A Measurement-Theoretic Foundation of Rule Interestingness Evaluation

Yiyu Yao, Yaohua Chen, Xuedong Yang
Department of Computer Science, University of Regina
Regina, Saskatchewan, Canada S4S 0A2
E-mail: {yyao, chen115y, yang}@cs.uregina.ca

Abstract

Many measures have been proposed and studied extensively in data mining for evaluating the usefulness or interestingness of discovered rules. They are normally defined based on structural characteristics or statistical information about the rules. The meaningfulness of each measure is interpreted based on some intuitive argument or mathematical properties. There does not exist a framework in which one is able to represent explicitly, precisely, and formally the user judgments. Since usefulness of discovered rules must be eventually judged by users, a framework that takes into consideration of users is needed. The objective of the paper is to propose such a framework based on the notion of user preference. The results are useful in establishing a measurement-theoretic foundation of rule evaluation.

1 Introduction

With rapidly increasing capabilities of accessing, collecting, and storing data, knowledge discovery in databases (KDD) has been emerged as a new area of research in computer science. The objective of KDD systems is to extract implicit, previously unknown, and potentially useful information and knowledge from databases. The discovered knowledge is often expressed in terms of a set of rules. They represent relationships, such as correlation, association, and causation, among concepts [37]. For example, the well known association rules deal with relationships among sale items [1, 3].

There are many types of rules embedded in a large database [33]. Furthermore, the number of rules is typically huge and only a small portion of rules is actually useful [26]. An important problem in data mining is the evaluation of *interestingness* of the mined rules and filtering out useless rules [26]. Many measures have been proposed and studied to quantify the usefulness or interestingness of rules [9, 12, 25, 26, 37]. The results lead to an in-depth understanding of different aspects of rules. It is recognized that each measure reflects a certain characteristics of rules. On the other hand, many studies investigate and compare rule interestingness measures based on intuitive argument or some mathematical properties. There is a lack of a well accepted framework for examining the issues of rule interestingness in a systematic and unified manner.

In this paper, we argue that measurement theory can be used to establish a solid foundation for rule interestingness evaluation. The theory provides the necessary concepts and methodologies for the representation, classification, characterization, and interpretation of user judgment of the usefulness of rules. A measure of rule interestingness is viewed as a quantitative representation of user judgment. The meaningfulness of a measure is determined by the users' perception of the usefulness of rules.

Existing studies of rule interestingness evaluation can be viewed as measure-centered approaches. Measures are used as primitive notions to quantify the interestingness of rules. In contrast, our method is a user-centered approach. User judgment, expressed by a user preference relation on a set of rules, is used as a primitive notion to model rule interestingness. Measures are treated as a derived notion that provides a quantitative representation of user judgment.

In order to achieve the above objective, the rest of the paper is organized as follows. In section 2, we review existing measures of rules interestingness, which reveals some difficulties with existing studies and provides motivations to the current study. Section 3 presents an overview of measurement theory. Section 4 applies measurement theory to build a framework of rule interestingness evaluation.

2 Rule Evaluation

As an active research area in data mining, rule evaluation has been considered by many authors from different perspectives. We present a critical review of studies on rule evaluation in order to observe their difficulties. This leads to a new direction for future research.

2.1 A critical review of existing studies

Studies related to rule evaluation can be divided into two classes. One class, the majority of studies, deals with the applications of quantitative measures to reduce the size of search space of rules in the mining process, to filter out mined but non-useful rules, or to evaluate the effectiveness of a data mining system. The other class, only a small portion of studies, is devoted solely to the investigations of rule evaluation on its own. We summarize the main results from the following different points of views.

The roles of rule evaluation

It is generally accepted that KDD is an interactive and iterative process consisting of the many phases [6, 11, 20, 27]. Fayyad *et al.* presented a KDD process consisting of the following steps: developing and understanding of the application domain, creating a target data set, data cleaning and preprocessing, data reduction and projection, choosing the data mining task, choosing the data mining algorithm(s), data mining, interpreting mined patterns, and consolidating, and acting on, the discovered knowledge [6, 7]. Rule evaluation plays different roles in different phases of the KDD process.

