In this article, we assume that the *support* and *confidence* measures yield fractions from [0, 1], rather than percentages. The support and confidence measures were the original interestingness measures proposed for association rules [Agrawal and Srikant 1994].

Suppose that D is the transaction table shown in Table II, which describes five transactions (rows) involving three items: milk, bread, and eggs. In the table, 1 signifies that an item occurs in the transaction and 0 means that it does not. The association rule ar_1: *Milk* $\rightarrow$ *Bread* can be mined from D. The support of this rule is 0.60 because the combination of milk and bread occurs in three out of five transactions, and the confidence is 0.75 because bread occurs in three of the four transactions that contain milk.

Recall that a classification rule is an implication of the form $X_1 \,\mathrm{op}\, x_1, X_2 \,\mathrm{op}\, x_2, \ldots, X_n \,\mathrm{op}\, x_n \rightarrow Y = y$, where X_i is a conditional attribute, x_i is a value that belongs to the domain of X_i, Y is the class attribute, y is a class value, and op is a relational operator such as = or >. The rule X_1 op x_1, X_2 op $x_2, \ldots, X_n$ op $x_n \rightarrow Y = y$ specifies that if an object satisfies the condition X_1 op x_1, X_2 op $x_2, \ldots, X_n$ op x_n, it can be classified into category y. Since a set of classification rules as a whole is used for the prediction of unseen data, the most common measure used to evaluate the quality of a set of classification rules is *predictive accuracy*, which is defined as

$$PreAcc = \frac{Number\ of\ testing\ examples\ correctly\ classified\ by\ the\ ruleset}{Total\ number\ of\ testing\ examples}.$$

In Table II, suppose that *Milk* and *Bread* are conditional attributes, *Eggs* is the class attribute, the first two tuples in the table are training examples, and the other tuples are testing examples. Suppose also that a ruleset is created which consists of the following two classification rules:

$$cr_1 : Bread = 0 \rightarrow Eggs = 1$$
$$cr_2 : Bread = 1 \rightarrow Eggs = 0$$

The predictive accuracy of the ruleset on the testing data is 0.33, since rule cr_1 gives one correct classification for the last testing example and rule cr_2 gives two incorrect classifications for the first two testing examples.

Although association and classification rules are both represented as if-then rules, we see five differences between them.

First, they have different purposes. Association rules are ordinarily used as descriptive tools. Classification rules, on the other hand, are used as a means of predicting classifications for unseen data.

Second, different techniques are used to mine these two types of rules. Association rule mining typically consists of two steps: (1) Finding frequent itemsets, that is, all itemsets with support greater than or equal to a threshold, and (2) generating association rules based on the frequent itemsets. Classification rule mining often consists of two different steps: (1) Using heuristics to select attribute-value pairs to use to form the conditions of rules, and (2) using pruning methods to avoid small disjuncts, that is, rules with antecedents that are too specific. The second pruning step is performed because although more specific rules tend to have higher accuracy on training data, they may not be reliable on unseen data, which is called *overfitting*. In some cases, classification rules are found by first constructing a tree (commonly called a *decision* tree), then pruning the tree, and finally generating the classification rules [Quinlan 1986].

Third, association rule mining algorithms often find many more rules than classification rule mining algorithms. An algorithm for association rule mining finds all rules that satisfy support and confidence requirements. Without postpruning and ranking, different algorithms for association rule mining find the same results. In contrast, most algorithms for classification rule mining find rules that together are sufficient to cover the training data, rather than finding all the rules that could be found for the dataset. Therefore, various algorithms for classification rules often find different rulesets.

Fourth, the algorithms for generating the two types of rules are evaluated differently. Since the results of association rule mining algorithms are the same, the running time and main memory used are the foremost issues for comparison. For classification rules, the comparison is based primarily on the predictive accuracy of the ruleset on testing data.

Fifth, the two types of rules are evaluated in different ways. Association rules are commonly evaluated by users, while classification rules are customarily evaluated by applying them to testing data.

Based on these differences between association and classification rules, interestingness measures play different roles in association and classification rule mining. In association rule mining, the user often needs to evaluate an overwhelming number of rules. Interestingness measures are very useful for filtering and ranking the rules presented to the user. In classification rule mining, interestingness measures can be used in two ways. First, they can be used during the induction process as heuristics to select attribute-value pairs for inclusion in classification rules. Second, they can be used to evaluate classification rules, similar to the way association rules are evaluated. However, the final evaluation of the results of classification rule mining is usually to measure the predictive accuracy of the whole ruleset on testing data because it is the ruleset, rather than a single rule, that determines the quality of prediction.

Despite the differences between association and classification rule mining, and the different roles the measures play in the two tasks, in this article, we survey the interestingness measures for these two kinds of rules together. When necessary, we identify which interestingness measures are used for each type of rule.

Table III. 2×2 Contingency for Rule $A \rightarrow B$

	B	$\overline{B}$	
A	$n(AB)$	$n(A\overline{B})$	$n(A)$
$\overline{A}$	$n(\overline{A}B)$	$n(\overline{A}\overline{B})$	$n(\overline{A})$
	$n(B)$	$n(\overline{B})$	N

2.1. Objective Measures for Association Rules or Classification Rules

In this section, we survey the objective interestingness measures for rules. We first describe measures based on probability in Section 2.1.1, the properties of such measures in Section 2.1.2, the strategies for selecting these measures in Section 2.1.3, and form-dependent measures in Section 2.1.4.

2.1.1. Objective Measures Based on Probability. Probability-based objective measures that evaluate the generality and reliability of association rules have been thoroughly studied by many researchers. They are usually functions of a 2×2 contingency table. A contingency table stores the frequency counts that satisfy given conditions. Table III is a contingency table for rule $A \rightarrow B$, where $n(AB)$ denotes the number of records satisfying both A and B, and N denotes the total number of records.

Table IV lists 38 common objective interestingness measures for association rules [Tan et al. 2002; Lenca et al. 2004; Ohsaki et al. 2004; Lavrac et al. 1999]. In the table, A and B represent the antecedent and consequent of a rule, respectively. $P(A) = \frac{n(A)}{N}$ denotes the probability of A; $P(B|A) = \frac{P(AB)}{P(A)}$ denotes the conditional probability of B, given A. These measures originate from various areas, such as statistics (correlation coefficient, odds ratio, Yule's Q, and Yule's Y), information theory (J-measure and mutual information), and information retrieval (accuracy and sensitivity/recall).

Given an association rule $A \rightarrow B$, the two main interestingness criteria for this rule are generality and reliability. *Support* $P(AB)$ or *coverage* $P(A)$ is used to represent the generality of the rule. *Confidence* $P(B|A)$ or a correlation factor such as the *added value* $P(B|A) - P(B)$ or *lift* $P(B/A)/P(B)$ is used to represent the reliability of the rule.

Some researchers have suggested that a good interestingness measure should include both generality and reliability. For example, Tan et al. [2000] proposed the IS measure: $IS = \sqrt{I \times support}$, where $I = \frac{P(AB)}{P(A)P(B)}$ is the ratio between the joint probability of two variables with respect to their expected probability under the independence assumption. This measure also represents the cosine angle between A and B. Lavrac et al. [1999] proposed weighted relative accuracy: $WRAcc = P(A)(P(B|A) - P(B))$. This measure combines the coverage $P(A)$ and the added value $P(B|A) - P(B)$. This measure is identical to Piatetsky-Shapiro's measure: $P(AB) - P(A)P(B)$ [Piatetsky-Shapiro 1991]. Other measures involving these two criteria include Yao and Liu's two-way support [Yao and Zhong 1999], Jaccard [Tan et al. 2002], Gray and Orlowska's interestingness weighting dependency [Gray and Orlowska 1998], and Klosgen's measure [Klosgen 1996]. All these measure combine either support $P(AB)$ or coverage $P(A)$ with a correlation factor of either $(P(B|A) - P(B)$ or lift $P(B/A)/P(B))$.

Tan et al. [2000] referred to a measure that includes both support and a correlation factor as an *appropriate* measure. They argued that any appropriate measure can be used to rank discovered patterns, and they also showed that the behaviors of such measures, especially where support is low, are similar.

Bayardo and Agrawal [1999] studied the relationships between support, confidence, and other measures from another angle. They defined a partial ordered relation based on support and confidence, as follows. For rules r_1 and r_2, if $support(r_1) \leq support(r_2)$ and $confidence(r_1) \leq confidence(r_2)$, we have $r_1 \leq_{sc} r_2$. Any rule r in the upper border,

