Available online at www.sciencedirect.com

ScienceDirect

Data & Knowledge Engineering 65 (2008) 373–400

DATA & KNOWLEDGE ENGINEERING

www.elsevier.com/locate/datak

On the evaluation of the decision performance of an incomplete decision table

Yuhua Qian [a,b,c], Chuangyin Dang [b], Jiye Liang [a,c,*], Haiyun Zhang [a,c], Jianmin Ma [d]

[a] Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Taiyuan 030006, China
[b] Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, Hong Kong
[c] School of Computer and Information Technology, Shanxi University, Taiyuan, 030006 Shanxi, China
[d] Institute for Information and System Sciences, Faculty of Science, Xi'an Jiaotong University, Xi'an 710049, China

Received 1 June 2007; received in revised form 10 December 2007; accepted 10 December 2007
Available online 31 December 2007

Abstract

As two classical measures, approximation accuracy and consistency degree can be extended for evaluating the decision performance of an incomplete decision table. However, when the values of these two measures are equal to zero, they cannot give elaborate depictions of the certainty and consistency of an incomplete decision table. To overcome this shortcoming, we first classify incomplete decision tables into three types according to their consistency and introduce four new measures for evaluating the decision performance of a decision-rule set extracted from an incomplete decision table. We then analyze how each of these four measures depends on the condition granulation and decision granulation of each of the three types of incomplete decision tables. Experimental analyses on three practical data sets show that the four new measures appear to be well suited for evaluating the decision performance of a decision-rule set extracted from an incomplete decision table and are much better than the two extended measures.

Keywords: Incomplete decision tables; Decision rules; Information granulation; Decision evaluation

1. Introduction

Rough set theory proposed by Pawlak in [35] is a relatively new soft computing tool for the analysis of a vague description of an object, and has become a popular mathematical framework for pattern recognition, image processing, feature selection, neuro computing, conflict analysis, decision support, data mining and knowledge discovery from large data sets [1,34,37–41]. The indiscernibility relation constitutes a mathematical basis of rough set theory [42]. It induces a partition of the universe into blocks of indiscernible objects, called elementary sets, which can be used to build knowledge about a real or abstract world [32,36,43,45,55].

* Corresponding author. Address: School of Computer and Information Technology, Shanxi University, Taiyuan, 030006 Shanxi, China.

E-mail addresses: jinchengqyh@126.com (Y. Qian), mecdang@cityu.edu.hk (C. Dang), ljy@sxu.edu.cn (J. Liang), seaddy@sxu.edu.cn (H. Zhang), majm@mail.xjtu.edu.cn (J. Ma).

doi:10.1016/j.datak.2007.12.002

Rough set-based data analysis starts from a data table, which is also called an information system and contains data about objects of interest that are characterized by a finite set of attributes. According to whether or not there are missing data (null values), information systems can be classified into two categories: complete and incomplete [2,3]. By an incomplete information system we mean a system with missing data (null values) [11,12]. In this paper, we will only deal with the case of unknown values in which a null value may be some value in the domain of the corresponding attribute [17,18,20,22,23,44]. For the case that a null value means an inapplicable value, it can be handled by adding to the attribute domains a special symbol for the inapplicable value. For an incomplete information system, if condition attributes and decision attributes are distinguished, then it is called an incomplete decision table. In general, one can extract some decision rules from an incomplete decision table [13,15].

For further developments, as follows, we briefly review some methods for rule extracting from incomplete decision tables. In the literature (e.g. CN2 [6], RIPPER [7], C4.5 [48]), many algorithms have been developed for learning a set of classification rules from a number of observations with their corresponding class labels (a decision table). Several solutions to the problem of generating a decision tree from a training set of examples with unknown values have been proposed in the area of artificial intelligence (AI). The simplest among them is to remove the examples with unknown values or replace the unknown values with the most common values. More complex approaches were presented in [14,47]. The problem of decision-rules generation from incomplete information systems was also investigated in the context of Rough Set Theory [5,16,52]. Modelling by means of fuzzy sets the uncertainty caused by the appearance of unknown values was discussed in [52]. Two methods of treating unknown values are available in the LERS system [5]. The methodology proposed in [16] allows to generate generalized rules directly from the original incomplete decision table. In [18], another method of computing all certain rules from an incomplete information decision table was presented, which does not require changing the size of the original system. In recent years, the generation of all optimal certain rules or a class of optimal certain rules from an incomplete decision table was also investigated in [21,24]. For decision problems in rough set theory, by various kinds of reduct techniques, a set of decision rules can be generated from a decision table for classification and prediction using information granules [14,49,56]. In the past twenty years, many kinds of reduct techniques for information systems and decision tables have been proposed in rough set theory [4,19,23,29–31,33,36,37,46,50,51,53–55,57]. β-reduct proposed by Ziarko provides a kind of attribute reduction methods in the variable precision rough set model [54]. α-reduct and α-relative reduct that allow the occurrence of additional inconsistency were proposed in [33] for information systems and decision tables, respectively. An attribute reduction method that preserves the class membership distribution of all objects in information systems was proposed by Slezak in [50,51]. Five kinds of attribute reducts and their relationships in inconsistent systems were investigated by Kryszkiewicz [19], Li et al. [24] and Mi et al. [31], respectively. By eliminating some rigorous conditions required by the distribution reduct, a maximum distribution reduct was introduced by Mi et al. in [31]. Unlike the possible reduct [24], the maximum distribution reduct can derive decision rules that are compatible with the original system.

Generally speaking, a set of decision rules can be generated from a decision table by adopting any kind of rule-extracting methods mentioned above. In recent years, how to evaluate the decision performance of a decision rule has become a very important issue in rough set theory. In [8], based on information entropy, Düntsch and Gediga suggested some uncertainty measures of a decision rule and proposed three criteria for model selection. In [10], Greco et al. applied some well-known confirmation measures within the rough set approach to discover relationships in data in terms of decision rules. For a decision-rule set consisting of every decision-rule induced from a decision table, three parameters are traditionally associated: the strength, the certainty factor and the coverage factor of the rule [10]. In many practical decision problems, we always adopt several rule-extracting methods for the same decision table. In this case, it is very important to check whether or not each of the rule-extracting approaches adopted is suitable for the given decision table. In other words, it is desirable to evaluate the decision performance of the decision-rule set extracted by each of the rule-extracting approaches. This strategy can help a decision maker to determine which of rule-extracting methods is preferred for a given decision table. However, all of the above measures for this purpose are only defined for a single decision rule and are not suitable for evaluating the decision performance of a decision-rule set. There are two more kinds of measures in the literature [36,39], which are approximation accuracy for decision classification and consistency degree for a decision table. Although these two measures, in some sense, could be regarded as measures for evaluating the decision performance of all decision-rules generated from a complete decision table, they have some limitations. For instance, the certainty and consistency of a rule set could not be well characterized by the approximation accuracy and consistency degree when their values reaches zero. As we know, when the approximation accuracy or consistency degree is equal to zero, it is only implied that there is no decision rule with the certainty of one in the complete decision table. This shows that the approximation accuracy and consistency degree of a complete decision table cannot give elaborate depictions of the certainty and consistency for a rule set. To overcome the shortcomings of the existing measures, three new evaluation measures were proposed for evaluating the decision performance of a set of decision-rules extracted from a complete decision table, which are certainty measure (α), consistency measure (β) and support measure (γ) [44]. To date, however, how to assess the decision performance of a decision-rule set extracted from an incomplete decision table has not been reported. Like the measures (α, β and γ), the certainty, consistency and support of a decision-rule set extracted from an incomplete decision table should be also studied to assess their decision performance. Moreover, the degree of the cover on the universe induced by the missing values in the condition part is also an important factor that affects the decision performance of a decision-rule set extracted from an incomplete decision table. In fact, the approximation accuracy and consistency degree can be extended for evaluating the decision performance of an incomplete decision table. Nevertheless, these two extensions have the same limitations, which still cannot give elaborate depictions of the certainty and consistency of a decision-rule set extracted from an incomplete decision table. To overcome this drawback, this paper introduces four new measures for evaluating the decision performance of a set of decision-rules extracted from an incomplete decision table, which are certainty measure (α), consistency measure (β), support measure (γ) and cover measure (ϑ).

The rest of this paper is organized as follows. Some preliminary concepts such as incomplete information systems, incomplete decision tables, the maximal consistent block technique and partial relation are briefly recalled in Section 2. In Section 3, some new concepts and two lemmas for further developments are introduced, which show how to classify incomplete decision tables into three types. In Section 4, through some examples, the limitations of the two extended measures are revealed. In Section 5, four new measures (α, β, γ and ϑ) are introduced for evaluating the decision performance of a set of rules extracted from an incomplete decision table, it is analyzed how each of these four measures depends on the condition granulation and decision granulation of each of the three types of incomplete decision tables, and experimental analyses of each of the measures (α, β and γ) are performed on three practical data sets. Finally, Section 6 concludes this paper with some remarks and discussion.

2. Preliminaries

In this section, we review some basic concepts such as incomplete information systems, incomplete decision tables, maximal consistent block technique and partial relation.

An information system is a pair $S = (U, A)$, where

(1) U is a non-empty finite set of objects;
(2) A is a non-empty finite set of attributes;
(3) for every $a \in A$, there is a mapping $a : U \to V_a$, where V_a is called the value set of a.

Each subset of attributes $P \subseteq A$ determines a binary indistinguishable relation $\mathrm{IND}(P)$ given by

$$\mathrm{IND}(P) = \{(u, v) \in U \times U \mid \forall a \in P, a(u) = a(v)\}.$$

It can be shown that $\mathrm{IND}(P)$ is an equivalence relation on the set U. For $P \subseteq A$, the relation $\mathrm{IND}(P)$ constitutes a partition of U, which is denoted by $U/\mathrm{IND}(P)$, or just U/P.

It may happen that some of the attribute values for an object are missing. For example, in medical information systems there may exist a group of patients for which it is impossible to perform all the required tests. These missing values can be represented by the set of all possible values for the attribute. To indicate such a situation, a distinguished value (the so-called null value) is usually assigned to those attributes.

If V_a contains a null value for at least one attribute $a \in A$, then S is called an incomplete information system; otherwise it is complete [17,18]. From now on, we will denote the null value by $*$. Let $S = (U, A)$ be an information system and $P \subseteq A$ an attribute set.

We define a binary relation on U by

$$\mathrm{SIM}(P) = \{(u, v) \in U \times U | \ \forall a \in P, \ a(u) = a(v) \text{ or } a(u) = * \text{ or } a(v) = *\}.$$

In fact, $\mathrm{SIM}(P)$ is a tolerance relation on U. The concept of a tolerance relation has a wide variety of applications in classifications [19,23]. It can be easily shown that $\mathrm{SIM}(P) = \bigcap_{a \in P} \mathrm{SIM}(\{a\})$. Let $S_P(u)$ denote the set $\{v \in U \mid (u, v) \in \mathrm{SIM}(P)\}$. Then, $S_P(u)$ is the maximal set of objects which are possibly indistinguishable by P with u. Let $U/\mathrm{SIM}(P)$ denote the family sets $\{S_P(u) \mid u \in U\}$, which are the classification or the knowledge induced by P. A member $S_P(u)$ from $U/\mathrm{SIM}(P)$ will be called a tolerance class or a granule of information. It should be noticed that the tolerance classes in $U/\mathrm{SIM}(P)$ do not constitute a partition of U in general. They constitute a cover of U, i.e., $S_P(u) \neq \emptyset$ for every $u \in U$, and $\bigcup_{u \in U} S_P(u) = U$.

