A Measurement-Theoretic Foundation of Rule Interestingness Evaluation

Yiyu Yao, Yaohua Chen, and Xuedong Yang

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

Summary. Many measures have been proposed and studied extensively in data mining for evaluating the interestingness (or usefulness) of discovered rules. They are usually defined based on structural characteristics or statistical information about the rules. The meaningfulness of each measure was interpreted based either on intuitive arguments or mathematical properties. There does not exist a framework in which one is able to represent the user judgment explicitly, precisely, and formally. Since the usefulness of discovered rules must be eventually judged by users, a framework that takes user preference or judgement into consideration will be very valuable. The objective of this 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 interestingness evaluation.

Key words: KDD, Rule Interestingness, Evaluation, Measurement Theory, User Preference

1 Introduction

With rapidly increasing capabilities of accessing, collecting, and storing data, knowledge discovery in databases (KDD) has emerged as a new area of research in computer science. The objective of KDD systems is to extract implicitly hidden, previously unknown, and potentially useful information and knowledge from databases [7]. A core task of the KDD field, called data mining, is the application of specific machine learning algorithms, knowledge representations, statistical methods, and other data analysis techniques for knowledge extraction and abstraction. The discovered knowledge is often expressed in terms of a set of rules. They represent relationships, such as correlation, association, and causation, among concepts [49]. For example, the well-known association rules deal with relationships among sale items [1, 3]. Some fundamental tasks of data mining process in KDD are the discovery, interpretation, and evaluation of those relationships.

There are many types of rules embedded in a large database [47]. Furthermore, the number of rules is typically huge and only a small portion of rules is actually useful [36]. An important problem in data mining is the evaluation of the *interestingness* of the mined rules and filtering out useless rules [36]. Many measures have been proposed and studied to quantify the interestingness (or usefulness) of rules [11, 15, 35, 36, 49]. The results lead to an in-depth understanding of different aspects of rules. It is recognized that each measure reflects a certain characteristic of rules. In addition, many studies investigate and compare rule interestingness measures based on intuitive arguments 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.

We argue that measurement theory can be used to establish a solid foundation for rule interestingness evaluation. The theory provides 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.

The rest of this chapter is organized as follows. In the next section, we introduce the basic notion of evaluation and related issues. A critical review of existing measures of rules interestingness is presented, which reveals some limitations with existing studies. The third section provides motivations to the current study. The fourth section presents an overview of measurement theory. The fifth section applies measurement theory to build a framework of rule interestingness evaluation. Finally, the conclusion in the sixth section gives the summary of this chapter and discusses the future research.

2 Introduction of Evaluation

The discussion of the basic notion of evaluation is aimed at improving our understanding to the rule interestingness evaluation methodologies.

2.1 What is the Evaluation?

Many approaches define the term of evaluation based on specific views [13, 32], such as qualitative assessments and detailed statistical analysis. Suchman analyzes various definitions of evaluation with regard to the conceptual and operational approaches [38]. Simply speaking, the evaluation can be defined as the determination of the results, which are attained by some activity for accomplishing valued goals or objectives. The practice of evaluation can in fact be applied to many processes and research areas, such as the systematic collection of information of programs, personnel and products for reducing uncertainties, improving effectiveness, and making decisions [28].

Three basic components of an evaluation are summarized by Geisler [13]. The first component is the subjects for evaluation, which is what or whom needs to be evaluated. For the discovered rules, the subjects for evaluation are the properties or characteristics of each rule such as the association relationship between sale items and a type of business profit. The formulation of the subjects is always done in the first step of the evaluation procedure. The more the subjects are distinguished precisely, the better the framework and measurement can be produced.

The users who are interested in and willing to perform the evaluation are considered as the second component of an evaluation. Knowing who will participate in judging or who will benefit from the evaluation will help to clarify why the evaluation is performed and which measures or methods of evaluation should be used. Since the qualities of objects or events must be eventually judged by users, an evaluation needs to consider the user judgment. The users can be humans, organizations, or even systems. Different types of participants may have different purposes of conducting an evaluation and lead to different results of an evaluation.

The processes for evaluation and concrete measures are the evaluation's third component. Clarification of the criteria for the measures and designing the implementation for the evaluation are the key points in this component. One must consider the first two components, the subjects and the users, and then develop the processes and measurements of an evaluation. As Suchman points out, an evaluation can be constructed for different purposes, by different methods, with different criteria with respect to different users and subjects [38].