Table IV. Probability Based Objective Interestingness Measures for Rules

Measure	Formula
Support	$P(AB)$
Confidence/Precision	$P(B\mid A)$
Coverage	$P(A)$
Prevalence	$P(B)$
Recall	$P(A\mid B)$
Specificity	$P(\neg B\mid\neg A)$
Accuracy	$P(AB) + P(\neg A \neg B)$
Lift/Interest	$P(B\mid A)/P(B)$ or $P(AB)/P(A)P(B)$
Leverage	$P(B\mid A) - P(A)P(B)$
Added Value/Change of Support	$P(B\mid A) - P(B)$
Relative Risk	$P(B\mid A)/P(B\mid\neg A)$
Jaccard	$P(AB)/(P(A) + P(B) - P(AB))$
Certainty Factor	$(P(B\mid A) - P(B))/(1 - P(B))$,
Odds Ratio	$\frac{P(AB)P(\neg A\neg B)}{P(A\neg B)P(\neg BA)}$
Yule's Q	$\frac{P(AB)P(\neg A\neg B) - P(A\neg B)P(\neg AB)}{P(AB)P(\neg A\neg B) + P(A\neg B)P(\neg AB)}$
Yule's Y	$\frac{\sqrt{P(AB)P(\neg A\neg B)} - \sqrt{P(A\neg B)P(\neg AB)}}{\sqrt{P(AB)P(\neg A\neg B)} + \sqrt{P(A\neg B)P(\neg AB)}}$
Klosgen	$\sqrt{P(AB)}(P(B\mid A) - P(B))$, $\sqrt{P(AB)}\max(P(B\mid A) - P(B), P(A\mid B) - P(A))$
Conviction	$\frac{P(A)P(\neg B)}{P(A\neg B)}$
Interestingness Weighting Dependency	$((\frac{P(AB))}{P(A)P(B)})^k - 1) * P(AB)^m$,where k, m are coefficients of dependency and generality, respectively, weighting the relative importance of the two factors.
Collective Strength	$\frac{P(AB)+P(\neg B\mid\neg A)}{P(A)P(B)+P(\neg A)*P(\neg B)} * \frac{1-P(A)P(B)-P(\neg A)*P(\neg B)}{1-P(AB)-P(\neg B\mid\neg A)}$
Laplace Correction	$\frac{N(AB)+1}{N(A)+2}$
Gini Index	$P(A) * \{P(B\mid A)^2 + P(\neg B\mid A)^2\} + P(\neg A) * \{P(B\mid\neg A)^2$ $+P(\neg B\mid\neg A)^2\} - P(B)^2 - P(\neg B)^2$
Goodman and Kruskal	$\frac{\sum_i \max_j P(A_i B_j)+\sum_j \max_i P(A_i B_j)-\max_i P(A_i)-\max_i P(B_j)}{2-\max_i P(A_i)-\max_i P(B_j)}$
Normalized Mutual Information	$\sum_i \sum_j P(A_i B_j) * \log_2 \frac{P(A_i B_j)}{P(A_i)P(B_j)} / \{-\sum_i P(A_i) * \log_2 P(A_i)\}$
J-Measure	$P(AB)\log(\frac{P(B\mid A)}{P(B)}) + P(A\neg B)\log(\frac{P(\neg B\mid A)}{P(\neg B)})$
One-Way Support	$P(B\mid A) * \log_2 \frac{P(AB)}{P(A)P(B)}$
Two-Way Support	$P(AB) * \log_2 \frac{P(AB)}{P(A)P(B)}$
Two-Way Support Variation	$P(AB) * \log_2 \frac{P(AB)}{P(A)P(B)} + P(A\neg B) * \log_2 \frac{P(A\neg B)}{P(A)P(\neg B)} +$ $P(\neg AB) * \log_2 \frac{P(\neg AB)}{P(\neg A)P(B)} + P(\neg A\neg B) * \log_2 \frac{P(\neg A\neg B)}{P(\neg A)P(\neg B)}$
$\emptyset$−Coefficient (Linear Correlation Coefficient)	$\frac{P(AB)-P(A)P(B)}{\sqrt{P(A)P(B)P(\neg A)P(\neg B)}}$
Piatetsky-Shapiro	$P(AB) - P(A)P(B)$
Cosine	$\frac{P(AB)}{\sqrt{P(A)P(B)}}$
Loevinger	$1 - \frac{P(A)P(\neg B)}{P(A\neg B)}$
Information Gain	$\log \frac{P(AB)}{P(A)P(B)}$
Sebag-Schoenauer	$\frac{P(AB)}{P(A\neg B)}$
Least Contradiction	$\frac{P(AB)-P(A\neg B)}{P(B)}$
Odd Multiplier	$\frac{P(AB)P(\neg B)}{P(B)P(A\neg B)}$
Example and Counterexample Rate	$1 - \frac{P(A\neg B)}{P(AB)}$
Zhang	$\frac{P(AB)-P(A)P(B)}{\max(P(AB)P(\neg B),P(B)P(A\neg B))}$

for which there is no r' such that $r \leq_{sc} r'$, is called an *sc*-optimal rule. For a measure that is monotone in both support and confidence, the most interesting rule is an *sc*-optimal rule. For example, the Laplace measure $\frac{n(AB)+1}{n(A)+2}$ can be transformed to $\frac{N \times support(A\to B)+1}{N \times support(A\to B)/confidence(A\to B)+2}$. Since N is a constant, the Laplace measure can be considered a function of *support*($A\to B$) and *confidence* ($A\to B$). It is easy to show that the Laplace measure is monotone in both support and confidence. This property is useful when the user is only interested in the single most interesting rule, since we only need to check the *sc*-optimal ruleset, which contains fewer rules than the entire ruleset.

Yao et al. [2006] identified a fundamental relationship between preference relations and interestingness measures for association rules: There exists a real valued interestingness measure that reflects a preference relation if and only if the preference relation is a *weak order*. A weak order is a relation that is *asymmetric* (i.e., $R_1 \succ R_2 \Rightarrow \neg R_2 \succ R_1$) and *negative-transitive* (i.e., $\neg R_1 \succ R_2 \wedge \neg R_2 \succ R_3 \Rightarrow \neg R_1 \succ R_3$). It is a special type of partial order and more general than a total order. Other researchers studied more general forms of interestingness measures. Jaroszewicz and Simovici [2001] proposed a general measure based on distribution divergence. The chi-square, Gini, and entropy-gain measures can be obtained from this measure by setting different parameters.

For classification rules, the most important role of probability-based interestingness measures in the mining process is to act as heuristics to choose the attribute-value pairs for inclusion. In this context, these measures are also called *feature-selection* measures [Murthy 1998]. In the induction process, two factors should be considered. First, a rule should have a high degree of accuracy on the training data. Second, the rule should not be too specific, covering only a few examples, and thus overfitting. A good measure should optimize these two factors. Precision (corresponding to confidence in association rule mining) [Pagallo and Haussler 1990], entropy [Quinlan 1986], Gini [Breiman et al. 1984], and Laplace [Clark and Boswell 1991] are the most widely used measures for selecting attribute-value pairs. Fürnkranz and Flach [2005] proved that the entropy and Gini measures are equivalent to precision, in the sense that they give either identical or reverse rankings for any ruleset. Clark and Boswell [1991] argued that the Laplace measure is biased towards more general rules with higher predictive accuracy than entropy, which is supported by their experimental results with CN2. Two comprehensive surveys of feature-selection measures for classification rules and decision trees are given in Murthy [1998] and Fürnkranz and Flach [2005].

All probability-based objective interestingness measures proposed for association rules can also be applied directly to classification rule evaluation, since they only involve the probabilities of the antecedent of a rule, the consequent of a rule, or both, and they represent the generality, correlation, and reliability between the antecedent and consequent. However, when these measures are used in this way, they assess the interestingness of the rule with respect to the given data (the training dataset), whereas the key focus in classification rule mining is on predictive accuracy.

In this survey, we do not elaborate on the individual measures. Instead, we emphasize the properties of these measures and discuss how to analyze and choose from among them for data mining applications.

2.1.2. Properties of Probability Based Objective Measures. Many objective measures have been proposed for different applications. To analyze these measures, some properties for the measures have been proposed. We consider three sets of properties that have been described in the literature.

Piatetsky-Shapiro [1991] proposed three principles that should be obeyed by any objective measure, F:

(P1) $F = 0$ if A and B are statistically independent, that is, $P(AB) = P(A)P(B)$.

(P2) F monotonically increases with $P(AB)$ when $P(A)$ and $P(B)$ remain the same.

(P3) F monotonically decreases with $P(A)$ (or $P(B)$) when $P(AB)$ and $P(B)$ (or $P(A)$) remain the same.

Principle (P1) states that an association rule which occurs by chance has zero interest value, that is, it is not interesting. In practice, this principle may seem too rigid. For example, the lift measure attains a value of 1 rather than 0 in the case of independent attributes, which corresponds to an association rule occurring by chance. A value greater than 1 indicates a positive correlation, and a value less than 1 indicates a negative correlation. To relax Principle (P1), some researchers propose a constant value for the independent situations [Tan et al. 2002]. Principle (P2) states that the greater the support for AB, the greater the interestingness value when the support for A and B is fixed, that is, the more positive correlation A and B have, the more interesting the rule. Principle (P3) states that if the supports for AB and B (or A) are fixed, the smaller the support for A (or B), the more interesting the pattern. According to Principles (P2) and (P3), when the covers of A and B are identical or the cover of A contains the cover of B (or vice versa), the interestingness measure should attain its maximum value.

Tan et al. [2002] proposed five properties based on operations for 2×2 contingency tables.

(O1) F should be symmetric under variable permutation.

(O2) F should be the same when we scale any row or column by a positive factor.

(O3) F should become $-F$ if either the rows or the columns are permuted, that is, swapping either the rows or columns in the contingency table makes interestingness values change their signs.

(O4) F should remain the same if both the rows and columns are permuted.

(O5) F should have no relationship with the count of the records that do not contain A and B.

Unlike Piatetsky-Shapiro's principles, these properties should not be interpreted as statements of what is desirable. Instead, they can be used to classify the measures into different groups. Property (O1) states that rules $A \rightarrow B$ and $B \rightarrow A$ should have the same interestingness values, which is not true for many applications. For example, confidence represents the probability of a consequent, given the antecedent, but not vice versa. Thus, it is an asymmetric measure. To provide additional symmetric measures, Tan et al. [2002] transformed each asymmetric measure F into a symmetric one by taking the maximum value of $F(A \rightarrow B)$ and $F(B \rightarrow A)$. For example, they defined a symmetric confidence measure as $\max(P(B|A), P(A|B))$. Property (O2) requires invariance with the scaling of rows or columns. Property (O3) states that $F(A \rightarrow B) = -F(A \rightarrow \neg B) = -F(\neg A \rightarrow B)$. This property means that the measure can identify both positive and negative correlations. Property (O4) states that $F(A \rightarrow B) = F(\neg A \rightarrow \neg B)$. Property (O3) is in fact a special case of Property (O4) because if permuting the rows (columns) causes the sign to change once and permuting the columns (rows) causes it to change again, the overall result of permuting both rows and columns will be to leave the sign unchanged. Property (O5) states that the measure should only take into account the number of records containing A, B, or both. Support does not satisfy this property, while confidence does.

Lenca et al. [2004] proposed five properties to evaluate association measures

(Q1) F is constant if there is no counterexample to the rule.

(Q2) F decreases with $P(A\neg B)$ in a linear, concave, or convex fashion around 0+.

(Q3) F increases as the total number of records increases.

(Q4) The threshold is easy to fix.

(Q5) The semantics of the measure are easy to express.

Lenca et al. claimed that Properties (Q1), (Q4), and (Q5) are desirable for measures, but that Properties (Q2) and (Q3) may or may not be desired by users. Property (Q1) states that rules with a confidence of 1 should have the same interestingness value, regardless of the support, which contradicts the suggestion of Tan et al. [2002] that a measure should combine support and association aspects. Property (Q2) describes the manner in which the interestingness value decreases as a few counterexamples are added. If the user can tolerate a few counterexamples, a concave decrease is desirable. If the system strictly requires a confidence of 1, a convex decrease is desirable.

In Table V, we indicate which properties hold for each of the measures listed in Table IV. For property (Q2), to simplify analysis, we assume the total number of records is fixed. When the number of records that match $A \neg B$ increases, the numbers that match AB decreases correspondingly. In Table V, we use 0, 1, 2, 3, 4, 5, and 6 to represent convex decreasing, linear decreasing, concave decreasing, invariant increasing, not applicable, and depending on parameters, respectively. We see that 32 measures decrease with the number of exceptions, and 23 measures both decrease with the number of exceptions and increase with support. *Loevinger* is the only measure that increases with the number of exceptions.

Property (Q3) describes the changes to the interestingness values that occur as the number of records in the dataset is increased, assuming that $P(A)$, $P(B)$, and $P(AB)$ are held constant. Property (Q4) states that when a threshold for an interestingness measure is used to separate interesting from uninteresting rules, the threshold should be easy to choose and the semantics easily expressed. Property (Q5) states that the semantics of the interestingness measure is understandable to the user.

To quantify the relationships between an appropriate interestingness measure and support and confidence as described in Section 2.1.1, here we propose two desirable properties for a measure F that is intended to measure the interestingness of association rules:

(S1) F should be an increasing function of support if the margins in the contingency table are fixed.