An incomplete information system $S = (U, C \cup D)$ is called an incomplete decision table if condition attributes and decision attributes are distinguished, where C is the condition attribute set and D is the decision attribute set. This is illustrated in the following example:

Example 1. Consider the descriptions of several cars in Table 1 [18]. This is an incomplete decision table, where $U = \{u_1, u_2, u_3, u_4, u_5, u_6\}$, $C = \{a_1, a_2, a_3, a_4\}$ with a_1 – price, a_2 – mileage, a_3 – size, a_4 – max-speed, and $D = \{d\}$. By computing, it follows that

$$U/\mathrm{SIM}(C) = \{S_C(u_1), S_C(u_2), S_C(u_3), S_C(u_4), S_C(u_5), S_C(u_6)\},$$

where $S_C(u_1) = \{u_1\}$, $S_C(u_2) = \{u_2, u_6\}$, $S_C(u_3) = \{u_3\}$, $S_C(u_4) = \{u_4, u_5\}$, $S_C(u_5) = \{u_4, u_5, u_6\}$ and $S_C(u_6) = \{u_2, u_5, u_6\}$.

It is easy to observe from Table 1 that the value of the generalized decision ∂_d for an object in an incomplete decision table is the superset of the object's value (see ∂_d in Table 1).

Now we define a partial order on the set of all classifications of U. Let $S = (U, A)$ be an incomplete information system, and $P, Q \subseteq A$. We say that Q is coarser than P (or P is finer than Q), denoted by $P \preceq Q$, if and only if $S_P(u_i) \subseteq S_Q(u_i)$ for $\forall i \in \{1, 2, \ldots, |U|\}$. If $P \preceq Q$ and $P \neq Q$, we say that Q is strictly coarser than P (or P is strictly finer than Q) and denoted by $P \prec Q$. In fact, $P \prec Q \Longleftrightarrow$ for $\forall i \in \{1, 2, \ldots, |U|\}$, $S_P(u_i) \subseteq S_Q(u_i)$, and $\exists j \in \{1, 2, \ldots, |U|\}$ such that $S_P(u_j) \subset S_Q(u_j)$.

In general, the tolerance classes are used to describe knowledge or information in incomplete information systems, however, as it has been pointed out in [9,23], they are not the minimal units. Let $S = (U, A)$ be an information system, $P \subseteq A$ an attribute set and $X \subseteq U$ a subset of objects. We say X is consistent with respect to P if $(u, v) \in \mathrm{SIM}(P)$ for any $u, v \in X$. If there does not exist a subset $Y \subseteq U$ such that $X \subset Y$, and Y is consistent with respect to P, then X is called a maximal consistent block of P.

Obviously, in a maximal consistent block, all objects are not indiscernible with available information provided by a similarity relation [23]. Henceforth, we denote by MC_P the set of all maximal consistent blocks determined by $P \subseteq A$, and by $MC_P(u)$ the set of all maximal consistent blocks of P which includes some object $u \in U$, respectively. It is clear that $X \in MC_P$ if and only if $X = \bigcap_{u \in X} S_P(u)$ [23]. This is illustrated in Example 2. In fact, the set of all the maximal consistent blocks, MC_P, will degenerate into the partition U/P induced by the attribute set P in a complete information system, i.e., $MC_P = U/P$.

Table 1
The incomplete decision table about car [18]

Car	Price	Mileage	Size	Max-speed	d	∂_d
u_1	High	Low	Full	Low	Good	{good}
u_2	Low	$*$	Full	Low	Good	{good}
u_3	$*$	$*$	Compact	Low	Poor	{poor}
u_4	high	$*$	Full	High	Good	{good, excellent}
u_5	$*$	$*$	Full	High	Excellent	{good, excellent}
u_6	Low	High	Full	$*$	Good	{good, excellent}

Example 2. Compute all the maximal consistent blocks of C in Table 1. From Example 1, it follows that

$$MC_C = \{\{u_1\}, \{u_2, u_6\}, \{u_3\}, \{u_4, u_5\}, \{u_5, u_6\}\},$$

where MC_C is the set of all the maximal consistent blocks determined by C on U.

Next, we define another partial relation in incomplete information systems. Let $S = (U, A)$ be an incomplete information system, $P, Q \subseteq A$, $MC_P = \{P_1, P_2, \ldots, P_m\}$ and $MC_Q = \{Q_1, Q_2, \ldots, Q_n\}$. Then, a partial relation $\preceq'$ is defined as follows:

$P \preceq' Q \iff$ for every $P_i \in MC_P$, there exists $Q_j \in MC_Q$ such that $P_i \subseteq Q_j$.

If $P \preceq' Q$ and $P \neq Q$, i.e., for some $P_{i_0} \in MC_P$, there exists $Q_{j_0} \in MC_Q$ such that $P_{i_0} \subset Q_{j_0}$, then we denote it as $P \prec' Q$.

3. Decision rule and information granulation in incomplete decision tables

In the first part of this section, we briefly review the notions of decision rules and certainty measure, support measure and converge measure of a decision rule in incomplete decision tables.

The knowledge hidden in incomplete decision tables may be discovered and expressed in the form of decision rules: $t = \wedge(a, v)$, $a \in C$, $v \in V_a \cup \{*\}$, and $s = (d, \omega)$, $\omega \in V_d$. In the sequel, we will call t and s the condition part and decision part of a rule, respectively. We will say that an object $u \in U$ supports a rule $t \to s$ iff u has both t and s properties in the given decision table.

Let $S = (U, C \cup D)$ be an incomplete decision table, $P \subseteq C$, $X_i \in MC_P$, $Y_j \in U/D$ and $X_i \cap Y_j \neq \emptyset$. By $des(X_i)$ and $des(Y_j)$, we denote the descriptions of the maximal consistent block X_i and the decision class Y_j in the decision table S. A decision rule is formally defined as

$$Z_{ij} : des(X_i) \to des(Y_j).$$

This is illustrated in the following example:

Example 3. As that in Example 2 for Table 1, let $X_1 = \{u_1\}$, $X_2 = \{u_2, u_6\}$, $X_3 = \{u_3\}$, $X_4 = \{u_4, u_5\}$, $X_5 = \{u_5, u_6\}$, $X_i \in MC_C$, and $Y_1 = \{u_1, u_2, u_4, u_6\}$, $Y_2 = \{u_3\}$, $Y_3 = \{u_5\}$. Then, the following decision rules can be induced from Table 1:

Z_{11}: $(P, \text{high}) \wedge (M, \text{low}) \wedge (S, \text{full}) \wedge (X, \text{low}) \to (d, \text{good})$,
Z_{21}: $((P, \text{low}) \wedge (S, \text{full})) \wedge ((M, \text{high}) \vee (X, \text{low})) \to (d, \text{good})$,
Z_{32}: $(S, \text{compact}) \wedge (X, \text{low}) \to (d, \text{poor})$,
Z_{41}: $((S, \text{full}) \wedge (X, \text{high})) \wedge ((P, \text{high}) \vee (P, *)) \to (d, \text{good})$,
Z_{43}: $((S, \text{full}) \wedge (X, \text{high})) \wedge ((P, \text{high}) \vee (P, *)) \to (d, \text{excellent})$,
Z_{51}: $(S, \text{full}) \wedge ((X, \text{high}) \vee ((P, \text{low}) \wedge (M, \text{high}))) \to (d, \text{good})$,
Z_{53}: $(S, \text{full}) \wedge ((X, \text{high}) \vee ((P, \text{low}) \wedge (M, \text{high}))) \to (d, \text{excellent})$.

In the condition parts of the rules Z_{41} and Z_{43}, the symbol "*" is used to factually characterize the description of the maximal consistent block. As we know, the symbol "*" is a missing value and can be filled with any value in its value field. Therefore, one cannot delete the description of the attribute in a decision rule as the value of an attribute is missing. In fact, the set $\{u_4, u_5\}$ is a maximal consistent block induced by the condition attributes and its description is $(S, \text{full}) \wedge (X, \text{high}) \wedge ((P, \text{high}) \vee (P, *))$. In this paper, we will not investigate how to generate all optimal decision rules or a class of optimal decision rules. In fact, these decision rules are equivalent to those in Example 4.2 of [18]. We will express the condition parts of these decision rules by using maximal consistent blocks in MC_C and the decision parts of them by using decision classes in U/d, respectively.

Like decision rules in complete information systems [36], the certainty measure, support measure and coverage measure of a decision rule Z_{ij} in an incomplete decision table can also be defined as

$$\mu(Z_{ij}) = |X_i \cap Y_j|/|X_i|, \quad s(Z_{ij}) = |X_i \cap Y_j|/|U| \quad \text{and} \quad \tau(Z_{ij}) = |X_i \cap Y_j|/|Y_j|,$$

respectively, where $|\cdot|$ is the cardinality of a set. It is clear that the value of each of $\mu(Z_{ij})$, $s(Z_{ij})$ and $\tau(Z_{ij})$ of a decision rule Z_{ij} falls into the interval $\left[\frac{1}{|U|}, 1\right]$. If the value of certainty measure of a decision rule is equal to one, then it is called certain; otherwise it is called uncertain. These three measures are illustrated in the following example.

Example 4. Continue from Example 3. By computing, we have that

$$\begin{aligned}
&Z_{11}\colon \mu(Z_{11}) = 1,\ \ s(Z_{11}) = \tfrac{1}{6},\ \ \tau(Z_{11}) = \tfrac{1}{4},\\
&Z_{21}\colon \mu(Z_{21}) = 1,\ \ s(Z_{21}) = \tfrac{1}{3},\ \ \tau(Z_{21}) = \tfrac{1}{2},\\
&Z_{32}\colon \mu(Z_{32}) = 1,\ \ s(Z_{32}) = \tfrac{1}{6},\ \ \tau(Z_{32}) = 1,\\
&Z_{41}\colon \mu(Z_{41}) = \tfrac{1}{2},\ \ s(Z_{41}) = \tfrac{1}{6},\ \ \tau(Z_{41}) = \tfrac{1}{4},\\
&Z_{43}\colon \mu(Z_{43}) = \tfrac{1}{2},\ \ s(Z_{43}) = \tfrac{1}{6},\ \ \tau(Z_{43}) = 1,\\
&Z_{51}\colon \mu(Z_{51}) = \tfrac{1}{2},\ \ s(Z_{51}) = \tfrac{1}{6},\ \ \tau(Z_{51}) = \tfrac{1}{4},\\
&Z_{53}\colon \mu(Z_{53}) = \tfrac{1}{2},\ \ s(Z_{53}) = \tfrac{1}{6},\ \ \tau(Z_{53}) = 1.
\end{aligned}$$

Example 4 shows that the decision rules Z_{11}, Z_{21} and Z_{32} are all certain, while others are all uncertain.

It is deserved to point out that, unlike a complete decision table, due to the generation of tolerance relation induced by condition attributes with null value "*", the sum of support measures of all decision rules induced by an incomplete decision table is not equal to one in general. For instance, $\sum s(Z_{ij}) = 6 \times \frac{1}{6} + \frac{1}{3} = \frac{4}{3} > 1$ in the above example.

In rough set theory, one can extract some decision rules from a given incomplete decision table. However, in some practical issues, it may happen that there does not exist any certain decision rule with the certainty of one in the decision-rule set extracted from a given incomplete decision table. In this situation, the lower approximation of the target decision is equal to an empty set in this incomplete decision table. To characterize this type of incomplete decision tables, in the following, incomplete decision tables are classified into three types according to their consistencies, which are consistent incomplete decision tables, conversely consistent incomplete decision tables and mixed incomplete decision tables.

As follows, we introduce several new concepts and notations, which will be applied in our further developments. We will denote by $|Z_{ij}|$ the cardinality of the set $X_i \cap Y_j$, which is called the support number of the rule Z_{ij}, and by $a(u)$ and $d(u)$ the values of an object u under a condition attribute $a \in C$ and a decision attribute $d \in D$, respectively.

Definition 1. Let $S = (U, C \cup D)$ be an incomplete decision table, $MC_C = \{X_1, X_2, \ldots, X_m\}$ and $U/D = \{Y_1, Y_2, \ldots, Y_n\}$. A maximal consistent block $X_i \in MC_C$ is said to be consistent if $d(u) = d(v)$ $\forall u, v \in X_i$ and $\forall d \in D$; a decision class $Y_j \in U/D$ is said to be conversely consistent if there exists a maximal consistent block X_i such that $u, v \in X_i$ $\forall u, v \in Y_j$.