From the existing studies, one can observe that rule evaluation plays at least three different types roles. In the data mining phase, quantitative measures can be used to reduce the size of search space. An example is the use of well known $support$ measure, which reduces the number of itemsets need to be examined [1]. In the phase of interpreting mined patterns, rule evaluation plays a role in selecting the useful or interesting rules from the set of discovered rules [25, 26]. For example, the $confidence$ measure of association rules is used to select only strongly associated itemsets [1]. In fact, many measures associated with rules are used for such a purpose [37]. Finally, in the phase of consolidating and acting on discovered knowledge, rule evaluation can be used to quantify the usefulness and effectiveness of discovered rules. Many measures such as cost, classification error, and classification accuracy play such a role [9]. Rule evaluation in this regard is related to the evaluation of a data mining system.

The process based approach captures the procedural aspects of KDD. Recently, Yao proposed a conceptual formulation of KDD in a three-layered framework [33]. They are the kernel level, technique level, and application level. The kernel level focuses on formal characterization, description, representation, and classification of knowledge embedded in a database without reference to mining algorithms. It provides answers to the questions "What is the knowledge embedded in a database?" The technique level concentrates on data mining algorithms without reference to specific applications. It provides answers to the questions "How to discover knowledge embedded in a database?" The application level focuses on the use of discovered knowledge with respect to particular domains. It provides answers to the question "How to apply the discovered knowledge?"

With respect to the three-layered framework, rule evaluation play the similarly roles as in the process based framework. In the kernel level, quantitative measures can be used to characterize and classify different types of rules. In the algorithm level, measures can be used to reduce search space. In the application level, measures can be used to quantify the utility, profit, effectiveness, or actionability of discovered rules.

Subjective vs. objective measures

Silberschatz and Tuzhilin suggested that measures can be classified into two categories consisting of objective measures and subjective measures [25]. Objective measures depend only on the structure of rules and the underlying data used in the discovery process. Subjective measures also depend on the user who examines the rules [25]. In comparison, there are limited studies on subjective measures. For example, Silberschatz and Tuzhilin proposed a subjective measure of rule interestingness based on the notion of unexpectedness and in terms of a user belief system [25, 26].

Statistical, structural vs. semantic measures

Many measures, such as support, confidence, independence, classification error, etc., are defined by based on statistical characteristics of rules. A systematic analysis of such measures is given by Yao *et al.* using a 2×2 contingency table induced by a rule [35, 37].

The structural characteristics of rules have been considered in many measures. For example, information, such as the disjunct size, attribute interestingness, the asymmetry of classification rules, etc., can be used [9]. These measures reflect the simplicity, easiness of understanding, or applicability of rules.

Although statistical and structural information provides an effective indicator of the potential effectiveness of a rule, its usefulness is limited. One needs to consider the semantics aspects of rules or explanations of rules [36]. Semantics centered approaches are application and user dependent. In addition to statistical information, one incorporates other domain specific knowledge such as user interest, utility, value, profit, actionability, and so on. Two examples of semantic based approaches are discussed below.

Profit or utility based mining is a special kind of constraint based mining, taking into account both statistical significance and profit significance [29, 15]. Doyle discussed the importance and usefulness of basic notions of economic rationality, such as utility functions, and suggests that economic rationality should play as large a role as logical rationality in rule reasoning [4]. For instance, one would not be interested in a frequent association that does not generate enough profit. The profit based measures allow the user to prune the rules with high statistical significance, but generate low profit or high risk. For example, Barber and Hamilton proposed the notion of share measures which consider the contribution, in terms of profit, of an item in an itemset [2].

Actionable rule mining deals with profit-driven actions required by business decision making [16, 18]. A rule is referred to as actionable if the user can do something about it. For example, a user may be able to change the non-desirable/non-profitable patterns to desirable/profitable patterns.

Measures defined by statistical and structural information may be viewed as objective measures. They are user, application and domain independent. For example, a pattern is deemed interesting if it has certain statistical properties. These measures may be useful in the kernel level of the three-layered framework. Different classes of rules can be identified based on statistical characteristics, such as peculiarity rules (low support and high confidence), exception rules (low support and high confidence, but complement to other high support and high confidence rules), and outlier patterns (far away from the statistical mean) [38].

Semantic based measures involve the user interpretation of domain specific notions such as profit and actionability. They may be viewed as subjective measures. Such measures are useful in the application level of the three-layered framework. The usefulness of rules are measured and interpreted based on domain specific notions.

Single rule vs. multiple rules

Rule evaluation can also be divided into measures for single rule and measures for a set of rules. Furthermore, a measure for a set of rules can be obtained from measures for single rules. For example, conditional probability can be used as a measure for a single classification rule, conditional entropy can be used as a measure for a set of classification rules [34]. The latter is defined in terms of the former.