(S2) F should be an increasing function of confidence if the margins in the contingency table are fixed.

For property (S1), we assume the margins in the contingency table are constant, that is, we assume $n(A) = a$, $n(\neg A) = N - a$, $n(B) = b$, and $n(\neg B) = N - b$. If we represent the *support* by x, then we have $P(AB) = x$, $P(\neg AB) = \frac{b}{N} - x$, $P(A \neg B) = \frac{a}{N} - x$, and $P(\neg A \neg B) = 1 - \frac{a+b}{N} + x$. By substituting these formulas in the measures, we obtain functions of the measures with the support x as a variable. For example, consider lift, which is defined as $lift = \frac{P(AB)}{P(A)P(B)} = \frac{x}{\frac{a}{n} \times \frac{b}{n}}$. Clearly, lift is an increasing function of the support x. In a similar fashion, we can determine the results for other measures, which are shown in Table V. We use 0, 1, 2, 3, and 4 to represent increasing with support, invariant with support, decreasing with support, not applicable, and depending on

Table V. Properties of Probability-Based Objective Interestingness Measures for Rules

Measure	P1	P2	P3	O1	O2	O3	O4	O5	Q1	Q2	Q3	S1
Support	N	Y	N	Y	N	N	N	N	N	1	N	0
Confidence/Precision	N	Y	N	N	N	N	N	N	Y	1	N	0
Coverage	N	N	N	N	N	N	N	N	N	3	N	1
Prevalence	N	N	N	N	N	N	N	N	N	1	N	1
Recall	N	Y	N	N	N	N	N	Y	N	2	N	0
Specificity	N	N	N	N	N	N	N	N	N	3	N	0
Accuracy	N	Y	Y	Y	N	N	Y	N	N	1	N	1
Lift/Interest	N	Y	Y	Y	N	N	N	N	N	2	N	0
Leverage	N	Y	Y	N	N	N	N	Y	N	1	N	0
Added Value	Y	Y	Y	N	N	N	N	N	N	1	N	0
Relative Risk	N	Y	Y	N	N	N	N	N	N	1	N	0
Jaccard	N	Y	Y	Y	N	N	N	Y	N	1	N	0
Certainty Factor	Y	Y	Y	N	N	N	Y	N	N	0	N	0
Odds ratio	N	Y	Y	Y	Y	Y	Y	N	Y	0	N	4
Yule's Q	Y	Y	Y	Y	Y	Y	Y	N	Y	0	N	4
Yule's Y	Y	Y	Y	Y	Y	Y	Y	N	Y	0	N	4
Klosgen	Y	Y	Y	N	N	N	N	N	N	0	N	0
Conviction	N	Y	N	N	N	N	Y	N	Y	0	N	0
Interestingness Weighting Dependency	N	Y	N	N	N	N	N	Y	N	6	N	0
Collective Strength	N	Y	Y	Y	N	Y	Y	N	N	0	N	0
Laplace Correction	N	Y	N	N	N	N	N	N	N	1	N	0
Gini Index	Y	N	N	N	N	N	Y	N	N	0	N	4
Goodman and Kruskal	Y	N	N	Y	N	N	Y	N	N	5	N	3
Normalized Mutual Information	Y	Y	Y	N	N	N	Y	N	N	5	N	3
J-Measure	Y	N	N	N	N	N	N	N	Y	0	N	4
One-Way Support	Y	Y	Y	N	N	N	N	Y	N	0	N	0
Two-Way Support	Y	Y	Y	Y	N	N	N	Y	N	0	N	0
Two-Way Support Variation	Y	N	N	Y	N	N	Y	N	N	0	N	4
$\phi-$Coefficient (Linear Correlation Coefficient)	Y	Y	Y	Y	N	Y	Y	N	N	0	N	0
Piatetsky-Shapiro	Y	Y	Y	Y	N	Y	Y	N	N	1	N	0
Cosine	N	Y	Y	Y	N	N	N	Y	N	2	N	0
Loevinger	Y	Y	N	N	N	N	N	N	Y	4	N	2
Information gain	Y	Y	Y	Y	N	N	N	Y	N	2	N	0
Sebag-Schoenauer	N	Y	Y	N	N	N	N	Y	Y	0	N	0
Least Contradiction	N	Y	Y	N	N	N	N	Y	N	2	N	0
Odd Multiplier	N	Y	Y	N	N	N	N	N	Y	0	N	0
Example and Counterexample Rate	N	Y	Y	N	N	N	N	Y	Y	2	N	0
Zhang	Y	N	N	N	N	N	N	N	N	0	N	4

Table VI. Analysis Methods for Objective Association Rule Interestingness Measures

Analysis Method	Based on Properties	Based on Data Sets
Ranking	Lenca et al. [2004]	Tan et al. [2002]
Clustering	Vaillant et al. [2004]	Vaillant et al. [2004]

parameters, respectively. Assuming the margins are fixed, 25 measures increase with support. Only the *Loevinger* measure decreases with support.

Property (S2) is closely related to property (Q2), albeit in an inverse manner, because if a measure decreases with $P(A\neg B)$, it increases with $P(AB)$. However, property (Q2) describes the relationship between measure F and $P(A\neg B)$, without constraining the other parameters $P(AB)$, $P(\neg AB)$, and $P(\neg A\neg B)$. This lack of constraint makes analysis difficult. With property (S2), constraints are applied to the margins of the contingency tables, which facilitates formal analysis.

2.1.3. Selection Strategies for Probability-Based Objective Measures. Due to the overwhelming number of interestingness measures shown in Table V, the means of selecting an appropriate measure for a given application is an important issue. So far, two methods have been proposed for comparing and analyzing the measures, namely, ranking and clustering. Analysis can be conducted based on either the properties of the measures or empirical evaluations on datasets. Table VI classifies the studies that are summarized here.

Tan et al. [2002] proposed a method to rank measures based on a specific dataset. In this method, the user is first required to rank a set of mined patterns, and the measure that has the most similar ranking results for these patterns is selected for further use. This method is not directly applicable if the number of patterns is overwhelming. Instead, this method selects the patterns that have the greatest standard deviations in their rankings by the measures. Since these patterns cause the greatest conflict among the measures, they should be presented to the user for ranking. The method then selects the measure that gives rankings most consistent with the manual ranking. This method is based on the specific dataset and needs the user's involvement.

Another method to select the appropriate measure is based on the multicriteria decision aid [Lenca et al. 2004]. In this approach, marks and weights are assigned to each property that the user considers to be of importance. For example, if a symmetric property is desired, a measure is assigned a 1 if it is symmetric, and 0 if it is asymmetric. With each row representing a measure and each column representing a property, a decision matrix is created. An entry in the matrix represents the mark for the measure according to the property. Applying the multicriteria decision process on the table, we can obtain a ranking of results. With this method, the user is not required to rank the mined patterns. Rather, he or she must identify the desired properties and specify their significance for a particular application.

An additional method for analyzing measures is to cluster the interestingness measures into groups [Vaillant et al. 2004]. As with the ranking method, this clustering method can be based on either the properties of the measures or the rulesets generated by experiments on datasets. *Property-based clustering*, which groups measures based on the similarity of their properties, works on a decision matrix with each row representing a measure, and each column representing a property. *Experiment-based clustering* works on a matrix with each row representing a measure and each column signifying a measure applied to a ruleset. Each entry represents a similarity value between the two measures on the specified ruleset. Similarity is calculated on the rankings of the two measures on the ruleset. Vaillant et al. [2004] showed consistent results using the two clustering methods with 20 measures on 10 rulesets.

2.1.4. Form-Dependent Objective Measures. A form-dependent measure is an objective measure that is based on the form of the rules. We consider form-dependent measures based on peculiarity, surprisingness, and conciseness.

The *neighborhood-based unexpectedness* measure for association rules [Dong and Li 1998] is based on peculiarity. The intuition for this method is that if a rule has a different consequent from neighboring rules, it is interesting. The distance $Dist(R_{1,} R_2)$ between two rules, R_1: $X_1 \rightarrow Y_1$ and R_2: $X_2 \rightarrow Y_2$, is defined as $Dist(R_{1,} R_2) = \delta_1|X_1Y_1 - X_2Y_2| + \delta_2|X_1 - X_2| + \delta_3|Y_1 - Y_2|$, where $X - Y$ denotes the symmetric difference between X and Y, $|X|$ denotes the cardinality of X, and δ_1, δ_2, and δ_3 are weights determined by the user. Based on this distance, the r-neighborhood of rule R_0, denoted as $N(R_0, r)$, is defined as $\{R$: Dist$(R, R_0) \leq r$, R is a potential rule$\}$, where $r > 0$ is the *radius* of the neighborhood. Dong and Li [1998] then proposed two interestingness measures. The first is called *unexpected confidence*: If the confidence of a rule r_0 is far from the average confidence of the rules in its neighborhood, this rule is interesting. Another measure is based on the sparsity of neighborhood, that is, if the number of mined rules in the neighborhood is far less than that of all potential rules in the neighborhood, it is considered interesting. This measure can be applied to classification rule evaluation if a distance function for the classification rules is defined.

Another form-dependent measure is called surprisingness, which is defined for classification rules. As described in Section 2.2, many researchers use subjective interestingness measures to represent the surprisingness of classification rules. Taking a different perspective, Freitas [1998] defined two objective interestingness measures for this purpose, on the basis of the form of the rules.

The first measure defind by Freitas [1998] is based on the generalization of the rule. Suppose there is a classification rule $A_1, A_2, \ldots, A_m \rightarrow C$. When we remove one of the conditions, say A_1, from the antecedent, the resulting antecedent $A_2, \ldots, A_m$ is more general than $A_1, A_2, \ldots, A_m$. Assume that when applied to the dataset, this antecedent predicts consequent C_1. We obtain the rule $A_2, \ldots, A_m \rightarrow C_1$, which is more general than $A_1, A_2, \ldots, A_m \rightarrow C$. If $C_1 = C$, we count 1, otherwise we count 0. Then, we do the same for each of $A_2, \ldots, A_m$ and count the sum of the times C_i differs from C. The result, an integer in the interval [0, m], is defined as the *raw surprisingness* of the rule, denoted as $Surp_{raw}$. *Normalized surprisingness* $Surp_{norm}$, defined as $Surp_{raw}/m$, takes on real values in the interval [0, 1]. If all the classes that the generalized rules predict are different from the original class C, $Surp_{norm}$ takes on a value of 1, which means the rule is most interesting. If all classes that the generalized rules predict are the same as C, $Surp_{norm}$ takes on a value of 0, which means that the rule is not interesting at all, since all of its generalized forms make the same prediction. This method can be regarded as neighborhood-based, where the neighborhood of a rule R is the set of rules with one condition removed from R.

Freitas' [1998] second measure is based on information gain, defined as the reciprocal of the average information gain for all the condition attributes in a rule. It is based on the assumption that a larger information gain indicates a better attribute for classification. The user may be more aware of it and consequently, the rules containing these attributes may be of less interest. This measure is biased towards the rules that have less than the average information gain for all their condition attributes.