Definition 2. Let $S = (U, C \cup D)$ be an incomplete decision table, $MC_C = \{X_1, X_2, \ldots, X_m\}$ and $U/D = \{Y_1, Y_2, \ldots, Y_n\}$. S is said to be consistent if every maximal consistent block $X_i \in MC_C$ is consistent; S is said to be conversely consistent if every decision class $Y_j \in U/D$ is conversely consistent.

An incomplete decision table is called a mixed decision table if it is neither consistent nor conversely consistent.

From the above definitions, it follows immediately that:

- an incomplete decision table S is consistent $\iff MC_C \preceq' MC_D$ $(MC_D = U/D)$,
- an incomplete decision table S is conversely consistent $\iff MC_D \preceq' MC_C$.

Obviously, a conversely consistent decision table is inconsistent. In addition to the above concepts and notations, we say that $S = (U, C \cup D)$ is strictly consistent (strictly and conversely consistent) if $MC_C \prec' MC_D$ $(MC_D \prec' MC_C)$, where $MC_D = U/D$. For convenience, we denote U/D by MC_D in the next part.

Remark. It is deserved to point out that these definitions are natural generalizations of Definitions 2 and 3 for a complete decision table in [44]. That is to say, if S is a complete decision table, then the maximal consistent blocks induced by the condition attribute set C will degenerate into the partition induced by C and the partial relation $\preceq'$ will degenerate into the partial relation on all partitions induced by the power set 2^C.

Granularity, a very important concept in rough set theory, is often used to indicate a partition or a cover of the universe of an information system or a decision table [22,24,25]. The decision performance of a decision rule depends directly on the condition granularity and decision granularity of a decision table [44]. In general, the change of granulation of a decision table can be realized through two ways [44]: (1) refining/coarsening the domain of attributes and (2) adding/reducing attributes. In general, information granulation is employed to measure the discernibility ability of a knowledge in information systems. The smaller information granulation of a knowledge is, the stronger its discernibility ability is [26,27]. In [28], Liang introduced an information granulation $G(A)$ to measure the discernibility ability of a knowledge in incomplete information systems, which is given in the following definition.

Definition 3. [28] Let $S=(U,A)$ be an incomplete information system and $U/SIM(A)=\{S_A(u_1),S_A(u_2),\ldots,S_A(u_{|U|})\}$. Information granulation of A is defined as

$$G(A)=\frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|S_A(u_i)|}{|U|}. \tag{1}$$

Following this definition, for a given decision table $S=(U,C\cup D)$, we call $G(C)$, $G(D)$ and $G(C\cup D)$ condition granulation, decision granulation and granulation of S, respectively.

As a result of the above discussions, we come to the following two lemmas.

Lemma 1. *Let $S=(U,C\cup D)$ be a strictly consistent decision table, i.e., $MC_C\prec' MC_D$. Then, there exists at least one decision class in MC_D such that it can be represented as the union of more than one maximal consistent blocks in MC_C.*

Proof. Let $MC_C=\{X_1,X_2,\ldots,X_m\}$ and $MC_D=\{Y_1,Y_2,\ldots,Y_n\}$. By the consistency of S, for any decision class $Y\in MC_D$, it is the union of some maximal consistent blocks $X\in MC_C$. Furthermore, since S is strictly consistent, there exist $X_0\in MC_C$ and $Y_0\in MC_D$ such that $X_0\subset Y_0$. It indicates that Y_0 is equal to the union of more than one maximal consistent blocks in MC_C. This completes the proof. □

Lemma 2. *Partial relation $\preceq'$ is a special instance of partial relation $\preceq$.*

Proof. Let $S=(U,A)$ be an incomplete information system, $P,Q\subseteq A$ with $P\preceq' Q$, $MC_P=\{P_1,P_2,\ldots,P_m\}$ and $MC_Q=\{Q_1,Q_2,\ldots,Q_n\}$. It follows from the definition of $\preceq\prime$ that for any $P_i\in MC_P$, there exists $Q_j\in MC_Q$ such that $P_i\subseteq Q_j$.Next, we prove that $S_P(u)\subseteq S_Q(u)$, $\forall u\in U$. Assume that $MC_P(u)=\{X_1,X_2,\ldots,X_m\}$ and $MC_Q(u)=\{Y_1,Y_2,\ldots,Y_n\}$. We know from Property 4 in [23] that $S_P(u)=\bigcup\{X_k\in MC_P\mid X_k\subseteq S_P(u)\}=\bigcup\{X_k\in MC_P(u)\}\quad(k\leqslant m)$ and $S_Q(u)=\bigcup\{Y_t\in MC_Q\mid Y_t\subseteq S_Q(u)\}=\bigcup\{Y_t\in MC_Q(u)\}\quad(t\leqslant n)$. From the definition of maximal consistent block, we have that $u\in MC_P(u)$, $u\in MC_Q(u)$, $u\notin MC_P-MC_P(u)$ and $u\notin MC_Q-MC_Q(u)$. Hence, it follows from $P\preceq' Q$ that for any $X_k\in MC_P(u)$, there exists $Y_t\in MC_Q(u)$ such that $X_k\subseteq Y_t$. Thus, for any $u\in U$, we can get that

$$\begin{aligned}S_P(u)&=\bigcup\{X_k\in MC_P|X_k\subseteq S_P(u)\}=\bigcup_{k=1}^{m}X_k\\&\subseteq\bigcup_{t=1}^{n}Y_t=\bigcup\{Y_t\in MC_Q|Y_t\subseteq S_Q(u)\}=S_Q(u),\end{aligned}$$

that is $P\preceq Q$. Therefore, partial relation $\preceq'$ is a special instance of partial relation $\preceq$. This completes the proof. □

By Lemma 2, one can easily obtain the following theorem.

Theorem 1. *Let $S=(U,C\cup D)$ be an incomplete decision table.*

(1) *If S is consistent, then $G(C)\leqslant G(D)$.*
(2) *If S is conversely consistent, then $G(C)\geqslant G(D)$.*

Proof. (1) If $S=(U,C\cup D)$ is consistent, we have that $MC_C\preceq' MC_D$. Hence, from Lemma 2, it follows that for any $u\in U$ one can obtain that $S_C(u)\subseteq S_D(u)$, i.e., $|S_C(u)|\subseteq|S_D(u)|$. Therefore,

$$G(C)=\frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|S_C(u_i)|}{|U|}\leqslant\frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|S_D(u_i)|}{|U|}=G(D),$$

that is $G(C)\leqslant G(D)$.

Analogously, (2) can be proved. □

It should be noted that the inverse propositions of Lemma 1 and Theorem 1 need not be true.

4. Limitations of traditional measures for incomplete decision tables

In this section, we reveal the limitations of some measures for evaluating the decision performance of an incomplete decision table.

In rough set theory, several measures for a decision rule $Z_{ij}:des(X_i)\rightarrow des(Y_j)$ have been introduced in [36], such as certainty measure $\mu(X_i,Y_j)=|X_i\cap Y_j|/|X_i|$, support measure $s(X_i,Y_j)=|X_i\cap Y_j|/|U|$ and coverage measure $\tau(X_i,Y_j)=|X_i\cap Y_j|/|Y_j|$. Naturally, their extensions in Section 2 of this paper are also suitable for evaluating the decision performance of a decision-rule extracted from an incomplete decision table. However, because $\mu(X_i,Y_j)$, $s(X_i,Y_j)$ and $\tau(X_i,Y_j)$ are only defined for a single decision rule and are not suitable for evaluating the decision performance of a decision-rule set extracted from an incomplete decision table.

In [40], approximation accuracy of a classification is introduced by Pawlak. Let $F=\{Y_1,Y_2,\ldots,Y_n\}$ be a classification or decision of the universe U (it can be regarded as a partition induced by decision attribute set D in a decision table, i.e., $F=U/D$) and C a condition attribute set. $\underline{C}F=\{\underline{C}Y_1,\underline{C}Y_2,\ldots,\underline{C}Y_n\}$ and $\overline{C}F=\{\overline{C}Y_1,\overline{C}Y_2,\ldots,\overline{C}Y_n\}$ are called C-lower and C-upper approximations of F, respectively, where $\underline{C}Y_i=\bigcup\{x\in U\mid [x]_C\subseteq Y_i\in F\}$ $(1\leqslant i\leqslant n)$ and $\overline{C}Y_i=\bigcup\{x\in U\mid [x]_C\cap Y_i\neq\varnothing, Y_i\in F\}$ $(1\leqslant i\leqslant n)$. The approximation accuracy of F by C is defined as

$$a_C(F)=\frac{\sum_{Y_i\in U/D}|\underline{C}Y_i|}{\sum_{Y_i\in U/D}|\overline{C}Y_i|}. \tag{2}$$

The approximation accuracy expresses the percentage of possible correct decisions when classifying objects by employing the attribute set C.

Definition 4. [23] Let $S=(U,A)$ be an incomplete information system and $P\subseteq A$. The approximation operators $\underline{apr}_P$ and $\overline{apr}_P$ are defined as

$$\underline{apr}_P(X)=\bigcup\{Y\in MC_P|Y\subseteq X\},$$
$$\overline{apr}_P(X)=\bigcup\{Y\in MC_P|Y\cap X\neq\varnothing\}.$$

Let $F=U/D=\{Y_1,Y_2,\ldots,Y_n\}$ be a classification of the universe U, and C a condition attribute set. In the view of maximal consistent block technique, we call $\underline{apr}_CF=\{\underline{apr}_C(Y_1),\underline{apr}_C(Y_2),\ldots,\underline{apr}_C(Y_n)\}$ and $\overline{apr}_CF=\{\overline{apr}_C(Y_1),\overline{apr}_C(Y_2),\ldots,\overline{apr}_C(Y_n)\}$ C-lower and C-upper approximations of F, respectively, where

$$\underline{apr}_C(Y_i)=\bigcup\{u\in U|MC_C(u)\subseteq Y_i,Y_i\in F\},\quad 1\leqslant i\leqslant n,$$

and

$$\overline{apr}_C Y_i = \bigcup\{u \in U | MC_C(u) \cap Y_i \neq \emptyset, Y_i \in F\}, \quad 1 \leqslant i \leqslant n.$$

Similar to formula (2), the approximation accuracy of F by C can be defined as

$$a_C(F) = \frac{\sum_{Y_i \in U/D} |\underline{apr}_C(Y_i)|}{\sum_{Y_i \in U/D} |\overline{apr}_C(Y_i)|}. \tag{3}$$

In some situations, $a_C(F)$ can be used to measure the certainty of an incomplete decision table. However, its limitations are revealed by the following example.

Example 5 (*Continued from Example* 1). By computing, one can obtain that

$MC_{a_2} = \{\{u_1, u_2, u_3, u_4, u_5\}, \{u_2, u_3, u_4, u_5, u_6\}\}$,
$MC_{a_2 \cup a_4} = \{\{u_1, u_2, u_3\}, \{u_4, u_5, u_6\}\}$ and
$U/d = \{\{u_1, u_2, u_4, u_6\}, \{u_3\}, \{u_5\}\}$.

From formula (3), we have that

$$a_{a_2}(U/d) = \frac{\sum_{Y_i \in U/d} |\underline{apr}_{a_2}(Y_i)|}{\sum_{Y_i \in U/d} |\overline{apr}_{a_2}(Y_i)|} = \frac{0}{6+6+6} = 0$$

and

$$a_{a_2 \cup a_4}(U/d) = \frac{\sum_{Y_i \in U/d} |\underline{apr}_{a_2 \cup a_4}(Y_i)|}{\sum_{Y_i \in U/d} |\overline{apr}_{a_2 \cup a_4}(Y_i)|} = \frac{0}{6+3+3} = 0.$$

That is to say $a_{a_2}(U/d) = a_{a_2 \cup a_4}(U/d)$.