Measures for multiple rules concentrate on properties of a set of rules. They are normally expressed as some kind of average. Hilderman and Hamilton examined many measures for multiple rules known as the summarizes of a database [12].

Axiomatic approaches

Instead of focusing on rules, the axiomatic approaches study the required properties of quantitative measures.

Suppose that the discovered knowledge is represented in terms of rules of the form, $E \rightarrow H$, and is paraphrased as "if E then H". Piatetsky-Shapiro [22] suggested that a quantitative measure of rule $E \rightarrow H$ may be computed as a function of $support(E)$, $support(H)$, $support(E \wedge H)$, rule complexity, and possibly other parameters such as the mutual distribution of E and H or the domain size of E and H. For the evaluation of rules, Piatetsky-Shapiro [22] introduced three axioms. Major and Mangano [19] added a fourth axioms. Klösgen [13] studied a special class of measures that are characterized by two quantities, $confidence(E \rightarrow H)$ and $support(E)$. The $support(H \wedge E)$ is obtained by $confidence(E \rightarrow H) support(E)$. Suppose $support(E, H)$ is a measure associated with rule $E \rightarrow H$. The version of the four axioms given by Klösgen [13] is:

(i). $Q(E, H) = 0$ if E and H are statistically independent,

(ii). $Q(E, H)$ monotonically increases in $confidence(E \rightarrow H)$ for a fixed $support(E)$,

(iii). $Q(E, H)$ monotonically decreases in $support(E)$ for a fixed $support(E \wedge H)$,

(iv). $Q(E, H)$ monotonically increases in $support(E)$ for a fixed $confidence(E \rightarrow H) > support(H)$.

The axiomatic approach is widely used in many other disciplines.

An axiomatic study of measures for multiple rules has been given by Hilderman and Hamilton [12].

2.2 A direction for future research

From the previous discussions, one can make several useful observations. Studies on rule evaluations can be classified in several ways. Each of them provides a different view. Most studies on rule evaluation concentrate on specific measures, with each measure reflects certain aspects of rules. Quantitative measures are typically interpreted using intuitively defined notions, such as novelty, usefulness, and non-trivialness, unexpectedness, and so on. Therefore, there is a need for a unified framework that enables us to define, interpreted, and compare different measures.

A research direction for rule evaluation is the study of its foundations. Several issues should be considered. One needs to link the meaningfulness of a measure to its usage. In theory, it may not be meaningful to argue which measure is better without reference to its roles and usages. It is also necessary to build a framework in which various notions of rule evaluation can be formally and precisely defined and interpreted. The study of rule evaluation needs to be connected to the study of foundations of data mining.

In the rest of the paper, we will demonstrate that measurement theory can be used to build a foundation of rule interestingness evaluation.

3 Overview of Measurement Theory

For completeness, we first give a brief review of the basic notions of measurement theory that are pertinent to our discussion. The contents of this section draw heavily from Krantz *et al.* [14], Roberts [23] and French [10].

When measuring an attribute of a class of objects or events, we may associate numbers with the individual objects so that the properties of the attribute are faithfully represented as numerical properties [14]. The properties are usually described by certain qualitative relations and operations. Consider an example discussed by Krantz *et al.* [14]. Suppose we are measuring the lengths of a set U of straight, rigid rods. One important property of length can be described by a qualitative relation "longer than". Such a relation can be obtained by first placing two rods, say a and b, side by side and adjusting them so that they coincide at one end, and then observing whether a extends beyond b at the other end. We say that a is longer than b, denoted by $a \succ b$, if a extends beyond b. In this case, we would like to assign numbers $\phi(a)$ and $\phi(b)$ with $\phi(a) > \phi(b)$ to reflect the results of the comparison. That is, we require the numbers assigned to the individual rods satisfy the condition: for all $a, b \in U$,

$$a \succ b \Longleftrightarrow \phi(a) > \phi(b), \qquad (1)$$

In other words, the qualitative relation "longer than", $\succ$, in the empirical system is faithfully reflected by the quantitative relation "greater than", $>$, in the numerical system. Another property of length is that we can concatenate two or more rods by putting them end to end in a straight line, and compare the length of this set with that of another set. The concatenation of a and b can be written as $a \circ b$. In order to reflect such a property, we require the numbers assigned to the individual rods be additive with respect to concatenation. That is, in addition to condition (1), the numbers assigned must also satisfy the following condition: for all $a, b \in U$,