These two measures cannot be applied to association rules unless all of them have only one item in the consequent.

Conciseness, a form-dependent measure, is often used for rulesets rather than single rules. We consider two methods for evaluating the conciseness of rules. The first is based on logical redundancy [Padmanabhan and Tuzhilin 2000; Bastide et al. 2000; Li and Hamilton 2004]. In this method, no measure is defined for conciseness; rather, algorithms are designed to find nonredundant rules. For example, Li and Hamilton

[2004] proposed both an algorithm to find a minimum ruleset and an inference system. The set of association rules discovered by the algorithm is minimum in that no redundant rules are present. All other association rules that satisfy confidence and support constraints can be derived from this ruleset using the inference system. This method is proposed for association rules with two-valued condition attributes, and is not suitable for classification rules with multivalued condition attributes.

The second method to evaluate the conciseness of a ruleset is called the minimum description-length (MDL) principle. It takes into account both the complexity and the accuracy of the theory (ruleset, in this context). The first part of the MDL measure, $L(H)$, is called the *theory cost*, which measures the theory complexity, where H is a theory. The second part, $L(D|H)$, measures the degree to which the theory fails to account for the data, where D denotes the data. For a group of theories (rulesets), a more complex theory tends to fit the data better than a simpler one, and therefore, the former has a higher $L(H)$ value and a smaller $L(D|H)$ value. The theory with the shortest description-length has the best balance between these two factors and is preferred. Detailed MDL measures for classification rules and decision trees can be found in Forsyth et al. [1994] and Vitanyi and Li [2000]. The MDL principle has been applied to evaluate both classification and association rulesets.

Objective interestingness measures indicate the support and degree of correlation of a pattern for a given dataset. However, they do not take into account the knowledge of the user who uses the data.

2.2. Subjective Interestingness Measures

In applications where the user has background knowledge, patterns ranked highly by objective measures may not be interesting. A subjective interestingness measure takes into account both the data and the user's knowledge. Such a measure is appropriate when: (1) The background knowledge of users varies, (2) the interests of the users vary, and (3) the background knowledge of users evolve. Unlike the objective measures considered in the previous section, subjective measures may not be representable by simple mathematical formulas because the user's knowledge may be represented in various forms. Instead, they are usually incorporated into the mining process. As mentioned previously, subjective measures are based on the surprisingness and novelty criteria. In this context, previous researchers have used the term *unexpectedness* rather than *surprisingness*, so we have adopted the same term.

2.2.1. Unexpectedness and Novelty. To find unexpected or novel patterns in data, three approaches can be distinguished based on the roles of unexpectedness measures in the mining process: (1) the user provides a formal specification of his or her knowledge, and after obtaining the mining results, the system chooses which unexpected patterns to present to the user [Liu et al. 1997, 1999; Silberschatz and Tuzhilin 1995, 1996]; (2) according to the user's interactive feedback, the system removes uninteresting patterns [Sahar 1999]; and (3) the system applies the user's specifications as constraints during the mining process to narrow down the search space and provide fewer results [Padmanabhan and Tuzhilin 1998]. Let us consider each of these approaches in turn.

2.2.2. Using Interestingness Measures to Filter Interesting Patterns from Mined Results. Silberschatz and Tuzhilin [1996] related unexpectedness to a belief system. To define beliefs, they used arbitrary predicate formulae in first-order logic, rather than if-then rules. They also classified beliefs as either *hard* or *soft*. A hard belief is a constraint that cannot be changed with new evidence. If the evidence (rules mined from data) contradicts hard beliefs, a mistake is assumed to have been made in acquiring the evidence. A soft belief is one that the user is willing to change as new patterns are discovered. The authors adopted a Bayesian approach and assumed that the degree of belief is measured with conditional probability. Given evidence E (patterns), the degree of belief in α is updated with Bayes' rule as follows:

$$P(\alpha|E,\xi) = \frac{P(E|\alpha,\xi)P(\alpha|\xi)}{P(E|\alpha,\xi)P(\alpha|\xi) + P(E|\neg\alpha,\xi)P(\neg\alpha|\xi)},$$

where ξ is the context representing the previous evidence supporting α. Then, the interestingness measure for pattern p, relative to a soft belief system B, is defined as the relative difference by the prior and posterior probabilities:

$$I(p,B) = \sum_{\alpha \in B} \frac{|P(\alpha|p,\xi) - P(\alpha|\xi)|}{P(\alpha|\xi)}.$$

Silberschatz and Tuzhilin [1996] presented a general framework for defining an interestingness measure for patterns. Let us consider how this framework can be applied to patterns in the form of association rules. For the example in Table II, we define the belief α as "people buy milk, eggs, and bread together." Here, ξ denotes the dataset D. Initially, suppose the user specifies the degree of belief in α as $P(\alpha|\xi) = 2/5 = 0.4$, based on the dataset, since two out of five transactions support belief α. Similarly, $P(\neg\alpha|\,\xi) = 0.6$. Suppose a pattern is mined in the form of an association rule p: milk→eggs with support = 0.4 and confidence = 2 / 3 ≈ 0.67. The new degree of belief in α, based on the new evidence p in the context of the old evidence ξ, is denoted $P(\alpha|\,p,\,\xi)$. It can be computed with Bayes' rule, as given previously, if we know the values of the $P(\alpha|\,\xi)$, $P(\neg\alpha|\,\xi)$, $P(p|\neg\alpha,\xi)$, and $P(p\,|\neg\alpha,\xi)$ terms. The values of the first two terms have already been calculated.

The other two terms can be computed as follows. The $P(p\,|\,\alpha,\,\xi)$ term represents the confidence of rule p, given belief α, that is, the confidence of the rule milk→eggs evaluated on transactions 3 and 4, where milk, eggs, and bread appear together. From Table II, we obtain $P(p\,|\,\alpha,\,\xi) = 1$. Similarly, the term $P(p\,|\,\neg\alpha,\,\xi)$ represents the confidence of rule p, given belief $\neg\alpha$, that is, the confidence of the rule milk→eggs evaluated on transactions 1, 2, and 5, where milk, eggs, and bread do not appear together. From Table II, we obtain $P(p\,|\,\neg\alpha,\,\xi) = 0.5$.

Using Bayes' rule, we calculate $P(\alpha|E,\xi) = \frac{1\times 0.4}{1\times 0.4 + 0.5\times 0.6} \approx 0.57$, and accordingly, the value of interestingness measure I for rule p is calculated as $I(p,B) = \frac{|0.57-0.4|}{0.4} \approx 0.43$.

To rank classification rules according to the user's existing knowledge, Liu et al. [1997] proposed two kinds of specifications (T1 and T2) for defining the user's vague knowledge, called *general impressions*. A general impression of type T1 can express a positive or negative relation between a condition variable and a class, a relation between a range (or subset) of values of condition variables and a class, or the vague impression that a relation exists between a condition variable and a class. T2 extends T1 by separating the user's knowledge into a *core* and *supplement*. The core refers to the user's knowledge that can be clearly represented and the supplement refers to the user's knowledge that can only be vaguely represented. The core and supplement are both soft beliefs because they may not be true, and thus need to be either verified or contradicted. Based on these two kinds of specifications, matching algorithms were proposed for obtaining confirming rules (called *conforming rules* in Liu et al. [1997]), and unexpected consequent rules and unexpected condition rules. These rules are ranked by the degree to which they match using interestingness measures. In the matching process for a rule R, the general impressions are separated into two sets: G_S and G_D. The set G_S consists of all general impressions with the same consequent as R, and G_D consists of all general impressions with different consequents from R. In the confirming rule case, R is matched with G_S. The interestingness measure calculates the similarity between the conditions of R and the conditions of G_S. In the unexpected consequent rule case, R is matched with G_D. The interestingness measure determines the similarity between the conditions of R and G_D. In the unexpected condition rule case, R is again matched with G_S, and the interestingness measure calculates the difference between the conditions of R and G_S. Thus, the rankings of unexpected condition rules are the reverse of those of confirming rules.

Let us use an example to illustrate the calculation of interestingness values for confirming rules using type T1 specifications. Assume we have discovered a classification rule r, and we want to use it to confirm the user's general impressions:

$$r\text{: jobless} = \text{no, saving} > 10{,}000 \rightarrow \text{approved},$$

which states that if a person is not jobless and his savings are more that $10,000, his loan will be approved.

Assume the user provides the following five general impressions:

(G1) saving> → approved

(G2) age | → {approved, not_approved}

(G3) jobless{no} → approved

(G4) jobless{yes} → not_approved

(G5) saving>, jobless{yes} → approved

General impression (G1) states that if an applicant's savings are large, the loan will be approved. Impression (G2) states that an applicant's age relates in an unspecified way to the result of his loan application, and (G3) states that if an applicant has a job, the loan will be approved. Impression (G4) states that if an applicant is jobless, the loan will not be approved, while (G5) states that if an applicant's savings are large and the applicant is jobless, the loan will be approved. Here, G_S is {(G4)}, and G_D is {(G1), (G2), (G3), (G5)}.

Since we want to use rule r to confirm these general impressions, we only consider (G1), (G2), (G3), and (G5) because (G4) has a different consequent from rule r. Impression (G2) does not match the antecedent of rule r, and is thus eliminated. Impression (G5) partially matches rule r and the degree of matching is represented as a value between 0 and 1. Assuming $10,000 is considered to be a large value, (G1) and (G3) together completely match rule r, so the degree of matching is 1. Finally, we take the maximum of the match values, which is 1, as the interestingness value for rule r. Thus, rule r strongly confirms the general impressions. If we wanted to find unexpected condition rules instead of confirming rules, rule r would have a low score because it is consistent with the general impressions.

Liu et al. [1999] also proposed another technique to rank classification rules according to the user's background knowledge, which is represented in fuzzy rules. Based on the user's existing knowledge, three kinds of interesting rules can be mined: *unexpected*, *confirming*, and *actionable patterns*. An unexpected pattern is one that is unexpected or previously unknown to the user, which corresponds to our terms *surprising* and *novel*. A rule can be an unexpected pattern if it has an unexpected condition, an unexpected consequent, or both. A confirming pattern is a rule that partially or completely matches the user's existing knowledge, while an actionable pattern is one that can help the user do something to his or her advantage. To allow actionable patterns to be identified, the user should describe the situations in which he or she can take actions. For all three categories, the user must provide some patterns, represented in the form of fuzzy rules, that reflect his or her knowledge. The system matches each discovered pattern against these fuzzy rules. The discovered patterns are then ranked according to the degree to which they match. Liu et al. [1999] proposed different interestingness measures for the three categories. All these measures are based on functions of fuzzy values that represent the match between the user's knowledge and the discovered patterns.

The advantage of the methods of Liu et al. [1997, 1999] is that they rank mined patterns according to the user's existing knowledge, as well as the dataset. The disadvantage is that the user is required to represent his or her knowledge in the specifications, which might not be an easy task.