In fact, the maximal consistent blocks induced by $a_2 \cup a_4$ should be much finer than those induced by a_2, i.e., the decision rules extracted by using $a_2 \cup a_4$ will have much higher certainty than those decision-rules extracted by using a_2. However, this example implies that the approximation accuracy cannot well characterize the certainty of an incomplete decision table when the lower approximation of the decision classification is an empty set. Therefore, a more comprehensive and effective measure for evaluating the certainty of an incomplete decision table is desired.

The consistency degree of a complete decision table $S = (U, C \cup D)$, another important measure proposed in [36], is defined as

$$c_C(D) = \frac{\sum_{i=1}^{n} |\underline{C} Y_i|}{|U|}. \tag{4}$$

The consistency degree expresses the percentage of objects which can be correctly classified to decision classes of U/D by a condition attribute set C. In some situations, $c_C(D)$ can be employed to measure the consistency of a decision table.

For an incomplete decision table, we can extend the consistency degree for measuring the consistency of a decision-rule set. Similar to formula (4), the consistency degree of an incomplete decision table is defined as

$$c_C(D) = \frac{\sum_{i=1}^{n} |\underline{apr}_C(Y_i)|}{|U|}. \tag{5}$$

Similar to Example 5, the consistency of an incomplete decision table also cannot be well characterized by the extended consistency degree because it only considers the lower approximation of a target decision. Therefore, a more comprehensive and effective measure for evaluating the consistency of an incomplete decision table is also needed.

From the definitions of the approximation accuracy and consistency degree, one can easily obtain the following property.

Property 1. *If* $S = (U, C \cup D)$ is a strictly and conversely consistent incomplete decision table, then $\alpha_C(U/D) = 0$ and $c_C(D) = 0$.

Property 1 shows that the extensions of the approximation accuracy and consistency degree cannot well characterize the certainty and consistency of a strictly and conversely consistent incomplete decision table.

Remark. From the above analyses, it is easy to see that the shortcomings of these two extended measures are mainly caused by the condition maximal consistent blocks that cannot be included in the lower approximation of the target decision in a given incomplete decision table. As we know, in an inconsistent incomplete decision table, there must exist some condition maximal consistent blocks that cannot be included in the lower approximation of the target decision. In fact, for a strictly and conversely consistent incomplete decision table, the lower approximation of the target decision is an empty set. Hence, we can draw the conclusion that the extensions of the approximation accuracy and consistency degree cannot be employed to effectively evaluate the decision performance of an inconsistent incomplete decision table. To overcome this drawback of the two extended measures, the effect of the condition maximal consistent blocks that are not included in the lower approximation of the target decision should be taken into account in evaluating the decision performance of an inconsistent incomplete decision table.

5. Performance evaluation of a decision-rule set

To evaluate the decision performance of a decision-rule set extracted from a complete decision table, one must take into consideration three important factors, that is, the certainty, consistency and support of the decision-rule set [44]. For decision problems in incomplete decision tables, these three factors also play an important role. Moreover, the degree of the cover induced by the missing values in the condition part can affect the decision performance of a decision-rule set extracted from an incomplete decision table. Although the extended approximation accuracy and consistency degree in Section 4, in some sense, can be used to measure the certainty and consistency of a decision-rule set extracted from an incomplete decision table, they will be invalid when the lower approximation of the target decision of an incomplete decision table equals an empty set.

To adequately evaluate the decision performance of an incomplete decision table, in this section, we introduce four new measures (α, β, γ and ϑ) and analyze how each of these four measures depends on the condition granulation and decision granulation of each of consistent incomplete decision tables and conversely consistent incomplete decision tables. Three incomplete decision tables from the real world are employed to demonstrate the advantage of the four new measures for evaluating the decision performance of a decision-rule set extracted from a general incomplete decision table.

Definition 5. Let $S = (U, C \cup D)$ be an incomplete decision table and $RULE = \{Z_{ij} \mid Z_{ij} : des(X_i) \rightarrow des(Y_j), X_i \in MC_C, Y_j \in MC_D\}$. Certainty measure α of S is defined as

$$\alpha(S) = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{N_i}\sum_{j=1}^{N_i}\frac{|X_i \cap Y_j|}{|X_i|}, \tag{6}$$

where N_i is the number of decision classes induced by the maximal consistent block X_i in the incomplete decision table.

In essence, the measure denotes the average value of the certainty measures of the decision-rules induced by each of maximal consistent blocks in an incomplete decision table. Note that the certainty measure of any decision rule is not equal to zero.

Theorem 2 (Extremum). *Let* $S = (U, C \cup D)$ *be an incomplete decision table and* $RULE = \{Z_{ij} \mid Z_{ij} : des(X_i) \rightarrow des(Y_j), X_i \in MC_C, Y_j \in MC_D\}$.

(1) *For any rule* $Z_{ij} \in RULE$, *if* $\mu(Z_{ij}) = 1$, *then the measure* α *achieves its maximum value* 1.
(2) *If* $m = 1$ *and* $n = |U|$, *then the measure* α *achieves its minimum value* $\frac{1}{|U|}$.

Proof. The results are straightforward and the proof is omitted. □

Remark. In fact, a decision table $S = (U, C \cup D)$ is consistent if and only if every decision rule from S is certain, i.e., its certainty measure of each of these decision rules is equal to one. So, (1) of Theorem 2 shows that the measure α achieves its maximum value 1 when S is consistent. When we want to distinguish any two objects of U without any condition information, (2) of Theorem 2 shows that α achieves its minimum value $\frac{1}{|U|}$.

In the following example, we show how the measure α overcomes the limitation of the extended measure $a_C(U/D)$.

Example 6 (*Continued from Example 5*). Let $S_1 = (U, \{a_2\} \cup d)$ and $S_2 = (U, \{a_2, a_4\} \cup d)$. Computing the measure α, we have that

$$\begin{aligned}\alpha(S_1) &= \frac{1}{m}\sum_{i=1}^{m}\frac{1}{N_i}\sum_{j=1}^{N_i}\frac{|X_i \cap Y_j|}{|X_i|}\\ &= \frac{1}{2}\left(\frac{1}{3}\times\left(\frac{3}{5}+\frac{1}{5}+\frac{1}{5}\right)+\frac{1}{3}\times\left(\frac{3}{5}+\frac{1}{5}+\frac{1}{5}\right)\right)=\frac{1}{3},\\ \alpha(S_2) &= \frac{1}{m}\sum_{i=1}^{m}\frac{1}{N_i}\sum_{j=1}^{N_i}\frac{|X_i \cap Y_j|}{|X_i|}\\ &= \frac{1}{2}\left(\frac{1}{2}\times\left(\frac{2}{3}+\frac{1}{3}\right)+\frac{1}{2}\times\left(\frac{2}{3}+\frac{1}{3}\right)\right)=\frac{1}{2}.\end{aligned}$$

Therefore, $\alpha(S_1) = \frac{1}{3} < \frac{1}{2} = \alpha(S_2)$, i.e., $\alpha(S_2) > \alpha(S_1)$.

Example 6 indicates that unlike the extended approximation accuracy $a_C(U/D)$, the measure α can be used to measure the certainty of a decision-rule set when $a_C(U/D) = 0$, i.e., the lower approximation of each decision class in the decision partition is equal to an empty set.

Remark. From the formula (3), it follows that $a_C(U/D) = 0$ if $\bigcup_{Y_i \in U/D} \underline{apr}_C(Y_i) = \emptyset$. In fact, in a broader sense, $\underline{apr}_C(Y_i) = \emptyset$ does not imply that the certainty of a decision rule concerning Y_i is equal to zero. So the measure α is much better than the extension of the approximation accuracy for measuring the certainty of a decision-rule set when an incomplete decision table is strictly and conversely consistent.

Corollary 1. *Let $S = (U, C \cup D)$ be an incomplete decision table. If S is consistent, then $\alpha(S) = 1$.*

Proof. It is straightforward from Definition 5 and (1) of Theorem 2.

In the following, we discuss the monotonicity of the measure α in a conversely consistent decision table. □

Theorem 3. *Let $S_1 = (U, C_1 \cup D_1)$ and $S_2 = (U, C_2 \cup D_2)$ be two conversely consistent incomplete decision tables. If $MC_{C_1} = MC_{C_2}$ and $MC_{D_2} \prec' MC_{D_1}$, then $\alpha(S_1) > \alpha(S_2)$.*

Proof. From $MC_1 = MC_2$ and the converse consistencies of S_1 and S_2, it follows that there exist $Y_q \in MC_{D_1}$ and $X_{p^1}, X_{p^2}, \ldots, X_{p^t} \in MC_{C_1}$ $(t \geqslant 1)$ such that $Y_q \subseteq X_{p^l}$ $(l \leqslant t)$. Since $MC_{D_2} \prec' MC_{D_1}$, there exist $Y_q^1, Y_q^2, \ldots, Y_q^s \in MC_{D_2}$ $(s > 1)$ such that $Y_q = \bigcup_{k=1}^{s} Y_q^k$. In other words, the rule $Z_{p^l q}$ $(l \leqslant t)$ in S_1 can be decomposed into a family of rules $Z_{p^l q}^1, Z_{p^l q}^2, \ldots, Z_{p^l q}^s$ in S_2. It is clear that $|Z_{p^l} q| = \sum_{k=1}^{s} |Z_{p^l q}^k|$ $(l \leqslant t)$.

Since S_1 and S_2 are all conversely consistent, one can see that the maximal consistent blocks of Y_q is the same as those of Y_q^k $(k \leqslant s)$, i.e., $X_{p^1}, X_{p^2}, \ldots, X_{p^t} \in MC_{C_1}$ $(t \geqslant 1)$. So $X_{p^l} \cap Y_q = Y_q$ and $X_{p^l} \cap Y_q^k = Y_q^k$ $(l \leqslant t)$, i.e., $| X_{p^l} \cap Y_q | = | Y_q |$ and $| X_{p^l} \cap Y_q^k | = | Y_q^k |$ $(l \leqslant t)$. Therefore, one can get that

$$\begin{aligned}
\alpha(S_1) &= \frac{1}{m}\sum_{i=1}^{m}\frac{1}{N_i}\sum_{j=1}^{N_i}\frac{|X_i\cap Y_j|}{|X_i|} = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{N_i}\sum_{j=1}^{N_i}\frac{|Y_j|}{|X_i|} = \frac{1}{m}\left(\sum_{i=1,i\neq p}^{m}\frac{1}{N_i}\sum_{j=1}^{N_i}\frac{|Y_j|}{|X_i|}+\frac{1}{N_p}\sum_{j=1}^{N_p}\frac{|Y_j|}{|X_p|}\right)\\
&= \frac{1}{m}\left(\sum_{i=1,i\neq p}^{m}\frac{1}{N_i}\sum_{j=1}^{N_i}\frac{|Y_j|}{|X_i|}+\frac{1}{N_p}\sum_{j=1}^{N_p}\frac{|\bigcup_{k=1}^{s}Y_j^k|}{|X_p|}\right) = \frac{1}{m}\left(\sum_{i=1,i\neq p}^{m}\frac{1}{N_i}\sum_{j=1}^{N_i}\frac{|Y_j|}{|X_i|}+\frac{1}{N_p}\sum_{j=1}^{N_p}\sum_{k=1}^{s}\frac{|Y_j^k|}{|X_p|}\right)\\
&> \frac{1}{m}\left(\sum_{i=1,i\neq p}^{m}\frac{1}{N_i}\sum_{j=1}^{N_i}\frac{|Y_j|}{|X_i|}+\frac{1}{N_p+s-1}\sum_{j=1}^{N_p}\sum_{k=1}^{s}\frac{|X_p\cap Y_j^k|}{|X_p|}\right) = \alpha(S_2),
\end{aligned}$$

that is $\alpha(S_1) > \alpha(S_2)$. This completes the proof. □

Theorem 3 states that the certainty measure α of a conversely consistent incomplete decision table decreases with its decision classes becoming finer.