2.2 How to Do the Evaluation?

According to the definition of evaluation, the procedure of evaluation can be simply and generally described as follows [13, 38]:

- Identification of the subjects to be evaluated.
- Collection of data for the evaluation.
- Users analyze and measure those data to summarize their judgments based on the criteria and conduct the process of the evaluation for decision making.

The real procedures of an evaluation can be very complicated and might be iterative [38]. Furthermore, identifying and accurately measuring or quantifying the properties of subjects is very difficult to achieve. More often than not, an approximate approach can be accepted by general users. In the processes of an evaluation, it is very important that users determine an appropriate evaluation as the means of measuring.

2.3 Measurement of Evaluation

During the procedure of an evaluation, the measurement always plays a crucial role and the measurement theory provides the necessary concepts and methodologies for the evaluation. The subjects of measurement in measurement theory are about estimating the attributes or properties of empirical objects or events, such as weight, color, or intelligence [29]. The measurement can be performed by assigning numbers to the objects or events in order that the properties or attributes can be represented as numerical properties [17]. In other words, the properties of the quantity are able to faithfully reflect the properties of objects or events to be evaluated.

2.4 Subjectivity of Evaluation

From the discussion of the definition and procedure of evaluation, it is recognized that evaluation is an inherently subjective process [38]. The steps, methods, and measures used in an evaluation depend on the users who participate in the evaluation. The selection of the criteria and measures reflects the principles and underlying beliefs of the users [13].

Mackie argues that subjective values are commonly used when one evaluates objects, actions, or events [24]. Objectivity is only related to the objective measures and implementation of the measurement. People always judge the subjects with their subjective interests. Different people have different judgments on the same object, action, or event because they always stand on their own interests or standards of evaluations. In other words, the objective measurement is relative to personal standards of evaluations. In this regard, there are no absolutely objective evaluations, only relatively objective evaluations for human beings.

Nevertheless, these standards of evaluations can be derived from human being's subjective interests. In fact, the user preference is indeed realized as a very important issue for an evaluation to occur [13]. It can be described as the user's discrimination on two different objects rationally [23]. The users can simply describe their preference as "they act upon their interests and desires they have" [6]. In measurement and decision theories, user preferences are used to present the user judgments or user interests and can be viewed as the standards of an evaluation [12, 23, 29, 33]. The user preference or judgment should be considered in the process of an evaluation.

3 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 the studies on rule evaluation in order to observe their difficulties. This leads to a new direction for future research.

3.1 A Model of Data Mining based on Granular Computing

In an information table, objects can be described by the conjunctions of attribute-value pairs [50]. The rows of the table represent the objects, the columns denote a set of attributes, and each cell is the value of an object on an attribute. In the model of granular computing, the objects in an information table are viewed as the universe and the information table can be expressed by a quadruple [46, 50]:

$$T = (U, At, \{V_a \mid a \in At\}, \{I_a \mid a \in At\}), \tag{1}$$

where U is a finite nonempty set of objects, At is a finite nonempty set of attributes, V_a is nonempty set of values for $a \in At$, and I_a is a function to map from U to V_a, that is, $I_a : U \rightarrow V_a$.

With respect to the notion of tables, we define a decision logic lanaguage [31]. In this language, an atomic formula is a pair (a, v), where $a \in At$ and $v \in V_a$. If ϕ and ψ are formulas, then $\neg\phi$, $\phi \wedge \psi$, $\phi \vee \psi$, $\phi \rightarrow \psi$, and $\phi \equiv \psi$ are also formulas. The set of objects that satisfy a formula ϕ are denoted by $m(\phi)$. Thus, given an atomic formula (a, v), the corresponding set of objects can be $m(a, v) = \{x \in U \mid I_a(x) = v\}$. The following properties hold:

(1) $m(\neg\phi) = \neg m(\phi)$,
(2) $m(\phi \wedge \psi) = m(\phi) \cap m(\psi)$,
(3) $m(\phi \vee \psi) = m(\phi) \cup m(\psi)$,
(4) $m(\phi \rightarrow \psi) = \neg m(\phi) \cup m(\psi)$,
(5) $m(\phi \equiv \psi) = (m(\phi) \cap m(\psi)) \cup (\neg m(\phi) \cap \neg m(\psi))$.