The specifications and matching algorithms of Liu et al. [1997, 1999] are designed for classification rules, and therefore cannot be applied to association rules. However, the general idea could be used for association rules if new specifications and matching algorithms were proposed for association rules with multiple item consequents.

2.2.3. Eliminating Uninteresting Patterns. To reduce the amount of computation and interactions with the user in filtering interesting association rules, Sahar [1999] proposed a method that removes uninteresting rules, rather than selecting interesting ones. In this method, no interestingness measures are defined; instead, the interestingness of a pattern is determined by the user via an interactive process. The method consists of three steps: (1) The *best candidate rule* is selected as the rule with exactly one condition attribute in the antecedent and exactly one consequence attribute in the consequent that has the largest *cover list*. The cover list of a rule R is all the mined rules that contain the condition and consequence of R. (2) The best candidate rule is presented to the user for classification into one of four categories: not-true-not-interesting, not-true-interesting, true-not-interesting, and true-and-interesting. Sahar [1999] described a rule as being not-interesting if it is "common knowledge," that is, not *novel* in our terminology. If the best candidate rule R is not-true-not-interesting or true-not-interesting, the system removes it and its cover list. If the rule is not-true-interesting, the system removes this rule as well as all the rules in its cover list that have the same antecedent, and keeps all the rules in its cover list that have more specific antecedents. Finally, if the rule is true-interesting, the system keeps it. This process iterates until the ruleset is empty or the user halts the process. The remaining patterns are true and interesting to the user.

The advantage of this method is that users are not required to provide specifications; rather, they work with the system interactively. They only need to classify simple rules as true or false and interesting or uninteresting, and then the system can eliminate a significant number of uninteresting rules. The drawback of this method is that although it makes the ruleset smaller, it does not rank the interestingness of the remaining rules. This method can also be applied to classification rules.

2.2.4. Constraining the Search Space. Instead of filtering uninteresting rules after the mining process, Padmanabhan and Tuzhilin [1998] proposed a method to narrow down the mining space on the basis of the user's expectations. In this method, no interestingness measure is defined. Here, the user's beliefs are represented in the same format as mined rules. Only surprising rules, that is, rules that contradict existing beliefs, are mined. The algorithm to find surprising rules consists of two parts: ZoominUR and ZoomoutUR. For a given belief $X \rightarrow Y$, ZoominUR finds all rules of the form $X,\ A \rightarrow \neg Y$ that have sufficient support and confidence in the dataset, which are more specific rules that have the contradictory consequence to the given belief. Then, ZoomoutUR generalizes the rules found by ZoominUR. For rule $X,\ A \rightarrow \neg Y$, ZoomoutUR finds all rules of the form $X',\ A \rightarrow \neg Y$, where X' is a subset of X.

This method is similar to the methods of Liu et al. [1997, 1999] in that the user needs to provide a specification of his or her knowledge. However, this method does not need to find all rules with sufficient support and confidence; instead, it only has to find any such rules that conflict with the user's knowledge, which makes the mining process more efficient. The disadvantage is that this method does not rank the rules. Although Padmanabhan and Tuzhilin [1998] proposed their method for association rules with only one item in their consequents, it can easily be applied to classification rules.

Based on the preceding analysis, we can see that if the user knows what kind of patterns he or she wants to confirm or contradict, the methods of Liu et al. [1997, 1999] and Padmanbhan and Tuzhilin [1989] are suitable. If the user does not want to explicitly represent knowledge about the domain, on the other hand, Sahar's [1999] interactive method is appropriate.

2.3. Semantic Measures

Recall that a semantic measure considers the semantics and explanations of the patterns. In this section, we consider semantic measures that are based on utility and actionability.

2.3.1. Utility Based Measures. A utility-based measure takes into consideration not only the statistical aspects of the raw data, but also the utility of the mined patterns. Motivated by decision theory, Shen et al. [2002] stated that "interestingness of a pattern = probability + utility." Based on both the user's specific objectives and the utility of the mined patterns, utility-based mining approaches may be more useful in real applications, especially in decision-making problems. In this section, we review utility-based measures for association rules. Since we use a unified notation for all methods, some representations differ from those used in the original articles.

The simplest method to incorporate utility is called *weighted association rule mining*, which assigns to each item a weight representing its importance [Cai et al. 1998]. These weights assigned to items are also called *horizontal* weights [Lu et al. 2001]. They can represent the price or profit of a commodity. In this scenario, two measures are proposed to replace support. The first is called weighted support, $(\sum_{i_j \in AB} w_j) Support(A \rightarrow B)$, where i_j denotes an item appearing in rule $A \rightarrow B$ and w_j denotes its corresponding weight. The first factor of the measure has a bias towards rules with more items. When the number of items is large, even if all the weights are small, the total weight may be large. The second measure, *normalized* weighted support, is proposed to reduce this bias and is defined as $\frac{1}{k}(\sum_{i_j \in AB} w_j) Support(A \rightarrow B)$, where k is the number of items in the rule. The traditional support measure is a special case of normalized weighted support because when all the weights for items are equal to 1, the normalized weighted support is identical to support.

Lu et al. [2002] proposed another data model by assigning a weight to each transaction. The weight represents the significance of the transaction in the dataset. Weights assigned to transactions are also called *vertical* weights [Lu et al. 2001]. For example, the weight can reflect the transaction time, that is, relatively recent transactions can be given greater weights. Based on this model, vertical weighted support is defined as

$$Support_v(A \rightarrow B) = \frac{\sum\limits_{AB \subseteq r} w\_v_r}{\sum\limits_{r \in D} w\_v_r},$$

where w_v_r denotes the vertical weight for transaction r.

The *mixed-weighted model* [Lu et al. 2001] uses both horizontal and vertical weights. In this model, each item is assigned a horizontal weight and each transaction a vertical

Table VII. Example Dataset

Treatment	Effectiveness	Side-Effects
1	2	4
2	4	2
2	4	2
2	2	3
2	1	3
3	4	2
3	4	2
3	1	4
4	5	2
4	4	2
4	4	2
4	3	1
5	4	1
5	4	1
5	4	1
5	3	1

weight. Mixed-weighted support is defined as:

$$Support_m(A \to B) = \frac{1}{k}\left(\sum_{i_j \in AB} w_j\right) Support_v(A \to B).$$

Both $support_v$ and $support_m$ are extensions of the traditional support measure. If all vertical and horizontal weights are set to 1, both $support_m$ and $support_m$ are identical to support.

Objective-oriented utility-based association (OOA) mining allows the user to set objectives for the mining process [Shen et al. 2002]. In this method, attributes are partitioned into two groups: *target* and *nontarget attributes*. A nontarget attribute (called a *nonobjective attribute* in Shen et al. [2002]) is only permitted to appear in the antecedents of association rules. A target attribute (called an *objective attribute* in Shen et al. [2002]) is only permitted to appear in the consequents of rules. The target attribute-value pairs are assigned utility values. The mining problem is to find frequent itemsets of nontarget attributes such that the utility values of their corresponding target attribute-value pairs are above a given threshold. For example, in Table VII, *Treatment* is a nontarget attribute, while *Effectiveness* and *Side-effect* are target attributes. The goal of the mining problem is to find treatments with high-effectiveness and little or no side-effects.

The utility measure is defined as

$$u = \frac{1}{support(A)} \sum_{A \subseteq r \wedge r \in DB} u_r(A),$$

where A is the nontarget itemsets to be mined (the *Treatment* attribute-value pairs in the example), $support(A)$ denotes the support of A in dataset D, r denotes a record that satisfies A, and $u_r(A)$ denotes the utility of A in terms of record r. The term $u_r(A)$ is defined as

$$u_r(A) = \sum_{A_i = v \in Cr} u_{A_i = v},$$

where Cr denotes the set of target items in record r, $A_i = v$ is an attribute-value pair of a target attribute, and $u_{A_i=v}$ denotes the latter's associated utility. If there is only one

Table VIII. Utility Values for *Effectiveness* and *Side-effects*

Effectiveness			Side-Effect		
Value	Meaning	Utility	Value	Meaning	Utility
5	Much better	1	4	Very serious	−0.8
4	Better	0.8	3	Serious	−0.4
3	No effect	0	2	A little	0
2	Worse	−0.8	1	Normal	0.6
1	Much worse	−1			

Table IX. Utilities of the Items

Itemset	Utility
Treatment = 1	−1.6
Treatment = 2	−0.25
Treatment = 3	−0.066
Treatment = 4	0.8
Treatment = 5	1.2

target attribute and its weight equals 1, then $\sum_{A \subseteq r \wedge r \in DB} u_r(A)$ is identical to *support*(A), and hence u equals 1.

Continuing the example, we assign the utility values to the target attribute-value pairs shown in Table VIII, and accordingly obtain the utility values for each treatment shown in Table IX. For example, Treatment 5 has the greatest utility value (1.2), and therefore, it best meets the user-specified target.

This data model was generalized in Zhang et al. [2004]. Attributes are again classified into nontarget and target attributes, called *segment* and *statistical attributes,* respectively, by the authors. For an itemset X composed of nontarget attributes, the interestingness measure, which is called the *statistic*, is defined as $statistic = f(D_x)$, where D_x denotes the set of records that satisfy X. Function f computes the statistic from the values of the target attributes in D_x. Based on this abstract framework, another detailed model, called *marketshare*, was proposed [Zhang et al. 2004]. In this model, the target attributes are *MSV* and P. The *MSV* attribute is a categorical attribute for which the market share values are to be computed, for example, *CompanyName*. The P is a continuous attribute, such as *GrossSales*, that is the basis for the market share computation for *MSV*. The interestingness measure called marketshare is defined as:

$$msh = \sum_{r \in D_x \wedge MSV_r = v} P_r \Bigg/ \sum_{r \in D_x} P_r,$$

where P_r denotes the P value for record r, and MSV_r denotes the *MSV* value for record r. A typical semantics for this measure is the percentage of sales P, for a specific company *MSV,* for given conditions X. If P_r is set to 1 for all records r, *msh* is equal to *confidence*$(X \rightarrow (MSV_r = v))$.