$$\phi(a \circ b) = \phi(a) + \phi(b). \qquad (2)$$

Thus, concatenation $\circ$ in the empirical system is preserved by addition $+$ in the numerical system. Many other properties of length comparison and of concatenation of rods can be similarly formulated. For instance, $\succ$ should be transitive and $\circ$ should be commutative and associative. The numbers assigned must reflect these properties as well. This simple example clearly illustrates the basic ideas of measurement theory, which is primarily concerned with choosing consistent quantitative representations of qualitative systems.

Based on the discussion of the above example, some important concepts of measurement theory are introduced as follows. A *relational system* (*structure*) is a set together with one or more relations (operations) on that set. That is, a relational system is an ordered $(p + q + 1)$-tuple $\mathcal{A} = (U, R_1, \ldots, R_p, \circ_1, \ldots, \circ_q)$, where U is a set, $R_1, \ldots, R_p$ are (not necessarily binary) relations on U, and $\circ_1, \ldots, \circ_q$ are binary operations on U. We call a relational system a numerical relational system if U is the set (or a subset) of real numbers. Operations can be considered as a special kind of relations. For convenience, we separate them from other relations. As illustrated by the example of rigid rods, in measurement, we start with an observed or empirical system $\mathcal{A}$ and seek a mapping into a numerical relational system $\mathcal{B}$ which preserves or faithfully reflects all the properties of the relations and operations in $\mathcal{A}$.

Consider two relational systems, an empirical (a qualitative) system $\mathcal{A} = (U, R_1, \ldots, R_p, \circ_1, \ldots, \circ_q)$, and a numerical system $\mathcal{B} = (V, R'_1, \ldots, R'_p, \circ'_1, \ldots, \circ'_q)$. A function $f : U \rightarrow V$ is called a *homomorphism* from $\mathcal{A}$ to $\mathcal{B}$ if, for all $a_1, \ldots, a_{r_i} \in \mathcal{A}$,

$$R_i(a_1, \ldots, a_{r_i}) \Longleftrightarrow R'_i[f(a_1), \ldots, f(a_{r_i})], \quad i = 1, \ldots, p,$$

and for all $a, b \in \mathcal{A}$,

$$f(a \circ_j b) = f(a) \circ'_j f(b), \quad j = 1, \ldots, q.$$

The empirical system for the earlier example is denoted by $(U, \succ, \circ)$, where U is the set of rigid rods and their finite concatenations, $\succ$ is the binary relation "longer than" and $\circ$ is the concatenation operation. The numerical relation system is $(\Re, >, +)$, where $\Re$ is the set of real numbers, $>$ is the usual "greater than" relation and $+$ is the arithmetic operation of addition. The numerical assignment $\phi(\cdot)$ is a homomorphism which maps U into $\Re$, $\succ$ into $>$, and $\circ$ into $+$ in such a way that $>$ preserves the properties of $\succ$, and $+$ preserves the properties of $\circ$ as stated by conditions (1) and (2).

There are three fundamental issues in measurement theory [10, 14, 23]. Suppose we are seeking a quantitative representation of an empirical system. The first step, naturally, is to define the relations and operations to be represented. We must describe the valid use of these relations and operations. The consistency properties to be preserved are known as *axioms*. The set of axioms characterizing the empirical system should be complete in the sense that every consistency property that we demand is either in the list or deducible from those in the list. The next task is to choose a numerical system. The final step is to construct an appropriate homomorphism. A *representation theorem* asserts that if a given empirical system satisfies certain axioms, then a homomorphism into the chosen numerical system can be constructed. A homomorphism into the set of real numbers is called a *scale*. The next question concerns the uniqueness of the scale. A uniqueness theorem is generally obtained by identifying a set of *admissible transformations*. If $\phi(\cdot)$ is a scale representing an empirical system and if $\lambda(\cdot)$ is an admissible transformation, then $\lambda(\phi(\cdot))$ is also a scale representing the same empirical system.

If the truth (falsity) of a numerical statement involving a scale remains unchanged under all admissible transformations, we say that it is quantitatively meaningful. A numerical statement may be quantitatively meaningful, but qualitatively meaningless. In order for a quantitative statement to be qualitatively meaningful, it must reflect or model a meaningful statement in the empirical system.