The following theorem shows the monotonicity of α with respect to the condition part of an incomplete decision table.

Theorem 4. *Let $S_1 = (U, C_1 \cup D_1)$ and $S_2 = (U, C_2 \cup D_2)$ be two conversely consistent incomplete decision tables. If $MC_{D_1} = MC_{D_2}$ and $MC_{C_2} \prec' MC_{C_1}$, then $\alpha(S_1) < \alpha(S_2)$.*

Proof. From $MC_{C_2} \prec' MC_{C_1}$, there exists $X_p \in MC_{C_1}$ and an integer $s > 1$ such that $X_p = \bigcup_{l=1}^{s} X_p^l$, where $X_p^l \in MC_{C_2}$. That is to say, $X_p^1, X_p^2, \ldots, X_p^s$ constitute a cover on the maximal consistent block X_p. Noticing that both S_1 and S_2 are conversely consistent, we have that $X_p \supset Y_q$ and $X_p^l \supset Y_q$ $(Y_q \in MC_{D_1})$, i.e., $|X_p \cap Y_q| = |Y_q|$ and $|X_p^l \cap Y_q| = |Y_q|$ $(Y_q \in MC_{D_1})$. Hence, one can get that

$$\begin{aligned}
\alpha(S_1) &= \frac{1}{m}\sum_{i=1}^{m}\frac{1}{N_i}\sum_{j=1}^{N_i}\frac{|X_i\cap Y_j|}{|X_i|} = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{N_i}\sum_{j=1}^{N_i}\frac{|Y_j|}{|X_i|} = \frac{1}{m}\left(\sum_{i=1,i\neq p}^{m}\frac{1}{N_i}\sum_{j=1}^{N_i}\frac{|Y_j|}{|X_i|}+\frac{1}{N_p}\sum_{j=1}^{N_p}\frac{|Y_j|}{|X_p|}\right)\\
&= \frac{1}{m}\left(\sum_{i=1,i\neq p}^{m}\frac{1}{N_i}\sum_{j=1}^{N_i}\frac{|Y_j|}{|X_i|}+\frac{1}{N_p}\sum_{j=1}^{N_p}\frac{|Y_j|}{|\bigcup_{k=1}^{s}X_p^k|}\right) < \frac{1}{m+s-1}\left(\sum_{i=1,i\neq p}^{m}\frac{1}{N_i}\sum_{j=1}^{N_i}\frac{|Y_j|}{|X_i|}+\sum_{k=1}^{s}\frac{1}{N_p^k}\sum_{j=1}^{N_p^k}\frac{|Y_j|}{|X_p^k|}\right)\\
&= \alpha(S_2),
\end{aligned}$$

i.e., $\alpha(S_1) < \alpha(S_2)$. This completes the proof. □

Theorem 4 states that the certainty measure α of a conversely consistent incomplete decision table increases with its maximal consistent blocks in the condition part becoming finer.

In the following, through experimental analyses, we illustrate the validity of the measure α for evaluating the decision performance of a decision-rule set extracted from a general incomplete decision table. In order to show the advantage of the measure α over the extended measure $a_C(U/D)$, we have downloaded three public data sets with practical applications from UCI Repository of machine learning databases [58], which are described in Table 2. All condition attributes and decision attributes in these three data sets are discrete.

Here, we compare the certainty measure α with the approximation accuracy $a_C(D)$ on these three practical data sets. The comparisons of values of two measures with the numbers of features in these three data sets are shown in Tables 3–5 and Figs. 1–3.

Table 2
Data sets description

Data sets	Samples	Condition features	Decision classes
Soybean-large	307	35	19
Mushroom	8124	22	2
Nursery	12960	8	5

Table 3
$a_C(D)$ and α with different numbers of features in the data set soybean-large

Measure	Features										
	1	3	6	9	12	15	18	20	25	30	35
$a_C(D)$	0.0000	0.0000	0.0000	0.0000	0.0251	0.2067	0.2091	0.2105	0.2121	0.2121	0.2121
α	0.0945	0.2605	0.4279	0.4850	0.5411	0.7637	0.7663	0.7677	0.7709	0.7715	0.7715

Table 4
$a_C(D)$ and α with different numbers of features in the data set mushroom

Measure	Features										
	1	3	5	7	9	11	13	15	17	19	22
$a_C(D)$	0.0000	0.0000	0.2399	0.4728	0.5386	0.9445	0.9931	0.9951	0.9961	0.9980	1.0000
α	0.5000	0.5000	0.7500	0.9412	0.9684	0.9909	0.9982	0.9986	0.9991	0.9996	1.0000

Table 5
$a_C(D)$ and α with different numbers of features in the data set nursery

Measure	Features							
	1	2	3	4	5	6	7	8
$a_C(D)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000
α	0.2611	0.3278	0.3319	0.3358	0.4096	0.4295	0.4530	1.0000

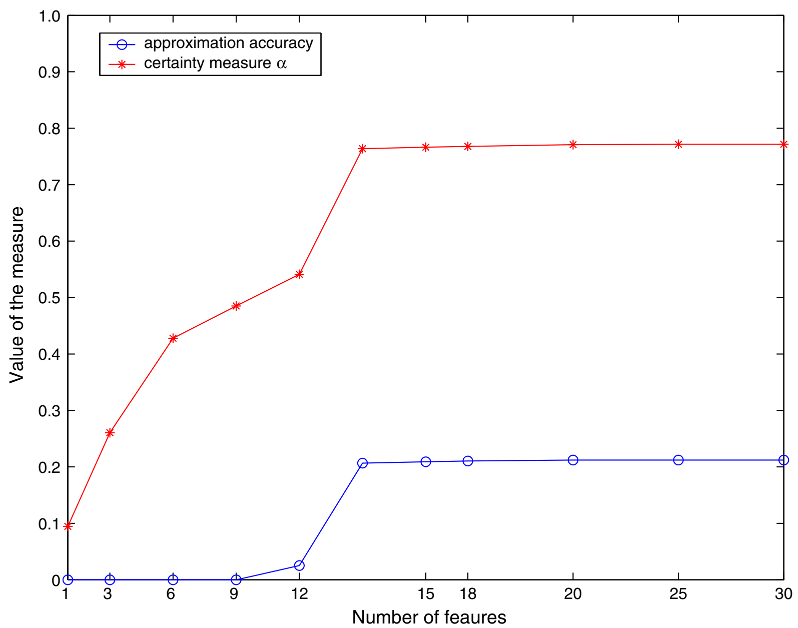

Fig. 1. Variation of the certainty measure α and the approximation accuracy with the number of features (data set soybean-large).

It can be seen from Tables 3–5 that the value of the certainty measure α is not smaller than that of the approximation accuracy $a_C(D)$ for the same number of selected features, and this value increases as the number of selected features becomes bigger in the same data set. The measure α and the extended approximation accuracy will achieve the same value 1 if the incomplete decision table becomes consistent after adding a

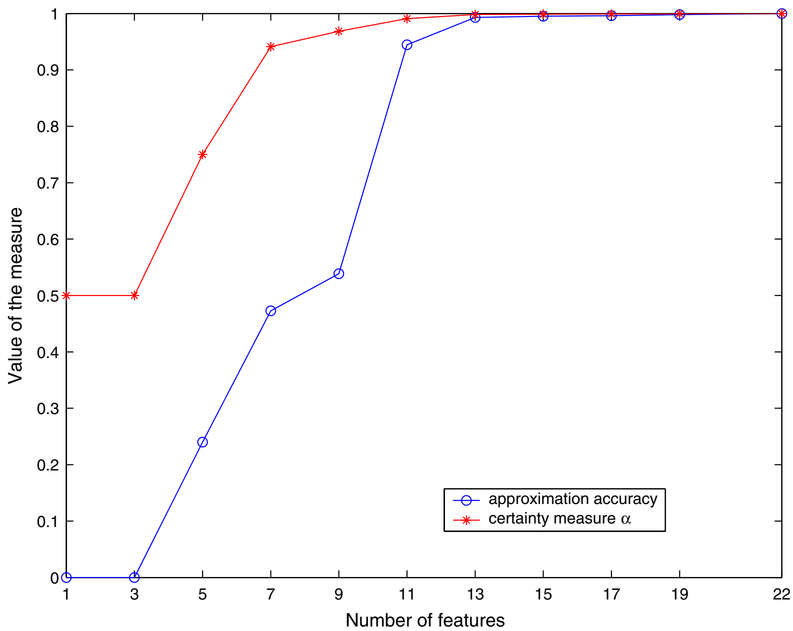

Fig. 2. Variation of the certainty measure α and the approximation accuracy with the number of features (data set mushroom).

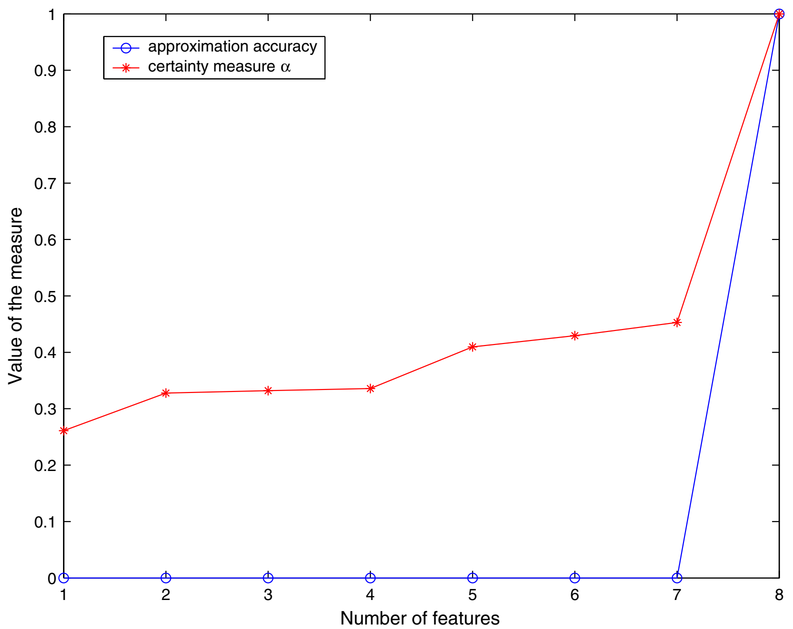

Fig. 3. Variation of the certainty measure α and the approximation accuracy with the number of features (data set nursery).

number of selected features. However, from Fig. 1, it is easy to see that the values of the extended approximation accuracy are equal to zero when the number of features falls in between 1 and 9. In this situation, the lower approximation of the target decision equals an empty set in the incomplete decision table. Hence, the extension of approximation accuracy cannot be used to effectively characterize the certainty of the incomplete decision table when the value of approximation accuracy equals zero. But, for the same situation, as the number of features varies from 1 to 9, the value of the certainty measure α changes from 0.0945 to 0.4850. It shows that unlike the extended approximation accuracy, the certainty measure α of the incomplete decision table with more features is higher than that of the incomplete decision table with fewer features. Thus, the measure α is much better than the extended approximation accuracy for this inconsistent incomplete decision table. One can draw the same conclusion from Figs. 2 and 3. In other words, when $a_C(D) = 0$ in Figs. 1–3, the measure α is still valid for evaluating the certainty of the set of decision rules obtained by using these selected features. Therefore, the measure α may be better than the extended approximation accuracy for evaluating the certainty of an incomplete decision table.

Based on the above analyses, we can conclude that if S is consistent, the evaluation ability of the measure α is the same as that of the accuracy measure $a_C(D)$ and that if S is inconsistent, the evaluation ability of the measure α is much higher than that of the extended accuracy measure $a_C(D)$.

Now we investigate how to measure the consistency of a decision-rule set extracted from an incomplete decision table.