Carter et al. [1997] proposed the share-confidence framework which allows specification of the weights on attribute-value pairs. For example, in a transaction dataset, the weight could represent the quantity of a given commodity in a transaction. More precisely, the *share* of an itemset is the ratio of the total weight of the items in the itemset when they occur together to the total weight of all items in the database. Share can be regarded as a generalization of support. The share-confidence framework was generalized by other researchers to take into account weights on both attributes and attribute-value pairs [Hilderman et al. 1998; Barber and Hamilton 2003]. For example, in a transaction dataset, the weight on an attribute could represent the price of a commodity, and the weight on an attribute-value pair could represent the quantity of the commodity in a transaction. Based on this model, both support and confidence are generalized. Let I be the set of all possible items, let $X = \{A_1, \ldots, A_n\}$ be an itemset, and let D_X denote the set of records where the weight of each item in X is positive, that is:

$$D_X = \{r \mid \forall A_i \in X, w(A_i, r) > 0\},$$

where $w(A_i, r)$ denotes the weight of attribute A_i for transaction r. The *count-share* for itemset X is defined as:

$$count\_share = \frac{\sum_{r \in D_X} \sum_{A_i \in X} w(A_i, r)}{\sum_{r \in D} \sum_{A \in I} w(A, r)}.$$

Accordingly, the *amount-share* is defined as:

$$amount\_share = \frac{\sum_{r \in D_X} \sum_{A_i \in X} w(A_i, r) w(A_i)}{\sum_{r \in D} \sum_{A \in I} w(A, r) w(A)},$$

where $w(A_i)$ is the weight for attribute A_i. Let $A \rightarrow B$ be an association rule, where $A = \{A_1, \ldots, A_n\}$ and $B = \{B_1, \ldots, B_m\}$, and let D_{AB} denote the set of records where the weight of each item in A and each item in B is positive, that is:

$$D_{AB} = \{r \mid \forall A_i \in A, \forall B_j \in B, \ w(A_i, r) > 0 \wedge w(B_j, r) > 0\}.$$

The *count-confidence* of $A \rightarrow B$ is defined as:

$$count\_conf = \frac{\sum_{r \in D_{AB}} \sum_{A_i \in A} w(A_i, r)}{\sum_{r \in D_A} \sum_{A_i \in A} w(A_i, r)}.$$

This measure is an extension of the confidence measure because if all weights are set to 1 (or any constant), it becomes identical to confidence. Finally, the *amount-confidence* is defined as:

$$amount\_conf = \frac{\sum_{r \in D_{AB}} \sum_{A_i \in A} w(A_i, r) w(A_i)}{\sum_{r \in D_A} \sum_{A_i \in A} w(A_i, r) w(A_i)}.$$

Based on the data model in Hilderman et al. [1998], other researchers proposed another utility function [Yao et al., 2004; Yao and Hamilton 2006], defined as:

$$u = \sum_{r \in D_X} \sum_{A_i \in X} w(A_i, r) w(A_i).$$

This utility function is similar to amount-share, except that it represents a utility value, such as the profit in dollars, rather than a fraction of the total weight of all transactions in the dataset.

Table X. Utility-Based Interestingness Measures

Measures	Data Models	Extension of
Weighted Support	Weights for items	Support
Normalized Weighted Support	Weights for items	Support
Vertical Weighted Support	Weights for transactions	Support
Mixed-Weighted Support	Weights for both items and transactions	Support
OOA	Target and non target attributes; weights for target attributes	Support
Marketshare	Weight for each transaction, stored in attribute P in dataset.	Confidence
Count-Share	Weights for items and cells in dataset	Support
Amount-Share	Weights for items and cells in dataset	Support
Count-Confidence	Weights for items and cells in dataset	Confidence
Amount-Confidence	Weights for items and cells in dataset	Confidence
Yao et al.	Weights for items and cells in dataset	Support

Table X summarizes the utility measures discussed in this section by listing the name of each measure and its data model. The data model describes how the information relevant to the utility is organized in the dataset. All these measures are extensions of the support and confidence measures, and most of them extend the standard Apriori algorithm by identifying upper-bound properties for pruning. No single utility measure is suitable for every application because applications have different objectives and data models. Given a dataset, we could choose a utility measure by examining the data models for the utility measures given in Table X. For example, if we have a dataset with weights for each row, then we might choose the vertical weighted support measure.

2.3.2. Actionability. As mentioned in Section 2.2.2, an actionable pattern can help the user do something that is to his or her advantage. Liu et al. [1997] proposed that to allow actionable patterns to be identified, the user should describe the situations in which he or she can take actions. With their approach, the user provides some patterns, in the form of fuzzy rules, representing both possible actions and the situations in which they are likely to be taken. As with confirming patterns, their system matches each discovered pattern against the fuzzy rules and then ranks them, according to the degrees to which they match. Actions with the highest degrees of matching are selected to be performed.

Ling et al. [2002] proposed a measure to find optimal actions for profitable customer relationship management. In this method, a decision tree is mined from the data. The nonleaf nodes correspond to the customer's conditions, while the leaf nodes relate to the profit that can be obtained from the customer. The cost for changing a customer's condition is assigned. Based on the cost and profit gain information, the system finds the *optimal action*, that is, the action that maximizes $profit\_gain - \Sigma cost$. Since this method works on a decision tree, it is readily applicable to classification rules, but not to association rules.

Wang et al. [2002] suggested an integrated method to mine association rules and recommend the best with respect to profit to the user. In addition to support and confidence, the system incorporates two other measures: *rule profit* and *recommendation profit*. The rule profit is defined as the total profit obtained in transactions for a rule that match the rule. The recommendation profit for a rule is the average profit for each transaction that matches the rule. The recommendation system chooses the rules in order of recommendation profit, rule profit, and conciseness. This method can be directly applied to classification rules if profit information is integrated into all relevant attributes.

3. MEASURES FOR SUMMARIES

Summarization is one of the major tasks in knowledge discovery [Fayyad et al. 1996] and the key issue in online analytical processing (OLAP) systems. The essence of summarization is the formation of interesting and compact descriptions of raw data at different concept levels, which are called *summaries*. For example, sales information in a company may be summarized to levels of area, such as *City*, *Province*, and *Country*. It can also be summarized to levels of time, such as *Week*, *Month*, and *Year*. The combination of all possible levels for all attributes produces many summaries. Accordingly, using measures to find interesting summaries is an important issue.

We study four interestingness criteria for summaries: diversity, conciseness, peculiarity, and surprisingness. The first three are objective and the last subjective.

3.1. Diversity

Diversity has been widely used as an indicator of the interestingness of a summary. Although diversity is difficult to define, it is widely accepted that it is determined by two factors: the proportional distribution of classes in the population, and the number of classes [Hilderman and Hamilton 2001]. Table XI lists 19 measures for diversity. The first 16 are taken from Hilderman and Hamilton [2001] and the remaining 3 are taken from Zbidi et al. [2006]. In this definition, p_i denotes the probability for class i, $\overline{q}$ denotes the average probability for all classes, n_i denotes the number of samples for class i, and N denotes the total number of samples in the summary.

Recall that the first three columns of Table I give a summary describing students majoring in computer science, where the first column identifies the program of study, the second identifies nationality, and the third shows the number of students. For reference, the fourth column shows the values for the uniform distribution of the summary.

If variance $\sum_{i=1}^{m}(p_i - \overline{q})^2/(m-1)$ is used as the interestingness measure, the interestingness value for this summary is determined as follows:

$$\frac{\left(\frac{15}{300}-\frac{75}{300}\right)^2+\left(\frac{25}{300}-\frac{75}{300}\right)^2+\left(\frac{200}{300}-\frac{75}{300}\right)^2+\left(\frac{60}{300}-\frac{75}{300}\right)^2}{4-1}=0.24$$

Hilderman and Hamilton [2001] proposed some general principles that a good measure should satisfy:

(1) *Minimum Value Principle*. Given a vector $(n_1, \ldots, n_m)$, where $n_i = n_j$ for all i, j, measure $f(n_1, \ldots, n_m)$ attains its minimum value. This property indicates that the uniform distribution is the most uninteresting.

(2) *Maximum Value Principle*. Given a vector $(n_1, \ldots, n_m)$, where $n_1 = N - m + 1$, $n_i = 1$, $i = 2, \ldots, m$, and $N > m$, measure $f(n_1, \ldots, n_m)$ attains its maximum value. This property shows that the most uneven distribution is the most interesting.

(3) *Skewness Principle*. Given a vector $(n_1, \ldots, n_m)$, where $n_1 = N - m + 1$, $n_i = 1$, $i = 2, \ldots,$ m, and $N > m$, and a vector $(n_1 - c, n_2, \ldots, n_m, n_{m+1}, \ldots, n_{m+c})$, where $n_1 - c > 1$, $n_i = 1$, $i = 2, \ldots, m + c$, then $f(n_1, \ldots, n_m) >$ f$(n_1 - c, n_2, \ldots n_{m+c})$. This property specifies that when the total frequency remains the same, the interestingness measure for the most uneven distribution decreases when the number of classes of tuples increases. This property has a bias for small numbers of classes.

Table XI. Interestingness Measures for Diversity

Measure	Definition
Variance	$\frac{\sum_{i=1}^{m}(p_i-\overline{q})^2}{m-1}$
Simpson	$\sum_{i=1}^{m} p_i^2$
Shannon	$-\sum_{i=1}^{m} p_i \log_2 p_i$
Total	$-m\sum_{i=1}^{m} p_i \log_2 p_i$
Max	$\log_2 m$
McIntosh	$\frac{N-\sqrt{\sum_{i=1}^{m} n_i^2}}{N-\sqrt{N}}$
Lorenz	$\overline{q}\sum_{i=1}^{m}(m-i+1)p_i$
Gini	$\frac{\overline{q}\sum_{i=1}^{m}\sum_{j=1}^{m}\lvert p_i-p_j\rvert}{2}$
Berger	$\max(p_i)$
Schutz	$\frac{\sum_{i=1}^{m}\lvert p_i-\overline{q}\rvert}{2m\overline{q}}$
Bray	$\sum_{i=1}^{m}\min(p_i,\ \overline{q})$
Whittaker	$1-\frac{1}{2}\sum_{i=1}^{m}\lvert p_i-\overline{q}\rvert$
Kullback	$\log_2 m-\sum_{i=1}^{m} p_i \log_2 \frac{p_i}{\overline{q}}$
MacArthur	$-\sum_{i=1}^{m}\frac{p_i+\overline{q}}{2}\log_2\frac{p_i+\overline{q}}{2}-\frac{(\log_2 m-\sum_{i=1}^{m} p_i \log_2 p_i)}{2}$
Theil	$\frac{\sum_{i=1}^{m}\lvert p_i\log_2 p_i-\overline{q}\log_2\overline{q}\rvert}{m\overline{q}}$
Atkinson	$1-\prod_{i=1}^{m}\frac{p_i}{\overline{q}}$
Rae	$\frac{\sum_{i=1}^{m} n_i(n_i-1)}{N(N-1)}$
CON	$\sqrt{\frac{\sum_{i=1}^{m} p_i^2-\overline{q}}{1-\overline{q}}}$
Hill	$1-\frac{1}{\sqrt{\sum_{i=1}^{m} p_i^3}}$

(4) *Permutation Invariance Principle*. Given a vector $(n_1, \ldots, n_m)$ and any permutation $(i_1, \ldots, i_m)$ of $(1, \ldots, m)$, then $f(n_1, \ldots, n_m) = f(n_{i1}, \ldots, n_{im})$. This property specifies that interestingness for diversity is unrelated to the order of the class; it is only determined by the distribution of the counts.