Examples of the discussed view of measurement theory include the axiomatization of probability and expected utility theory [21, 24], the axiomatization of possibility functions [5] and the axiomatization of belief functions [30].

4 Application of Measurement Theory to Rule Evaluation

Given a database, in theory, there exists a set of rules embedded in it, independent of whether one has an algorithm to mine them. For a particular application, the user may only be interested in a certain type of rules. For rule evaluation, we assume that we are given a set of rules.

Let $\mathbf{R}$ be a set of rules derivable from a databases. User judgment of the usefulness or interestingness of rules can be formally described by a user preference relation $\succ$ on $\mathbf{R}$. Given two rules $R', R'' \in \mathbf{R}$, if a user judges R' to be more useful than R'', we say that the user prefers R' to R'' and write $R' \succ R''$. That is,

$$R' \succ R'' \Leftrightarrow \text{ the user prefers } R' \text{ to } R''. \qquad (3)$$

In the absence of strict preference, i.e., if both $\neg(R' \succ R'')$ and $\neg(R' \succ R'')$ hold, we say that R' and R'' are indifferent. An indifference relation $\sim$ on $\mathbf{R}$ can be defined as follows:

$$R' \sim R'' \Leftrightarrow (\neg(R' \succ R''), \neg(R'' \succ R')). \qquad (4)$$

In our formulation, we treat the user preference relation $\succ$ as a primitive notion. At this stage, we will not attempt to define and interpret a user preference relation using other notions.

The next issue is to identify the desired properties of a preference relation so that it can be measured quantitatively. Such consistency properties are known as axioms. We consider the following two axioms:

- **Asymmetry**:
  $R' \succ R'' \Rightarrow \neg(R'' \succ R')$,
- **Negative transitivity**:
  $(\neg(R' \succ R''), \neg(R'' \succ R''')) \Rightarrow \neg(R' \succ R''')$.

The first axiom is the asymmetry axiom, which requires that a user cannot prefer R' to R'' and at the same time prefers R'' to R'. The second is the negative transitivity axiom, which means that if a user does not prefer R' to R'', nor R'' to R''', the user should not prefer R' to R'''. A preference relation satisfying these two axioms is called a *weak order*. If a preference relation is a weak order, it is transitive, i.e., $R' \succ R''$ and $R'' \succ R'''$ imply $R' \succ R'''$. It seems reasonable that a user preference relation should satisfy these two axioms.

A few additional properties of a weak order are summarized in the following lemma.

Lemma 1 *Suppose a preference relation $\succ$ on a finite set of rules $\mathbf{R}$ is a weak order. Then,*

- *the relation $\sim$ is an equivalence relation,*
- *exactly one of $R' \succ R''$, $R'' \succ R'$ and $R' \sim R''$ holds for every $R', R'' \in \mathbf{R}$.*
- *the relation $\succ'$ on $\mathbf{R}/\sim$ defined by $X \succ' Y \Leftrightarrow \exists R'$, $R''(R' \succ R''$, $R' \in X$, $R'' \in Y)$, is a linear order, where X and Y are elements of $\mathbf{R}/\sim$.*

A linear order is a weak order in which any two different elements are comparable. This lemma implies that if $\succ$ is a weak order, the indifference relation $\sim$ divides the set of rules into disjoint subsets.

In the measurement-theoretic terminology, the requirement of a weak order indeed suggests the use of an ordinal scale for the measurement of user preference of rules, as shown by the following representation theorem [23]. That is, we can find a real-valued function u as a measure.

Theorem 1 *Suppose $\mathbf{R}$ is a finite non-empty set of rules and $\succ$ a relation on $\mathbf{R}$. There exists a real-valued function $u : \mathbf{R} \longrightarrow R$ satisfying the condition,*

$$R' \succ R'' \Leftrightarrow u(R') > u(R'') \qquad (5)$$

if and only if $\succ$ is a weak order. Moreover, u is defined up to a strictly monotonic increasing transformation.

The numbers $u(R')$, $u(R'')$,... as ordered by $>$ reflect the order of R', R'', ... under $\succ$. The function u is referred to as an order-preserving utility function. It quantifies a user preference relation and provides a measurement of user judgments. According to Theorem 1, the axioms of a weak order are the conditions which allow the measurement. Thus, to see if we can measure a user's preference to the extent of producing an ordinal utility function, we just check if this preference satisfies the conditions of asymmetry and negative transitivity. A rational user's judgments must allow the measurement in terms of a quantitative utility function. On the other hand, another interpretation treats the axioms as testable conditions. Whether can measure the user judgments depends on whether the user preference relation is a weak order [32].