The formula ϕ can be viewed as the description of the set of objects $m(\phi)$.

In formal concept analysis, every concept consists of the intention and the extension [41, 42]. A set of objects is referred to as the extension, and the corresponding set of attributes as the intention of a concept. Therefore, a formula ϕ can represent the intention of a concept and a subset of objects $m(\phi)$ can be the extension of the concept. The pair $(\phi, m(\phi))$ is denoted as a concept.

One of the important functions of data mining of KDD is to find the strong relationships between concepts [49]. A rule can be represented as $\phi \Rightarrow \psi$, where ϕ and ψ are intensions of two concepts [46]. The symbol $\Rightarrow$ in the rules are interpreted based on the types of knowledge and rules can be classified according to the interpretations of $\Rightarrow$. In other words, different kinds of rules represent different types of knowledge extracted from a large database. Furthermore, based on the extensions $m(\phi)$, $m(\psi)$, and $m(\phi \wedge \psi)$, various quantitative measures can be used for the rules evaluation. A systematic analysis of quantitative measures associated with rules is given by Yao and Zhong [49].

3.2 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 many phases [7, 14, 22, 26, 37, 54]. 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 [7, 8]. 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 of roles. In the data mining phase, quantitative measures can be used to reduce the size of search space. An example is the use of the well known *support* measure, which reduces the number of item sets 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 [35, 36]. For example, the *confidence* measure of association rules is used to select only strongly associated item sets [1]. In fact, many measures associated with rules are used for such a purpose [49]. 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 [11]. 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 [47]. They are the philosophy level, technique level, and application level. The philosophy 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 question: 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 question: 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 plays the similar roles. In the philosophy level, quantitative measures can be used to characterize and classify different types of rules. In the technique 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 [35]. 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 [35]. 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 [35, 36].

Statistical, structural vs. semantic measures

Many measures, such as support, confidence, independence, classification error, etc., are defined based on statistical characteristics of rules. A systematic analysis of such measures can be performed by using a 2×2 contingency table induced by a rule [49, 51].

The structural characteristics of rules have been considered in many measures. For example, information, such as the size of disjunct (rule), attribute interestingness, the asymmetry of classification rules, etc., can be used [11]. 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 semantic aspect of rules or explanations of rules [53]. 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-based or utility-based mining is one example of a special kind of constraint-based mining, taking into account both statistical significance and profit significance [18, 40]. Doyle discusses the importance and usefulness of the 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 low profit or high risk. For example, Barber and Hamilton propose the notion of share measure which considers the contribution, in terms of profit, of an item in an item set [2].

Actionable rule mining is another example of dealing with profit-driven actions required by business decision making [19, 21]. 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 philosophical 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) [52].

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 a 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 [48]. 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 summarization of a database [15].

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 [30] suggests 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 size of E and H. For the evaluation of rules, Piatetsky-Shapiro [30] introduces three axioms. Major and Mangano [25] add the fourth axioms. Klösgen [16] studies 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 [16] 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 [15].

3.3 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 and each measure reflects certain aspects of rules. Quantitative measures are typically interpreted by 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, interpret, and compare different measures.

A very interesting 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 usage. 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.

4 Overview of Measurement Theory

For completeness, we 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.* [17], Roberts [33] and French [12].

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 [17, 29]. The properties are usually described by certain qualitative relations and operations. Consider an example discussed by Krantz *et al.* [17]. 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 $f(a)$ and $f(b)$ with $f(a) > f(b)$ to reflect the results of the comparison. That is, we require the numbers assigned to the individual rods satisfy the condition: for all $a, b \in U$,

$$a \succ b \Longleftrightarrow f(a) > f(b). \tag{2}$$

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 (2), the numbers assigned must also satisfy the following condition: for all $a, b \in U$,

$$f(a \circ b) = f(a) + f(b). \tag{3}$$

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 description of the basic notions of measurement theory in the above example, some basic concepts and notations are introduced and the formal definitions and formulations of the theory are reviewed.

4.1 Relational Systems

Suppose U is a set. The Cartesian product of U with U, denoted $U \times U$, is a set of all ordered pairs (a, b) so that $a, b \in U$. A binary relation R on a set U, simply denote (U, R), is a subset of the Cartesian product $U \times U$. For $a, b \in U$, if a is related to b under R, we write aRb or $(a, b) \in R$. For example, consider the binary relation "less than" ($<$) relation on real numbers. An ordered pair (a, b) is in the binary relation if and only if $a < b$. Similarly, "greater than" and "equals" also can be defined as the binary relations on real numbers.