At first, we discuss the consistency of the decision-rules induced by a maximal consistent block X in the condition part of a given incomplete decision table.

Let $S = (U, C \cup D)$ be an incomplete decision table, $X \in MC_C$ a maximal consistent block and $MC_D = U/D = \{[u]_D : u \in U\}$. For an object $u \in U$, a membership function of u in X is denoted as

$$\delta_X(u) = \frac{|X \cap [u]_D|}{|X|},$$

where $\delta_X(u)$ $(0 \leqslant \delta_X(u) \leqslant 1)$ represents a fuzzy concept. In fact, if $\delta_X(u) = 1$, then X can be said to be consistent with respect to $[u]_D$. In other words, if X is a consistent set with respect to $[u]_D$, then one has $X \subseteq [u]_D$. Given this function, one can generate a fuzzy set $F_X^D = \{(u, \delta_X(u)) \mid u \in U\}$ on the universe U.

Definition 6. Let $S = (U, C \cup D)$ be an incomplete decision table, $X \in MC_C$ a maximal consistent block and $MC_D = U/D = \{[u]_D : u \in U\}$. Inconsistency measure of X is defined as

$$E(F_X^D) = \sum_{i=1}^{|U|} \delta_X(u_i)(1 - \delta_X(u_i)), \tag{7}$$

where $\delta_X(u_i)$ is the membership function of $u_i \in U$ in X.

The class of all fuzzy (crisp, respectively) sets of U is denoted by $F(U)$ ($P(U)$, respectively). For $A \in F(U)$ and $u \in U$, $\delta_A(u)$ is the degree of u in A. If $A \in P(U)$, then $A(\cdot)$ expresses the characteristic function of A. Denote by $\underline{\underline{a}}$ $\forall a \in [0, 1]$, the constant fuzzy set with its membership function given by $\underline{\underline{a}}(u) = a$ $\forall u \in U$.

Definition 7 [25]. A real function $e : F(U) \to [0, 1]$ is referred to as an entropy on $F(U)$ if it satisfies the following conditions:

(1) $e(A) = 0$ iff $A \in P(U)$;
(2) $e(A) = \max_{A \in F(U)} e(A)$ iff $A = \underline{\underline{0.5}}$;
(3) for any $A, B \in F(U)$, if $\delta_B(u) \geqslant \delta_A(u)$ for $\delta_A(u) \geqslant \frac{1}{2}$ or if $\delta_B(u) \leqslant \delta_A(u)$ for $\delta_A(u) \leqslant \frac{1}{2}$, then $e(A) \geqslant e(B)$; and
(4) $e(A) = e(A^c)$ $\forall A \in F(U)$.

Theorem 5. *The inconsistency measure E is an entropy on $F(U)$.*

Proof. By Definition 7, we have that:

(1) If $X \in P(U)$, then, for all $u_i \in U$, either $\delta_X(u_i) = 0$ or $\delta_X(u_i) = 1$. Therefore, $E(X) = 0$. On the other hand, let $E(X) = 0$, then, for all $u_i \in U$, $\delta_X(u_i)(1 - \delta_X(u_i)) = 0$. It follows that either $\delta_X(u_i) = 0$ or $\delta_X(u_i) = 1$, i.e., X is a crisp set.

(2) Since $0 \leqslant \delta_X(u) \leqslant 1$, we have that $\max_{X \in F(U)}(\delta_X(u)(1-\delta_X(u))) = (\delta_{X_0}(u)(1-\delta_{X_0}(u))) = \frac{1}{4}$, where $X_0 \in F(U)$, and $\delta_X(u) = \frac{1}{2}$ for any $u \in U$. Hence, $E(\underline{\underline{0.5}}) = \max_{X \in F(U)} E(X)$.
(3) Let $X, Y \in F(U)$. If $\delta_X(u_i) \geqslant \frac{1}{2}$ and $\delta_Y(u_i) \geqslant \delta_X(u_i)$ for all $u_i \in U$, then

$$E(X) = \sum_{i=1}^{|U|} \delta_X(u_i)(1-\delta_X(u_i)) = \sum_{i=1}^{|U|}(-(\delta_X(u_i)-0.5)^2 + 0.25) = \frac{|U|}{4} - \sum_{i=1}^{|U|}(\delta_X(u_i)-0.5)^2$$
$$\geqslant \frac{|U|}{4} - \sum_{i=1}^{|U|}(\delta_Y(u_i)-0.5)^2 = E(Y).$$

If $\delta_X(u_i) \leqslant \frac{1}{2}$ and $\delta_Y(u_i) \leqslant \delta_X(u_i)$ for all $u_i \in U$, similar to the above proof, we have $E(X) \geqslant E(Y)$.
(4) $\forall X \in F(U)$, since $\delta_{\sim X}(u_i) = 1 - \delta_X(u_i)$, it follows that for all $u_i \in U$, $\delta_{\sim X}(u_i)(1-\delta_{\sim X}(u_i)) = (1-\delta_X(u_i))\delta_X(u_i)$. Therefore, $E(X) = E(\sim X)$.

Summarizing (1)–(4) above, we conclude that the inconsistency measure E is an entropy on $F(U)$. This completes the proof. □

Theorem 6. *The inconsistency measure of a consistent set is* 0.

Proof. Let $S = (U, C \cup D)$ be an incomplete decision table, $X \in MC_C$ a maximal consistent block and $MC_D = U/D = \{[u]_D : u \in U\}$. If X is a consistent set, then, for any $u \in X$, there exists a decision class $[u]_D$ such that $X \subseteq [u]_D$. So $\delta_X(u) = \frac{|X \cap [u]_D|}{|X|} = \frac{|X|}{|X|} = 1$. For any $u \in U - X$, we have $[u]_D \cap X = \varnothing$ and $\delta_X(u) = \frac{|X \cap [u]_D|}{|X|} = \frac{|\varnothing|}{|X|} = 0$. Therefore, $\delta_X(u_i)(1-\delta_X(u_i)) = 0\ \forall u_i \in U$, i.e., $E(F_X^D) = 0$. Thus, the inconsistency measure of a consistent set is 0. This completes the proof. □

Based on the above analyses, we propose a new measure β for measuring the consistency of a set of decision-rules extracted from an incomplete decision table, which is given in the following definition.

Definition 8. Let $S = (U, C \cup D)$ be an incomplete decision table and $RULE = \{Z_{ij} \mid Z_{ij} : des(X_i) \rightarrow des(Y_j), X_i \in MC_C, Y_j \in MC_D\}$. Consistency measure β of S is defined as

$$\beta(S) = \frac{1}{m}\sum_{i=1}^{m}\left[1 - \frac{4}{|X_i|}\sum_{j=1}^{N_i}|X_i \cap Y_j|\mu(Z_{ij})(1-\mu(Z_{ij}))\right], \tag{8}$$

where N_i is the number of decision-rules determined by the maximal consistent block X_i and $\mu(Z_{ij})$ is the certainty measure of the rule Z_{ij}.

To evaluate the consistency of an incomplete decision table, β computes the average value of the consistency measures for all the maximal consistent blocks in the condition part of the incomplete decision table.

Theorem 7 (Extremum). *Let* $S = (U, C \cup D)$ *be an incomplete decision table and* $\mathrm{RULE} = \{Z_{ij} \mid Z_{ij} : des(X_i) \rightarrow des(Y_j), X_i \in MC_C, Y_j \in MC_D\}$.

(1) *For every* $Z_{ij} \in \mathrm{RULE}$, *if* $\mu(Z_{ij}) = 1$, *then the measure* β *achieves its maximum value* 1, *and*
(2) *for every* $Z_{ij} \in \mathrm{RULE}$, *if* $\mu(Z_{ij}) = \frac{1}{2}$, *then the measure* β *achieves its minimum value* 0.

Proof. The results are straightforward and the proof is omitted. □

In the following example, we show how the measure β overcomes the limitation of the extended measure $c_C(D)$.

Example 7 (*Continued from Example* 5). Let $S_1 = (U, \{a_2\} \cup d)$ and $S_2 = (U, \{a_2, a_4\} \cup d)$. Computing the measure β, we have that

$$\beta(S_1) = \frac{1}{m}\sum_{i=1}^{m}\left[1 - \frac{4}{|X_i|}\sum_{j=1}^{N_i}|X_i \cap Y_j|\mu(Z_{ij})(1-\mu(Z_{ij}))\right]$$
$$= \frac{1}{2} \times \left[\left(1 - \frac{104}{125}\right) + \left(1 - \frac{80}{125}\right)\right] = \frac{33}{125},$$
$$\beta(S_2) = \frac{1}{m}\sum_{i=1}^{m}\left[1 - \frac{4}{|X_i|}\sum_{j=1}^{N_i}|X_i \cap Y_j|\mu(Z_{ij})(1-\mu(Z_{ij}))\right]$$
$$= \frac{1}{2} \times \left[\left(1 - \frac{8}{9}\right) + \left(1 - \frac{8}{9}\right)\right] = \frac{1}{9}.$$

Therefore, $\beta(S_1) = \frac{33}{125} > \frac{1}{9} = \beta(S_2)$.

Remark. Unlike the consistency degree $c_C(D)$, the measure β can be used to measure the consistency of a decision-rule set when $c_C(D) = 0$, i.e., the lower approximation of each of the decision classes in the decision part is equal to an empty set. From formula (5), it follows that $c_C(D) = 0$ if $\bigcup_{Y_i \in MC_D} \underline{apr}_C(Y_i) = \varnothing$. In fact, in a broader sense, $\underline{apr}_C(Y_i) = \varnothing$ does not imply that the certainty of a rule concerning Y_i is equal to zero. So, the measure β is much better than the extended measure $c_C(D)$ for evaluating the consistency of a decision-rule set when an incomplete decision table is strictly and conversely consistent.

Corollary 2. *Let $S = (U, C \cup D)$ be an incomplete decision table. If S is consistent, then $\beta(S) = 1$.*

Proof. It is straightforward from Definition 8 and (1) of Theorem 7.

The monotonicity of the measure β on conversely consistent incomplete decision tables can be found in the following Theorems 8 and 9.

Theorem 8. *Let $S_1 = (U, C_1 \cup D_1)$ and $S_2 = (U, C_2 \cup D_2)$ be two conversely consistent incomplete decision tables. If $MC_{C_1} = MC_{C_2}$ and $MC_{D_2} \prec' MC_{D_1}$, then $\beta(S_1) < \beta(S_2)$ when $\forall \mu(Z_{ij}) \leqslant \frac{1}{2}$, and $\beta(S_1) > \beta(S_2)$ when $\forall \mu(Z_{ij}) \geqslant \frac{1}{2}$.*

Proof. Since $MC_{C_1} = MC_{C_2}$ and the converse consistencies of S_1 and S_2, hence, for any $Y \in MC_{D_1}$, there exists $X \in MC_{C_1}$ such that $Y \subseteq X$. From $MC_{D_2} \prec' MC_{D_1}$, there exist $Y_q^1, Y_q^2, \ldots, Y_q^s \in MC_{D_2}$ $(s > 1)$ such that $Y_q = \bigcup_{k=1}^{s} Y_q^k$, where $Y_q \in MC_{D_1}$ with $Y_q \subseteq X_p$, $X_p \in MC_{C_1}$. In other words, the rule Z_{pq} in S_1 can be decomposed into a family of rules $Z_{pq}^1, Z_{pq}^2, \ldots, Z_{pq}^s$ in S_2. It is clear that $|Z_{pq}| = \sum_{k=1}^{s}|Z_{pq}^k|$, where

$$|Z_{pq}| = \frac{|X_p \cap Y_q|}{|X_p|} = \frac{|Y_q|}{|X_p|}, \qquad |Z_{pq}^k| = \frac{|X_p \cap Y_q^k|}{|X_p|} = \frac{|Y_q^k|}{|X_p|}, \quad k \leqslant s.$$