In the above discussion, only the asymmetry and negative transitivity axioms must be satisfied. This implies that the ordinal scale is used for the measurement of user preference. For the ordinal scale, it is meaningful to examine the order or compare the order induced by the utility function.

The main ideas can be illustrated by a simple example. Suppose a user preference relation $\succ$ on a set of rules $\mathbf{R} = \{r_1, r_2, r_3, r_4\}$ is specified by the following weak order:

$$r_3 \succ r_1, \;\; r_4 \succ r_1, \;\; r_3 \succ r_2, \;\; r_4 \succ r_2, \;\; r_4 \succ r_3.$$

This relation $\succ$ satisfies the asymmetry and negative transitivity conditions (axioms). We can find three equivalence classes $\{r_4\}$, $\{r_3\}$, and $\{r_1, r_2\}$. In turn, they can be arranged as three levels:

$$\{r_4\} \succ' \{r_3\} \succ' \{r_1, r_2\}.$$

Obviously, we can defined the utility function u_1 as follows:

$$u_1(r_1) = 0, \;\; u_1(r_2) = 0, \;\; u_1(r_3) = 1, \;\; u_1(r_4) = 2.$$

Another utility function u_2 also may also be used:

$$u_2(r_1) = 5, \;\; u_2(r_2) = 5, \;\; u_2(r_3) = 6, \;\; u_2(r_4) = 7.$$

The two utility functions preserve the same order for any pair of rules, although they use different values.

Based on the formal model of measurement on rules interestingness, we can study different types of user preference relations. In order to do so, we need to impose more axioms on the user preference relation. The axioms on user preference relations can be easily interpreted, and be related to domain specific notions.

5 Conclusion

A critical review of rule evaluation suggests that we can study the topic from different points of views. Each view leads to different perspectives and different issues. It is recognized that there is a need for a unified framework for rule evaluation, in which various notions can be defined and interpreted formally and precisely.

Measurement theory is used to establish a solid foundation for rule evaluation. Fundamental issues are discussed based on the user preference of rules. Conditions on a user preference relation are discussed so that one can obtain a quantitative measure that reflects the ordering of rules by the user.

The proposed framework provides a solid basis for future research. We will investigate additional qualitative properties on the user preference relation. Furthermore, we will identify the qualitative properties on user preference relations that justify the use of many existing measures.

References

[1] R. Agrawal, T. Imielinski and A. Swami, "Mining association rules between sets of items in large databases", *Proceedings of ACM SIGMOD*, 207-216, 1993.

[2] B. Barber and H. Hamilton, "Extracting share frequent itemsets with infrequent subsets", *Data Mining and Knowledge Discovery*, **7**, 153-185, 2003.

[3] M.S. Chen, J. Han, and P.S. Yu, "Data mining: an overview from database perspective", *IEEE Transactions on Knowledge and Data Engineering*, **8**, 866-833, 1996.

[4] J. Doyle, "Rationality and its role in reasoning", *Computational Intelligence*, **8**, 376-409, 1992.

[5] D. Dubois, "Belief structures, possibility theory and decomposable confidence measures on finite sets", *Computers and Artificial Intelligence*, **5**, 403-416, 1986.

[6] U.M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery: an overview", in: U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, (Eds.), *Advances in Knowledge Discovery and Data Mining, AAAI/MIT Press*, 1-34, 1996.

[7] U.M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases", *AI Magazine*, **17**, 37-54, 1996.

[8] W. Frawely, G. Piatetsky-Shapiro, and C. Matheus, "Knowledge discovery in database: an overview", *Knowledge Discovery in Database, AAAI/MIT Press*, 1-27, 1991.

[9] A.A. Freitas, "On rule interestingness measures", *Knowledge-Based Systems*, **12**, 309-315, 1999.

[10] S. French, *Decision Theory: An Introduction to the Mathematics of Rationality*, Ellis Horwood Limited, Chichester, West Sussex, England, 1988.

[11] J. Han and M. Kamber, *Data mining: Concept and Techniques*, Morgan Kaufmann, Palo Alto, CA, 2000.