With the set U, a function $f : U \to U$ can in fact also be thought of as a binary relation (U, R). A function $f : U^n \to U$ can be an (n+1)-ary relation (U, R). The functions from U into U is called binary operations, or just operations for short. For example, for addition (+), given a pair of real numbers a and b, there exists a third real number c so that $a + b = c$.

A *relational system* (*structure*) is a set of one or more relations (operations) on an arbitrary 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. If the binary operations are considered as a special type of relations, a relational system can be simply denoted as a $(p+1)$-tuple $\mathcal{A} = (U, R_1, \ldots, R_p)$. For convenience, we separate the operations from other relations.

If U is the set (or a subset) of real numbers, such a relational system is called as a numerical relational systems. As illustrated by the example of rigid rods, for measuring the property of length, we can 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 $\mathcal{A}$.

4.2 Axioms of the Empirical System

Based on the definitions of the relations and operations in the relation systems, we should describe the valid use or properties of these relations and operations in order to find the appropriate corresponding numerical systems. Many properties are common to well-defined relations. The consistency properties to be preserved are known as *axioms*. For example, if U is a set of real numbers and R is the relation of "equality" on U, R is reflexive, symmetric, and transitive. However, if U is the set of people in the real world and R is the relation "father of" on U, R is irreflexive, asymmetric, and nontransitive.

The set of axioms characterizing the relations in an empirical system should be complete so that every consistency property for the relations that is required is either in the list or deducible from those in the list [12, 17, 33].

4.3 Homomorphism of Relational Systems

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 $a_1, \ldots, a_{r_i} \in \mathcal{A}$,

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

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

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

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 $f(\cdot)$ 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 (2) and (3).

In general, a measurement has been performed if a homomorphism can be assigned from an empirical (observed) relational system $\mathcal{A}$ to a numerical relational system $\mathcal{B}$. The homomorphism is said to give a *representation*, and the triple $(\mathcal{A}, \mathcal{B}, f)$ of the empirical relational system $\mathcal{A}$, the numerical relational system $\mathcal{B}$, and the function f is called a *scale* or *measure*. Sometimes, a homomorphism from an empirical relational system into the set of real numbers is referred alone as a scale (measure).

With given numerical scales (measures), new scales or measures defined in terms of the old ones are called derived scales or derived measures. For example, density d can be defined in terms of mass m and volume v as $d = m/v$. The density d is the derived scale (measure), and the mass m and volume v are called as primitive scales (measures).

4.4 Procedure of Measurement

Generally, there are three fundamental steps in measurement theory [12, 17, 33]. 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. The axioms of the empirical system are determined. The next step 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.

The next question concerns the uniqueness of the scale. A uniqueness theorem is generally obtained by identifying a set of *admissible transformations*. If $f(\cdot)$ is a scale representing an empirical system and if $\lambda(\cdot)$ is an admissible transformation, then $\lambda(f(\cdot))$ is also a scale representing the same empirical system.

If the truth (falsity) of a numerical statement involving a scale or measure 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 [27, 34], the axiomatization of possibility functions [5] and the axiomatization of belief functions [43].

5 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. Therefore, the key issue of rules evaluation is in fact the measurement of rules' usefulness or interestingness expressed by a user preference relation. According to the procedure of measurement, for rule evaluation, we follow the three steps to seek a quantitative representation of an empirical system.

5.1 User Preference Relations

In the measurement theory, the user judgment or user preference can be modeled as a kind of binary relation, called user preference relation [33]. If the user prefers a rule to another rule, then we can say that one rule is more useful or interesting than the other rule.

Assume we are given a set of discovered rules. Let $\mathbf{R}$ be a set of rules. Since the usefulness or interestingness of rules should be finally judged by users, we focus on user preference relation as a binary relation on the set of discovered rules. 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 denote it by $r' \succ r''$. That is,

$$r' \succ r'' \Leftrightarrow \text{ the user prefers } r' \text{ to } r''. \tag{4}$$

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:

$$r' \sim r'' \Leftrightarrow (\neg(r' \succ r''), \neg(r'' \succ r')). \tag{5}$$

The empirical relational system can be defined as following:

Definition 1. *Given a set of discovered rules* $\mathbf{R}$ *and user preference* $\succ$*, the pair* $(\mathbf{R}, \succ)$ *is called the (empirical) relational system of the set of discovered rules.*

The user judgment on rules can be formally described by a user preference relation $\succ$ on $\mathbf{R}$. 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.