(5) *Transfer Principle*. Given a vector $(n_1, \ldots, n_m)$ and $0 < c < n_j < n_i$, then $f(n_1, \ldots, n_i + c, \ldots, n_j - c, \ldots, n_m) > f(n_1, \ldots, n_i, \ldots, n_j, \ldots, n_m)$. This property specifies that interestingness increases when a positive transfer is made from the count of one tuple to another whose count is greater.

These principles can be used to identify the interestingness of a summary according to its distribution.

3.2. Conciseness and Generality

Concise summaries are easily understood and remembered, and thus they are usually more interesting than ones that are complex. Typically, a summary at a more general level is more concise than one at a more specific level.

Fabris and Freitas [2001] defined interestingness measures for attribute-value pairs in a data cube. For a single attribute, the I_1 measure reflects the difference between the observed probability of an attribute-value pair and the average probability in the summary, that is, $I_1(A = v) = |P(A = v) - 1/Card(A)|$, where $P(A = v)$ denotes the probability of attribute-value pair $A = v$, and $Card(A)$ denotes the cardinality of the attribute A, that is, the number of unique values for A in the summary. For the interaction of two attributes, the I_2 measure reflects the degree of correlation, on the assumption that dependencies are of interest. It is defined as:

$$I_2(A = v_a,\ B = v_b) = |P(A = v_a,\ B = v_b) - P(A = v_a)P(B = v_b)|,$$

where $P(A = v_a, B = v_b)$ denotes the observed probability of both attribute A taking value v_a and attribute B taking value v_b.

To deal with the conceptual levels introduced by hierarchies, Fabris and Freitas [2001] use coefficients to introduce a bias towards general concepts, which occur at higher levels of the hierarchies. In the data cube, the summaries corresponding to these concepts tend to be concise. Suppose that a summary to be analyzed is at level L_A in the hierarchy for attribute A, which has NHL_A levels, numbered from 0 to $NHL_A - 1$. The coefficient for the one-attribute case is defined as $CF_1 = \sqrt{\frac{NHL_A - L_A}{NHL_A}}$. At the most general level in the hierarchy L_A is 0, and thus CF_1 takes its maximum value of 1. At the most specific level, L_A is $NHL_A - 1$, and thus CF_1 takes its minimum value. For the two-attribute case, the coefficient is defined as $CF_2 = \sqrt{\frac{NHL_{\max} - (L_A + L_B)/2}{NHL_{\max}}}$, where $NHL_{\max} = \max(NHL_A, NHL_B)$, NHL_A and NHL_B are the total number of levels in the hierarchies for attributes A and B, respectively, and L_A and L_B denote the levels being analyzed for attributes A and B, respectively. Again, CF_2 takes its maximum value of 1 at the most general levels of A and B and its minimum value at the most specific levels of A and B.

The corrected measure for the one-attribute case is defined as:

$$F_1(A = v) = I_1(A = v)CF_1,$$

and the corrected measure for the two-attribute case is:

$$F_2(A = v_a,\ B = v_b) = I_2(A = v_a,\ B = v_b)CF_2.$$

3.3. Peculiarity

In data cube systems, a cell in the summary, rather than the summary itself, might be interesting due to its peculiarity. *Discovery-driven exploration* guides the exploitation process by providing users with interestingness values for measuring the peculiarity of the cells in a data cube, according to statistical models [Sarawagi et al. 1998]. Initially, the user specifies a starting summary, and the tool automatically calculates three kinds of interestingness values for each cell in the summary, based on statistical models. The first value, denoted SelfExp, indicates the interestingness of this cell relative to all other cells in the same summary. The second value, denoted InExp, indicates the maximum interestingness value if we drilled down from the cell to a more detailed summary somewhere beneath this cell. Also, for each path available for drilling down from this cell, a third kind of value, denoted PathExp, indicates the maximum interestingness value anywhere on the path. The value of SelfExp for a cell is defined as the difference between the observed and anticipated values. The anticipated value is calculated according to a table-analysis method from statistics [Hoaglin et al. 1985]. For example, in the three-dimensional cube $A - B - C$, the anticipated value for a cell could be calculated as the mean of several means: the mean of the summary for each individual attribute, the mean of the summary for each pair of attributes, and the overall mean, that is, $\overline{A} + \overline{B} + \overline{C} + \overline{AB} + \overline{AC} + \overline{BC} + \overline{ABC}$. InExp is obtained as the maximum of the SelfExp values over all cells that are under this cell. Each PathExp value is calculated as the maximum value of SelfExp over all cells reachable by drilling down along the path. The user can be guided by these three measures to navigate through the space of the data cube.

The process of automatically finding the underlying reasons for a peculiarity can be simplified [Sarawagi 1999]. The user identifies an interesting difference between two cells, and the system presents the most relevant data in more detailed cubes that account for the difference.

3.4. Surprisingness/Unexpectedness

Surprisingness is a suitable subjective criterion for evaluating the interestingness of summaries. A straightforward way to define a surprisingness measure is to incorporate the user's expectations into an objective interestingness measure. Most objective interestingness measures for summaries can be transformed into subjective ones by replacing the average probability with the expected probability. For example, variance $\sum_{i=1}^{m} (p_i - \overline{q})^2/(m-1)$ becomes $\sum_{i=1}^{m} (p_i - e_i)^2/(m-1)$, where p_i is the observed probability for a cell i, $\overline{q}$ is the average probability, and e_i is the expected probability for cell i.

Suppose that a user gives expectations for the distribution of students shown in the fifth column in Table I. The interestingness value of the summary in the context of these expectations is:

$$\frac{\left(\frac{15}{300} - \frac{20}{300}\right)^2 + \left(\frac{25}{300} - \frac{30}{300}\right)^2 + \left(\frac{200}{300} - \frac{180}{300}\right)^2 + \left(\frac{60}{300} - \frac{70}{300}\right)^2}{4-1} = 0.06.$$

Comparing this example with that in Section 3.1, we can see that the user's expectations are closer to the real distribution than the uniform distribution is. Therefore, when the expectations are added, the interestingness value of the summary decreases from 0.24 to 0.06. A summary may be interesting to a user who has some relevant background knowledge.

It is difficult for the user to specify all expectations quickly and consistently. The user may prefer to specify expectations for just one or a few summaries in the data cube. Therefore, a method is needed to propagate the expectations to all other summaries. Hamilton et al. [2006] proposed a propagation method for this purpose.

4. CONCLUSIONS

To reduce the number of mined results, many interestingness measures have been proposed for various kinds of patterns. In this article, we surveyed interestingness measures used in data mining. We summarized nine criteria to determine and define interestingness. Based on the form of the patterns produced by the data mining method, we distinguished measures for association rules, classification rules, and summaries. We distinguished objective, subjective, and semantics-based measures. Objective interestingness measures are based on probability theory, statistics, and information theory. Therefore, they have strict principles and foundations and their properties can be formally analyzed and compared. We surveyed the properties of objective measures, as well as relevant analysis methods and strategies for selecting such measures for applications. However, objective measures take into account neither the context of the domain of application nor the goals and background knowledge of the user. Subjective and semantics-based measures incorporate the user's background knowledge and goals, respectively, and are suitable both for more experienced users and interactive data mining. It is widely accepted that no single measure is superior to all others or suitable for all applications.

Of the nine criteria for interestingness, novelty (at least in the way we have defined it) has received the least attention. The prime difficulty is in modeling what the user does not know in order to identify what is new. Nonetheless, novelty remains a crucial factor in the appreciation for interesting results. Diversity is a major criterion for measuring summaries, but no work has been done so far to study the diversity of either association or classification rules. We consider this a possible research direction. For example, suppose we have two sets of association rules mined from a dataset. We might say that the set with more diverse rules is more interesting, and that a ruleset containing too many similar rules conveys less knowledge to the user. Compared with rules, much less research has been conducted on the interestingness of summaries. In particular, the utility and actionability of summaries could be investigated.

Existing subjective and semantics-based measures employ various representations of the user's background knowledge, which lead to different measures and procedures for determining interestingness. A general framework for representing knowledge that is related to data mining would be useful for defining a unifying view of subjective and semantics-based measures.

Choosing interestingness measures that reflect real human interest remains an open issue. One promising approach is to use metalearning to automatically select or combine appropriate measures. Another possibility is to develop an interactive user interface based on visually interpreting the data using a selected measure to assist the selection process. Extensive experiments comparing the results of interestingness measures with actual human interest could be used as another method of analysis. Since user interactions are indispensable in the determination of rule interestingness, it is desirable to develop new theories, methods, and tools to facilitate the user's involvement.

REFERENCES

AGRAWAL, R. AND SRIKANT, R. 1994. Fast algorithms for mining association rules. In *Proceedings of the 20th International Conference on Very Large Databases*. Santiago, Chile. 487–499.

BARBER, B. AND HAMILTON, H. J. 2003. Extracting share frequent itemsets with infrequent subsets. *Data Mining Knowl. Discovery 7*, 2, 153–185.

BASTIDE, Y., PASQUIER, N., TAOUIL, R., STUMME, G., AND LAKHAL, L. 2000. Mining minimal nonredundant association rules using frequent closed itemsets. In *Proceedings of the Ist International Conference on Computational Logic*. London, UK. 972–986.

BAY, S. D. AND PAZZANI, M. J. 1999. Detecting change in categorical data: Mining contrast sets. In *Proceedings of the 5th International Conference on Knowledge Discovery and Data Mining* (*KDD-99*). San Diego, CA. 302–306.

BAYARDO, R. J. AND AGRAWAL R. 1999. Mining the most interesting rules. In *Proceedings of the 5th International Conference on Knowledge Discovery and Data Mining* (*KDD-99*). San Diego, CA. 145–154.

BREIMAN, L., FREIDMAN, J., OLSHEN, R., AND STONE, C. 1984. *Classification and Regression Trees*. Wadsworth and Brooks, Pacific Grove, CA.

CAI, C. H., FU, A. W., CHENG, C. H., AND KWONG, W. W. 1998. Mining association rules with weighted items. In *Proceedings of the International Database Engineering and Applications Symposium* (*IDEAS '98*). Cardiff, UK. 68–77.

CARTER, C. L., HAMILTON, H. J., AND CERCONE, N. 1997. Share-Based measures for itemsets. In *Proceedings of the Ist European Symposium on Principles of Data Mining and Knowledge Discovery* (*PKDD '97*). Trondheim, Norway. 14–24.

CARVALHO, D. R. AND FREITAS, A. A. 2000. A genetic algorithm-based solution for the problem of small disjuncts. In *Proceedings of the 4th European Conference on Principles of Data Mining and Knowledge Discovery* (*PKDD 2000*). Lyon, France. 345–352.

CHAN, R., YANG, Q., AND SHEN, Y. 2003. Mining high-utility itemsets. In *Proceedings of the 3rd IEEE International Conference on Data Mining* (*ICDM '03*). Melbourne, FL. 19–26.

CLARK, P. AND BOSWELL, R. 1991. Rule induction with CN2: Some recent improvements. In *Proceedings of the 5th European Working Session on Learning* (*EWSL '91*). Porto, Portugal. 151–163.