[12] R.J. Hilderman and H.J. Hamilton, *Knowledge Discovery and Measures of Interest*, Kluwer Academic Publishers, Boston, 2001.

[13] W. Klösgen, "Explora: a multipattern and multistrategy discovery assistant", in: U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, (Eds.), *Advances in Knowledge Discovery and Data Mining*, AAAI Press / MIT Press, California, 249-271, 1996.

[14] D.H. Krantz, R.D. Luce, P. Suppes, and A. Tversky, *Foundations of Measurement*, Academic Press, New York, 1971.

[15] T.Y. Lin, Y.Y. Yao, and E. Louie, "Value added association rules", *Proceedings of PAKDD*, 328-333, 2002.

[16] C. Ling, T. Chen, Q. Yang and J. Chen, "Mining optimal actions for profitable CRM", *Proceedings of ICDM'02*, 767-770, 2002.

[17] B. Liu, W. Hsu, and S. Chen, "Using general impressions to analyze discovered classification rules", *Proceedings of KDD*, 31-36, 1997.

[18] B. Liu, W. Hsu, and Y. Ma, "Identifying nonactionable association rules", *Proceedings of KDD'01*, 329-334, 2001.

[19] J. Major and J. Mangano, "Selecting among rules induced from a hurricane database", *The Journal of Intelligent Information Systems*, **4**, 1995.

[20] H. Mannila, "Methods and problems in data mining", *Proceedings of International Conference on Database Theory*, 41-55, 1997.

[21] J. Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, 1944, 1947, 1953.

[22] G. Piatetsky-Shapiro, "Discovery, analysis, and presentation of strong rules", in: G. Piatetsky-Shapiro, and W.J. Frawley, (Eds.), *Knowledge Discovery in Databases*, AAAI/MIT Press, 229-238, 1991.

[23] F. Roberts, *Measurement Theory*, Addison Wesley, Massachusetts, 1979.

[24] L.J. Savage, *The Foundations of Statistics*, Dover, New York, 1972.

[25] A. Silberschatz and A. Tuzhilin, "On subjective measures of interestingness in knowledge discovery", *Proceedings of KDD*, 275-281, 1995.

[26] A. Silberschatz and A. Tuzhilin, "What makes patterns interesting in knowledge discovery systems", *IEEE Transactions on Knowledge and Data Engineering*, **8**, 970-974, 1996.

[27] E. Simoudis, "Reality check for data mining", *IEEE Expert*, **11**, 26-33, 1996.

[28] K. Wang and Y. He, "User-defined association mining", *Proceedings of PAKDD*, 387-399, 2001.

[29] K. Wang, S. Zhou, and J. Han, "Profit mining: from patterns to actions", *Proceedings of EDBT*, 70-87, 2002.

[30] S.K.M. Wong, Y.Y. Yao, P. Bollmann, and H.C. Bürger, "Axiomatization of qualitative belief structure", *IEEE Transaction on Systems, Man, and Cybernetics*, **21**, 726-734, 1991.

[31] S.K.M. Wong, P. Bollmann, and Y.Y. Yao, "A measurement-theoretic axiomatization of fuzzy sets", *Fuzzy Sets and Systems*, **60**, 295-308, 1993.

[32] Y.Y. Yao, "Measuring retrieval performance based on user preference of documents", *Journal of the American Society for Information Science*, **46**, 133-145, 1995.

[33] Y.Y. Yao, "A step towards the foundations of data mining", *Data Mining and Knowledge Discovery: Theory, Tools, and Technology V, The International Society for Optical Engineering*, 254-263, 2003.

[34] Y.Y. Yao, "Information-theoretic measures for knowledge discovery and data mining", in: Karmeshu (Ed.), *Entropy Measures, Maximum Entropy and Emerging Applications*, Springer, Berlin, 115-136, 2003.

[35] Y.Y. Yao and C.J. Liau, "A generalized decision logic language for granular computing", *FUZZ-IEEE on Computational Intelligence*, 1092-1097, 2002.

[36] Y.Y. Yao, Y. Zhao, and R.B. Maguire, "Explanation-oriented association mining using a combination of unsupervised and supervised learning algorithms", *Proceedings of Advances in Artificial Intelligence*, 527-532, 2003.

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

[38] Y.Y. Yao, N. Zhong, and M. Ohshima, "An analysis of Peculiarity oriented multi-database mining", *IEEE Transactions on Knowledge and Data Engineering*, **15**, 952-960, 2003.