5.2 Axioms of User Preference Relations

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 requires that a user cannot prefer r' to r'' and at the same time prefers r'' to r'. In other words, the result of a user preference on two different discovered rules is not contradictive. In fact, this axiom ensures the user preference or user judgement is rational. 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'''.

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.

5.3 Homomorphism based on Real-valued Function

In the measurement-theoretic terminology, the requirement of a weak order indeed suggests the use of an ordinal scale (homomorphism) for the measurement of user preference of rules, as shown by the following representation theorem [33]. That is, we can find a real-valued function u as a measure.

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

$$r' \succ r'' \Leftrightarrow u(r') > u(r'') \tag{6}$$

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 [45].

5.4 Ordinal Measurement of Rules Interestingness

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:

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

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:

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

Another utility function u_2 also may also be used:

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

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 related to domain specific notions.

Luce and Suppes discuss the user preference and the closely related areas of utility and subjective probability from mathematical psychology point of view [23]. The utility is defined as a type of property of any object, whereby it tends to produce benefit, advantage, pleasure, good, and happiness, or to prevent the happening of mischief, pain, evil, or unhappiness. In other words, utility is a type of subjective measure, not objective measure. The utility of an item depends on the user preference and differs among the individuals. In the theory of decision making, utility is viewed as essential elements of a user preference on a set of decision choices or candidates [9, 12].

6 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 user-preferred ordering of rules.

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 massive databases", *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, 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. P.A. Facione, D. Scherer, and T. Attig, *Values and Society: An Introduction to Ethics and Social Philosophy*, Prentice-Hall, Inc., New Jersey, 1978.
7. 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.
8. U.M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “From data mining to knowledge discovery in databases”, *AI Magazine*, **17**, 37-54, 1996.
9. P.C. Fishburn, *Utility Theory for Decision Making*, John Wiley & Sons, Inc., New York, 1970.
10. W. Frawely, G. Piatetsky-Shapiro, and C. Matheus, “Knowledge discovery in database: an overview”, *Knowledge Discovery in Database, AAAI/MIT Press*, 1-27, 1991.
11. A.A. Freitas, “On rule interestingness measures”, *Knowledge-Based Systems*, **12**, 309-315, 1999.
12. S. French, *Decision Theory: An Introduction to the Mathematics of Rationality*, Ellis Horwood Limited, Chichester, West Sussex, England, 1988.
13. E. Geisler, *Creating Value with Science and Technology*, Quorum Books, London, 2001.
14. J. Han and M. Kamber, *Data mining: Concept and Techniques*, Morgan Kaufmann, Palo Alto, CA, 2000.
15. R.J. Hilderman and H.J. Hamilton, *Knowledge Discovery and Measures of Interest*, Kluwer Academic Publishers, Boston, 2001.
16. 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*, 249-271, 1996.
17. D.H. Krantz, R.D. Luce, P. Suppes, and A. Tversky, *Foundations of Measurement*, Academic Press, New York, 1971.
18. T.Y. Lin, Y.Y. Yao, and E. Louie, “Value added association rules”, *Advances in Knowledge Discovery and Data Mining, Proceedings of 6th Pacific-Asia Conference (PAKDD 2002)*, 328-333, 2002.
19. C. Ling, T. Chen, Q. Yang and J. Chen, “Mining optimal actions for profitable CRM”, *Proceedings of the 2002 IEEE International Conference on Data Mining (ICDM 2002)*, 767-770, 2002.
20. B. Liu, W. Hsu, and S. Chen, “Using general impressions to analyze discovered classification rules”, *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining (KDD-97)*, 31-36, 1997.
21. B. Liu, W. Hsu, and Y. Ma, “Identifying non-actionable association rules”, *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 329-334, 2001.
22. C. Liu, N. Zhong, and S. Ohsuga,“ A multi-agent based architecture for distributed KDD process”, *Foundations of Intelligent Systems, Proceedings of 12th International Symposium (ISMIS 2000)*, 591-600, 2000
23. R.D. Luce and P. Suppes, “Preference, utility, and subjective probability”, in: R.D. Luce, R.R. Bush and E. Galanter (Eds.), *Handbook of Mathematical Psychology, John Wiley and Sons, Inc., New York*, 249-410, 1965.