DONG, G. AND LI, J. 1998. Interestingness of discovered association rules in terms of neighborhood-based unexpectedness. In *Proceedings of the 2nd Pacific Asia Conference on Knowledge Discovery in Databases* (*PAKDD-98*). Melbourne, Australia. 72–86.

FABRIS, C. C. AND FREITAS, A. A. 2001. Incorporating deviation-detection functionality into the OLAP paradigm. In *Proceedings of the 16th Brazilian Symposium on Databases* (*SBBD 2001*). Rio de Janeiro, Brazil. 274–285.

FAYYAD, U. M., PIATETSKY-SHAPIRO, G., AND SMYTH, P. 1996. From data mining to knowledge discovery: An overview. In *Advances in Knowledge Discovery and Data Mining*, U. M. Fayyad et al., Eds. MIT Press. Cambridge, MA, 1–34.

FORSYTH, R. S., CLARKE, D. D., AND WRIGHT, R. L. 1994. Overfitting revisited: An information-theoretic approach to simplifying discrimination trees. *J. Exp. Theor. Artif. Intell. 6*, 289–302.

FREITAS, A. A. 1998. On objective measures of rule surprisingness. In *Proceedings of the 2nd European Symposium on Principles of Data Mining and Knowledge Discovery* (*PKDD '98*). Nantes, France. 1–9.

FÜRNKRANZ, J. AND FLACH, P. A. 2005. ROC 'n' rule learning: Towards a better understanding of covering algorithms. *Mach. Learn. 58*, 1, 39–77.

GRAY, B. AND ORLOWSKA, M. E. 1998. CCAIIA: Clustering categorical attributes into interesting association rules. In *Proceedings of the 2nd Pacific Asia Conference on Knowledge Discovery and Data Mining* (*PAKDD-98*). Melbourne, Australia. 132–143.

HAMILTON, H. J., GENG, L., FINDLATER, L., AND RANDALL, D. J. 2006. Efficient spatio-temporal data mining with GenSpace graphs. *J. Appl. Logic 4*, 2, 192–214.

HILDERMAN, R. J., CARTER, C. L., HAMILTON, H. J., AND CERCONE, N. 1998. Mining market basket data using share measures and characterized itemsets. In *Proceedings of the 2nd Pacific Asia Conference on Knowledge Discovery in Databases* (*PAKDD-98*). Melbourne, Australia. 72–86.

HILDERMAN, R. J. AND HAMILTON, H. J. 2001. *Knowledge Discovery and Measures of Interest*. Kluwer Academic, Boston, MA.

HOAGLIN, D. C., MOSTELLER, F., AND TUKEY, J. W., EDS. 1985. *Exploring Data Tables, Trends, and Shapes*. Wiley, New York.

JAROSZEWICZ, S. AND SIMOVICI, D. A. 2001. A general measure of rule interestingness. In *Proceedings of the 5th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD 2001)*. Freiburg, Germany. 253–265.

KLOSGEN, W. 1996. Explora: A multipattern and multistrategy discovery assistant. In *Advances in Knowledge Discovery and Data Mining*, U. M. Fayyad et al., Eds. MIT Press, Cambridge, MA, 249–271.

KNORR, E. M., NG, R. T., AND TUCAKOV, V. 2000. Distance based outliers: Algorithms and applications. *Int. J. Very Large Databases 8*, 237–253.

LAVRAC, N., FLACH, P., AND ZUPAN, B. 1999. Rule evaluation measures: A unifying view. In *Proceedings of the 9th International Workshop on Inductive Logic Programming* (*ILP '99*). Bled, Slovenia. Springer-Verlag, 174–185.

LENCA, P., MEYER, P., VAILLANT, B., AND LALLICH, S. 2004. A multicriteria decision aid for interestingness measure selection. Tech. Rep. LUSSI-TR-2004-01-EN, May 2004. LUSSI Department, GET/ENST, Bretagne, France.

LI, G. AND HAMILTON, H. J. 2004. Basic association rules. In *Proceedings of the 4th SIAM International Conference on Data Mining*. Orlando, FL. 166–177.

LING, C., CHEN, T., YANG, Q., AND CHEN, J. 2002. Mining optimal actions for profitable CRM. In *Proceedings of the 2002 IEEE International Conference on Data Mining* (*ICDM '02*). Maebashi City, Japan. 767–770.

LIU, B., HSU, W., AND CHEN, S. 1997. Using general impressions to analyze discovered classification rules. In *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining* (*KDD-97*). Newport Beach, CA. 31–36.

LIU, B., HSU, W., MUN, L., AND LEE, H. 1999. Finding interesting patterns using user expectations. *IEEE Trans. Knowl. Data Eng. 11*, 6, 817–832.

LU, S., HU, H., AND LI, F. 2001. Mining weighted association rules. *Intell. Data Anal. 5*, 3, 211–225.

MCGARRY, K. 2005. A survey of interestingness measures for knowledge discovery. *Knowl. Eng. Review 20*, 1, 39–61.

MURTHY, S. K. 1998. Automatic construction of decision trees from data: A multi-disciplinary survey. *Data Mining Knowl. Discovery 2*, 4, 345–389.

OHSAKI, M., KITAGUCHI, S., OKAMOTO, K., YOKOI, H., AND YAMAGUCHI, T. 2004. Evaluation of rule interestingness measures with a clinical dataset on hepatitis. In *Proceedings of the 8th European Conference on Principles of Data Mining and Knowledge Discovery* (*PKDD 2004*). Pisa, Italy. 362–373.

PADMANABHAN, B. AND TUZHILIN, A. 1998. A belief-driven method for discovering unexpected patterns. In *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining* (*KDD-98*). New York. 94–100.

PADMANABHAN, B. AND TUZHILIN, A. 2000. Small is beautiful: Discovering the minimal set of unexpected patterns. In *Proceedings of the 6th International Conference on Knowledge Discovery and Data Mining* (*KDD 2000*). Boston, MA. 54–63.

PAGALLO, G. AND HAUSSLER, D. 1990. Boolean feature discovery in empirical leaning. *Mach. Learn. 5*, 1, 71–99.

PIATETSKY-SHAPIRO, G. 1991. Discovery, analysis, and presentation of strong rules. In *Knowledge Discovery in Databases*, G. Piatetsky-Shapiro and W. J. Frawley, Eds. MIT Press, Cambridge, MA, 229–248.

PIATETSKY-SHAPIRO, G. AND MATHEUS, C. 1994. The interestingness of deviations. In *Proceedings of the AAAI-94 Workshop on Knowledge Discovery in Databases* (*KDD-94*). Seattle, WA. 25–36.

QUINLAN, J. R. 1986. Induction of decision trees. *Mach. Learn. 1*, 1, 81–106.

SAHAR, S. 1999. Interestingness via what is not interesting. In *Proceedings of the 5th International Conference on Knowledge Discovery and Data Mining* (*KDD-99*). San Diego, CA. 332–336.

SARAWAGI, S. 1999. Explaining differences in multidimensional aggregates. In *Proceedings of the 25th International Conference on Very Large Databases* (*VLDB '99*). Edinburgh, U. K. 42–53.

SARAWAGI, S., AGRAWAL, R., AND MEGIDDO, N. 1998. Discovery-driven exploration of OLAP data cubes. In *Proceedings of the 6th International Conference of Extending Database Technology* (*EDBT '98*). Valencia, Spain. 168–182.

SHEN, Y. D., ZHANG, Z., AND YANG, Q. 2002. Objective-Oriented utility-based association mining. In *Proceedings of the 2002 IEEE International Conference on Data Mining* (*ICDM '02*). Maebashi City, Japan. 426–433.

SILBERSCHATZ, A. AND TUZHILIN, A. 1995. On subjective measures of interestingness in knowledge discovery. In *Proceedings of the Ist International Conference on Knowledge Discovery and Data Mining* (*KDD-95*). Montreal, Canada. 275–281.

SILBERSCHATZ, A. AND TUZHILIN, A. 1996. What makes patterns interesting in knowledge discovery systems. *IEEE Trans. Knowl. Data Eng. 8*, 6, 970–974.

TAN, P. AND KUMAR, V. 2000. Interestingness measures for association patterns: A perspective. Tech. Rep. 00-036, Department of Computer Science, University of Minnesota.

TAN, P., KUMAR, V., AND SRIVASTAVA, J. 2002. Selecting the right interestingness measure for association patterns. In *Proceedings of the 8th International Conference on Knowledge Discovery and Data Mining* (*KDD 2002*). Edmonton, Canada. 32–41.

VAILLANT, B., LENCA, P., AND LALLICH, S. 2004. A clustering of interestingness measures. In *Proceedings of the 7th International Conference on Discovery Science* (*DS 2004*). Padova, Italy. 290–297.

VITANYI, P. M. B. AND LI, M. 2000. Minimum description length induction, Bayesianism, and Kolmogorov complexity. *IEEE Trans. Inf. Theory 46*, 2, 446–464.

WANG, K., ZHOU, S., AND HAN, J. 2002. Profit mining: From patterns to actions. In *Proceedings of the 8th Conference on Extending Database Technology* (*EDBT 2002*). Prague, Czech Republic. 70–87.

WEBB, G. I. AND BRAIN, D. 2002. Generality is predictive of prediction accuracy. In *Proceedings of the 2002 Pacific Rim Knowledge Acquisition Workshop* (*PKAW 2002*). Tokyo. 117–130.

YAO, Y. Y., CHEN, Y. H., AND YANG, X. D. 2006. A measurement-theoretic foundation of rule interestingness evaluation. In *Foundations and Novel Approaches in Data Mining*, T. Y. Lin et al., Eds. Springer-Verlag, Berlin, 41–59.

YAO, Y. Y. AND ZHONG, N. 1999. An analysis of quantitative measures associated with rules. In *Proceedings of the 3rd Pacific-Asia Conference on Knowledge Discovery and Data Mining* (*PAKDD-99*). Beijing, China. 479–488.

YAO, H., HAMILTON, H. J., AND BUTZ, C. J. 2004. A foundational approach for mining itemset utilities from databases. In *Proceedings of the SIAM International Conference on Data Mining*. Orlando, FL. 482–486.

YAO, H. AND HAMILTON, H. J. 2006. Mining itemset utilities from transaction databases. *Data Knowl. Eng. 59*, 3.

ZBIDI, N., FAIZ, S., AND LIMAM, M. 2006. On mining summaries by objective measures of interestingness. *Mach. Learn. 62*, 3, 175–198.

ZHANG, H., PADMANABHAN, B., AND TUZHILIN, A. 2004. On the discovery of significant statistical quantitative rules. In *Proceedings of the 10th International Conference on Knowledge Discovery and Data Mining* (*KDD 2004*). Seattle, WA. 374–383.

ZHONG, N., YAO, Y. Y., AND OHSHIMA, M. 2003. Peculiarity oriented multidatabase mining. *IEEE Trans. Knowl. Data Engi. 15*, 4, 952–960.

Received June 2005; revised March 2006; accepted March 2006