Let $\delta_D(Z_{il}) = \frac{|X_i \cap [x_l]_D|}{|X_i|}$ $(x_l \in X_i)$, where $[x_l]_D$ is the decision class of x_l induced by D. Then, we know that if $x_l \in X_i \cap Y_j$, it holds that $\delta_D(Z_{il}) = \mu(Z_{ij})$. Thus, one can obtain that

$$\beta(S) = \frac{1}{m}\sum_{i=1}^{m}\left[1 - \frac{4}{|X_i|}\sum_{j=1}^{N_i}|X_i \cap Y_j|\mu(Z_{ij})(1-\mu(Z_{ij}))\right] = \frac{1}{m}\sum_{i=1}^{m}\left[1 - \frac{4}{|X_i|}\sum_{l=1}^{|X_i|}\delta_D(Z_{il})(1-\delta_D(Z_{il}))\right]$$
$$= \frac{1}{m}\sum_{i=1}^{m}\frac{4}{|X_i|}\sum_{l=1}^{|X_i|}\left(\delta_D(Z_{il}) - \frac{1}{2}\right)^2.$$

Therefore, when $\forall\mu(Z_{ij}) \leqslant \frac{1}{2}$, we have that

$$\begin{aligned}\beta(S_1) &= \frac{1}{m}\sum_{i=1}^{m}\left[1-\frac{4}{|X_i|}\sum_{j=1}^{N_i}|X_i\cap Y_j|\mu(Z_{ij})(1-\mu(Z_{ij}))\right] = \frac{1}{m}\sum_{i=1}^{m}\frac{4}{|X_i|}\sum_{l=1}^{|X_i|}\left(\delta_{D_1}(Z_{il})-\frac{1}{2}\right)^2\\ &= \frac{1}{m}\left(\sum_{i=1,i\neq p}^{m}\frac{4}{|X_i|}\sum_{l=1}^{|X_i|}\left(\delta_{D_1}(Z_{il})-\frac{1}{2}\right)^2+\frac{4}{|X_p|}\sum_{l=1}^{|X_p|}\left(\delta_{D_1}(Z_{pl})-\frac{1}{2}\right)^2\right)\\ &< \frac{1}{m}\left(\sum_{i=1,i\neq p}^{m}\frac{4}{|X_i|}\sum_{l=1}^{|X_i|}\left(\delta_{D_2}(Z_{il})-\frac{1}{2}\right)^2+\frac{4}{|X_p|}\sum_{l=1}^{|X_p|}\left(\delta_{D_2}(Z_{pl})-\frac{1}{2}\right)^2\right)\\ &= \frac{1}{m}\sum_{i=1}^{m}\frac{4}{|X_i|}\sum_{l=1}^{|X_i|}\left(\delta_{D_2}(Z_{il})-\frac{1}{2}\right)^2 = \beta(S_2).\end{aligned}$$

Similar to the above, one can show that $\beta(S_1) > \beta(S_2)$ when $\forall\mu(Z_{ij}) \geqslant \frac{1}{2}$. This completes the proof. □

Theorem 8 states that the consistency measure β of a conversely consistent incomplete decision table increases with its decision classes becoming finer when $\forall\mu(Z_{ij}) \leqslant \frac{1}{2}$, and decreases with its decision classes becoming finer when $\forall\mu(Z_{ij}) \geqslant \frac{1}{2}$.

Theorem 9. *Let $S_1 = (U, C_1 \cup D_1)$ and $S_2 = (U, C_2 \cup D_2)$ be two conversely consistent incomplete decision tables. If $MC_{D_1} = MC_{D_2}$ and $MC_{C_2} \prec' MC_{C_1}$, then $\beta(S_1) > \beta(S_2)$ when $\forall\mu(Z_{ij}) \leqslant \frac{1}{2}$, and $\beta(S_1) < \beta(S_2)$ when $\forall\mu(Z_{ij}) \geqslant \frac{1}{2}$.*

Proof. Let $\delta_C(Z_{il}) = \frac{|X_i\cap[x_l]_D|}{|X_i|}$ $(x_l \in X_i, X_i \in U/C)$, where $[x_l]_D$ is the decision class of x_l induced by D. Then, we know that if $x_l \in X_i \cap Y_j$, it holds that $\delta_C(Z_{il}) = \mu(Z_{ij})$.

From $MC_{C_2} \prec' MC_{C_1}$, there exist $X_p \in MC_{C_1}$ and an integer $s > 1$ such that $X_p = \bigcup_{k=1}^{s} X_p^k$, where $X_p^k \in MC_{C_2}$. Clearly, we have $X_p^k \subset X_p$ for every $X_p^k \in MC_{C_2}$. Hence, $|X_p^k| < |X_p|$. From the converse consistencies of S_1 and S_2, it follows that

$$\mu(Z_{pj}) = \frac{|X_p\cap Y_j|}{|X_p|} = \frac{|Y_j|}{|X_p|} < \frac{|Y_j|}{|X_p^k|} = \frac{|X_p^k\cap Y_j|}{|X_p^k|} = \mu(Z_{pj}^k), \quad k = \{1,2,\ldots,s\}.$$

That is $\delta_{C_1}(Z_{il}) < \delta_{C_2}(Z_{il})$.

Thus, when $\forall\mu(Z_{ij}) \leqslant \frac{1}{2}$, we get that

$$\begin{aligned}\beta(S_1) &= \frac{1}{m}\sum_{i=1}^{m}\left[1-\frac{4}{|X_i|}\sum_{j=1}^{N_i}|X_i\cap Y_j|\mu(Z_{ij})(1-\mu(Z_{ij}))\right] = \frac{1}{m}\sum_{i=1}^{m}\frac{4}{|X_i|}\sum_{l=1}^{|X_i|}\left(\delta_{C_1}(Z_{il})-\frac{1}{2}\right)^2\\ &= \frac{1}{m}\left(\sum_{i=1,i\neq p}^{m}\frac{4}{|X_i|}\sum_{l=1}^{|X_i|}\left(\delta_{C_1}(Z_{il})-\frac{1}{2}\right)^2+\frac{4}{|X_p|}\sum_{l=1}^{|X_p|}\left(\delta_{C_1}(Z_{pl})-\frac{1}{2}\right)^2\right)\\ &> \frac{1}{m+s-1}\left(\sum_{i=1,i\neq p}^{m}\frac{4}{|X_i|}\sum_{l=1}^{|X_i|}\left(\delta_{C_2}(Z_{il})-\frac{1}{2}\right)^2+\frac{4}{|X_p|}\sum_{l=1}^{|X_p|}\left(\delta_{C_2}(Z_{pl})-\frac{1}{2}\right)^2\right)\\ &= \frac{1}{m+s-1}\left(\sum_{i=1,i\neq p}^{m}\frac{4}{|X_i|}\sum_{l=1}^{|X_i|}\left(\delta_{C_2}(Z_{il})-\frac{1}{2}\right)^2+\frac{4}{|X_p|}\sum_{k=1}^{s}\sum_{l=1}^{|X_p^k|}\left(\delta_{C_2}(Z_{pl})-\frac{1}{2}\right)^2\right) = \beta(S_2),\end{aligned}$$

that is $\beta(S_1) < \beta(S_2)$.

Similarly, one can prove that $\beta(S_1) < \beta(S_2)$ when $\forall\mu(Z_{ij}) \geqslant \frac{1}{2}$. This completes the proof. □

Theorem 9 states that the consistency measure β of a conversely consistent incomplete decision table decreases with its maximal consistent blocks in the condition part becoming finer when $\forall\mu(Z_{ij}) \leqslant \frac{1}{2}$, and increases with its maximal consistent blocks in the condition part becoming finer when $\forall\mu(Z_{ij}) \geqslant \frac{1}{2}$.

For general incomplete decision tables, to illustrate the differences between the consistency measure β and the consistency degree $c_C(D)$, the three practical data sets in Table 2 will be used again. The comparisons of values of the two measures with the numbers of features in these three data sets are shown in Tables 6–8, and Figs. 4–6.

From Tables 6–8, it can be seen that the value of the consistency measure β is not smaller than that of the extended consistency degree $c_C(D)$ for the same number of selected features, and this value increases as the number of selected features becomes bigger in the same data set. In particular, if the incomplete decision table becomes consistent after adding a number of selected features, the measure β and the extended consistency degree will have the same value 1.

Whereas, from Fig. 4, it is easy to see that the values of the consistency degree equal 0 when the number of features falls in between 1 and 9. In this situation, the lower approximation of the target decision in the incomplete decision table equals an empty set. Hence, the extension of consistency degree cannot be used to effectively characterize the consistency of the incomplete decision table when the value of the consistency degree equals zero. But, for the same situation, as the number of features varies from 1 to 9, the value of the consistency measure β changes within the interval $[0.0181, 0.4737]$. It shows that unlike the extended consistency degree, the consistency measure β is still valid for evaluating the consistency of the incomplete decision table when the lower approximation of the target decision is an empty set. Therefore, the measure β is much better than the extended consistency degree for this inconsistent incomplete decision table. Obviously, one can draw the same conclusion from Figs. 7 and 8. In other words, the measure β is still valid for evaluating the consistency of a set of decision rules obtained by using these selected features when the value of the consistency degree $c_C(D)$ is equal to zero. Given this advantage, we may conclude that the measure β is much better than the extended consistency degree for evaluating the consistency of an incomplete decision table.

Based on the above discussion, we can draw conclusions that if S is consistent, the evaluation ability of the measure β is the same as that of the extended consistency degree $c_C(D)$ and that if S is inconsistent, the evaluation ability of the measure β is much higher than that of the extended consistency degree $c_C(D)$.

In the following, we define a new measure γ for measuring the support degree of an incomplete decision table.

Table 6
$c_C(D)$ and β with different numbers of features in the data set soybean-large

Measure	Features										
	1	3	6	9	12	15	18	20	25	30	35
$c_C(D)$	0.0000	0.0000	0.0000	0.0000	0.1238	0.6678	0.6743	0.6840	0.6840	0.6840	0.6840
β	0.4737	0.2414	0.0765	0.0181	0.1003	0.5414	0.5465	0.5431	0.5263	0.5275	0.5275

Table 7
$c_C(D)$ and β with different numbers of features in the data set mushroom

Measure	Features										
	1	3	5	7	9	11	13	15	17	19	22
$c_C(D)$	0.0000	0.0000	0.3870	0.6421	0.7001	0.9714	0.9966	0.9975	0.9980	0.9990	1.0000
β	0.2734	0.6587	0.7518	0.9449	0.9737	0.9857	0.9971	0.9974	0.9984	0.9991	1.0000

Table 8
$c_C(D)$ and β with different numbers of features in the data set nursery

Measure	Features							
	1	2	3	4	5	6	7	8
$c_C(D)$	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000
β	0.13777	0.11119	0.11122	0.11126	0.11120	0.11111	0.11111	1.00000

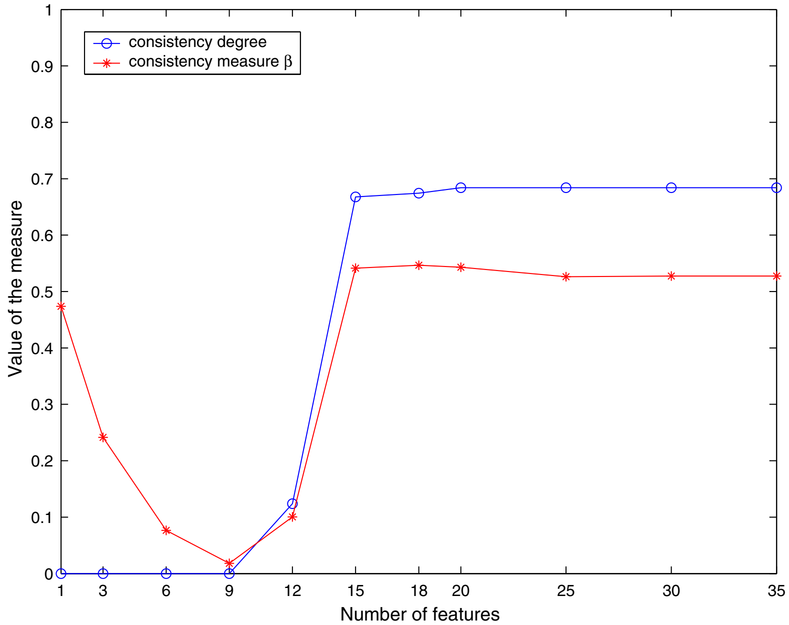

Fig. 4. Variation of the consistency measure β and the consistency degree with the number of features (data set soybean-large).