24. Machie, J.L. *Ethics: Inventing Right and Wrong*, Penguin Books Ltd., Harmondsworth, 1977.
25. J. Major and J. Mangano, "Selecting among rules induced from a hurricane database", *The Journal of Intelligent Information Systems*, **4**, 1995.
26. H. Mannila, "Methods and problems in data mining", *Database Theory, Proceedings of 6th International Conference (ICDT-97)*, 41-55, 1997.
27. J. Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, 1944, 1947, 1953.
28. M.Q. Patton, *Practical Evaluation*, Sage Publications, Newbury Park, 1982.
29. J. Pfanzagl, *Theory of Measurement*, John Wiley & Sons, New York, 1968.
30. 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.
31. Z. Pawlak, *Rough Sets, Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Dordrecht, 1991.
32. D. Reith, "Evaluation, a function of practice", in: J. Lishman, (Ed.), *Evaluation, 2nd Edition*, Kingsley Publishers, London, 23-39, 1988.
33. F. Roberts, *Measurement Theory*, Addison Wesley, Massachusetts, 1979.
34. L.J. Savage, *The Foundations of Statistics*, Dover, New York, 1972.
35. A. Silberschatz and A. Tuzhilin, "On subjective measures of interestingness in knowledge discovery", *Proceedings of the 1st International Conference on Knowledge Discovery and Data Mining (KDD-95)*, 275-281, 1995.
36. A. Silberschatz and A. Tuzhilin, "What makes patterns interesting in knowledge discovery systems", *IEEE Transactions on Knowledge and Data Engineering*, **8**, 970-974, 1996.
37. E. Simoudis, "Reality check for data mining", *IEEE Expert*, **11**, 26-33, 1996.
38. E.A. Suchman, *Evaluation Research*, Russell Sage Foundation, USA, 1967.
39. K. Wang and Y. He, "User-defined association mining", *Knowledge Discovery and Data Mining, Proceedings of 5th Pacific-Asia Conference (PAKDD 2001)*, 387-399, 2001.
40. K. Wang, S. Zhou, and J. Han, "Profit mining: from patterns to actions", *Advances in Database Technology, Proceedings of 8th International Conference on Extending Database Technology (EDBT 2002)*, 70-87, 2002.
41. Wille, R. Concept lattices and concept knowledge systems, *Computers Mathematics with Applications*, **23**, 493-515, 1992.
42. Wille, R. Restructuring lattice theory: an approach based on hierarchies of concepts, in: Ivan Rival (ed.), *Ordered sets*, Reidel, Dordecht-Boston, 445-470, 1982.
43. 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.
44. 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.
45. Y.Y. Yao, "Measuring retrieval performance based on user preference of documents", *Journal of the American Society for Information Science*, **46**, 133-145, 1995.
46. Y.Y. Yao, "Modeling data mining with granular computing", *Proceedings of the 25th Annual International Computer Software and Applications Conference*, 638-643, 2001.

47. 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.
48. 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.
49. Y.Y. Yao and N. Zhong, "An analysis of quantitative measures associated with rules", *Methodologies for Knowledge Discovery and Data Mining, Proceedings of the 3rd Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD-99)*, 479-488, 1999.
50. Y.Y. Yao. and N. Zhong, "Potential applications of granular computing in knowledge discovery and data mining", *Proceedings of World Multiconference on Systemics, Cybernetics and Informatics*, 573-580, 1999.
51. Y.Y. Yao and C.J. Liau, "A generalized decision logic language for granular computing", *FUZZ-IEEE on Computational Intelligence*, 1092-1097, 2002.
52. 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.
53. Y.Y. Yao, Y. Zhao, and R.B. Maguire, "Explanation-oriented association mining using a combination of unsupervised and supervised learning algorithms", *Advances in Artificial Intelligence, Proceedings of the 16th Conference of the Canadian Society for Computational Studies of Intelligence (AI 2003)*, 527-532, 2003.
54. N. Zhong, C. Liu, and S. Ohsuga, "Dynamically organizing KDD processes", *Journal of Pattern Recognition and Artificial Intelligence* **15**, 451-473, 2001.