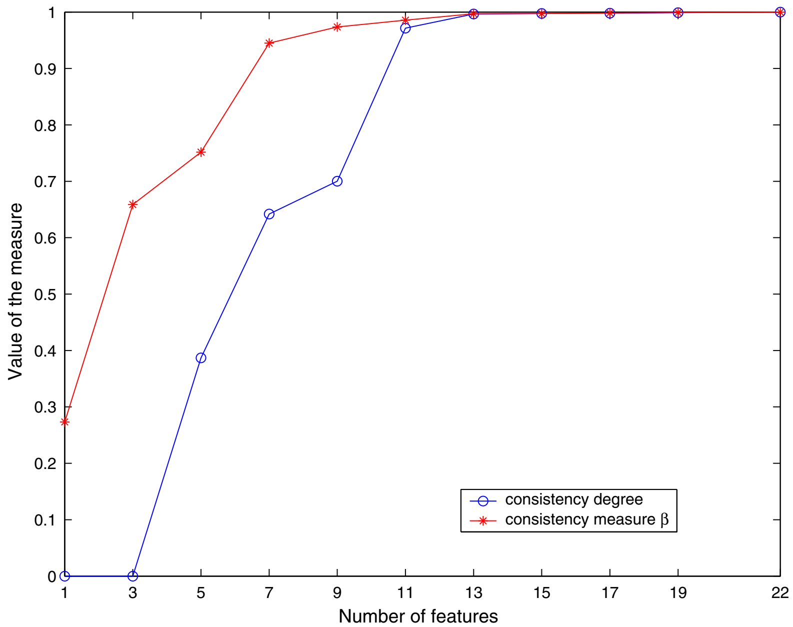

Fig. 5. Variation of the consistency measure β and the consistency degree with the number of features (data set mushroom).

Definition 9. Let $S = (U, C \cup D)$ be an incomplete decision table and $\mathrm{RULE} = \{Z_{ij} \mid Z_{ij} : des(X_i) \rightarrow des(Y_j), X_i \in MC_C, Y_j \in MC_D\}$. Support measure γ of S is defined as

$$\gamma(S) = \sum_{j=1}^{n} \frac{|Y_j|}{N_j|U|} \sum_{k=1}^{N_j} \frac{|X_k \cap Y_j|}{|U|}, \tag{9}$$

where N_j is the number of maximal consistent blocks in the condition part with respect to Y_j.

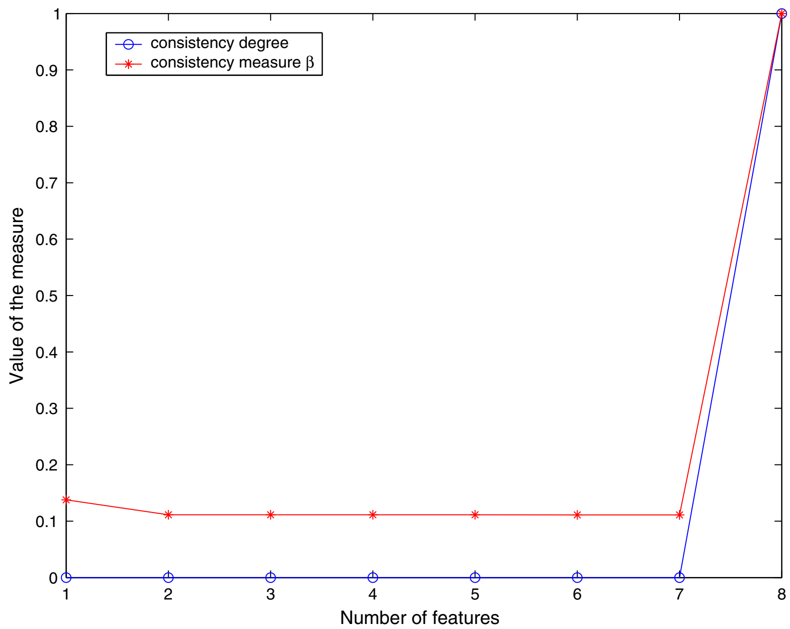

Fig. 6. Variation of the consistency measure β and the consistency degree with the number of features (data set nursery).

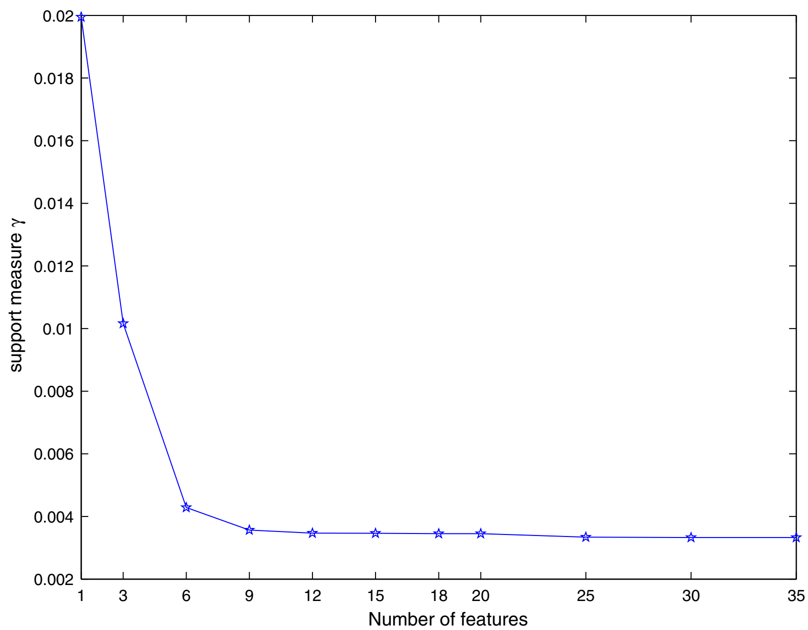

Fig. 7. Variation of the support measure γ with the number of features (data set soybean-large).

The measure γ is given by the weighted average value of the support measures of the decision rules with Y_j extracted from an incomplete decision table.

Theorem 10 (Extremum). *Let* $S = (U, C \cup D)$ *be an incomplete decision table and* $\mathrm{RULE} = \{Z_{ij} | Z_{ij} : des(X_i) \rightarrow des(Y_j), X_i \in MC_C, Y_j \in MC_D\}$.

(1) If $X_i = U$ and $Y_j = U$, then the measure γ achieves its maximum value 1, and
(2) if $| X_i \cap Y_j |= 1$ for any i, j, then the measure γ achieves its minimum value $\frac{1}{|U|}$.

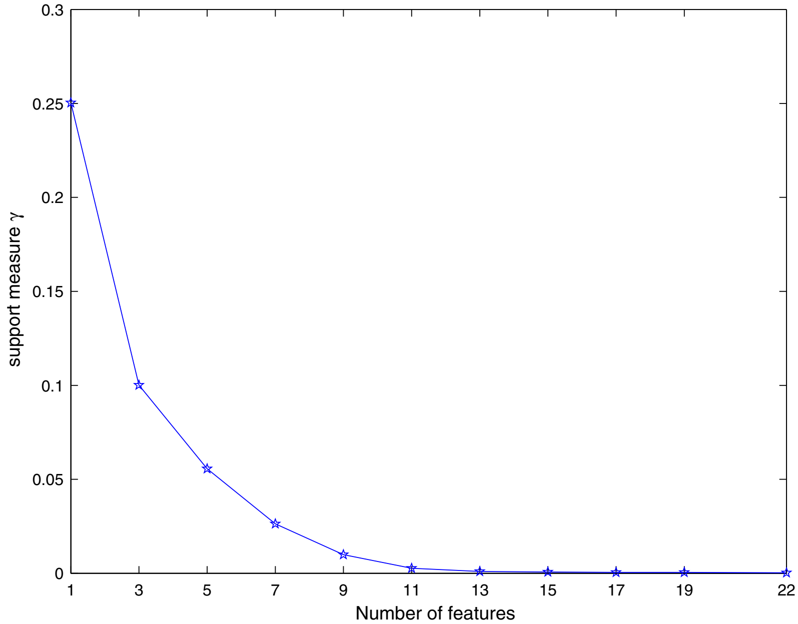

Fig. 8. Variation of the support measure γ with the number of features (data set mushroom).

Proof. The results are straightforward and the proof is omitted. □

Example 8 (*Continued from Example 5*). By computing the measure γ, it follows that

$$\begin{aligned}\gamma(S_1) &= \sum_{j=1}^{3}\frac{|Y_j|}{N_j|U|}\sum_{k=1}^{N_j}\frac{|X_k\cap Y_j|}{|U|}\\ &= \frac{4}{2\times 6}\left(\frac{3}{6}+\frac{3}{6}\right)+\frac{1}{2\times 6}\left(\frac{1}{6}+\frac{1}{6}\right)+\frac{1}{2\times 6}\left(\frac{1}{6}+\frac{1}{6}\right)=\frac{7}{18},\\ \gamma(S_2) &= \sum_{j=1}^{3}\frac{|Y_j|}{N_j|U|}\sum_{k=1}^{N_j}\frac{|X_k\cap Y_j|}{|U|}\\ &= \frac{4}{2\times 6}\left(\frac{2}{6}+\frac{2}{6}\right)+\frac{1}{1\times 6}\times\frac{1}{3}+\frac{1}{1\times 6}\times\frac{1}{3}=\frac{1}{3}.\end{aligned}$$

Hence, $\gamma(S_1)=\frac{7}{18}>\frac{6}{18}=\frac{1}{3}=\gamma(S_2)$, i.e., $\gamma(S_1)>\gamma(S_2)$.

Theorem 11. *Let $S_1=(U,C_1\cup D_1)$ and $S_2=(U,C_2\cup D_2)$ be two consistent incomplete decision tables. If $MC_{C_1}\prec' MC_{C_2}$ and $MC_{D_1}=MC_{D_2}$, then $\gamma(S_1)<\gamma(S_2)$.*

Proof. Since both S_1 and S_2 are all consistent, from $MC_{C_1}\prec' MC_{C_2}$, we have $X_p\subseteq Y_q$ and $X_p=\bigcup_{l=1}^{s}X_p^l$, where $X_p\in MC_{C_2}$, $Y_q\in MC_{D_2}$, and $X_p^l\in MC_{C_1}$. In other words, the rule Z_{pq} in S_2 can be decomposed into a family of rules $Z_{pq}^1,Z_{pq}^2,\ldots,Z_{pq}^s$ in S_1. Therefore,

$$\begin{aligned}\gamma(S_2) &= \sum_{j=1}^{n}\frac{|Y_j|}{N_j|U|}\sum_{k=1}^{N_j}\frac{|X_k\cap Y_j|}{|U|}=\sum_{j=1}^{n}\frac{|Y_j|}{|U|}\frac{1}{N_j}\sum_{k=1}^{N_j}\frac{|X_k|}{|U|}=\sum_{j=1,j\neq q}^{n}\frac{|Y_j|}{|U|}\frac{1}{N_j}\sum_{k=1}^{N_j}\frac{|X_k|}{|U|}+\frac{|Y_q|}{|U|}\frac{1}{N_q}\sum_{k=1}^{N_q}\frac{|X_k|}{|U|}\\ &= \sum_{j=1,j\neq q}^{n}\frac{|Y_j|}{|U|}\frac{1}{N_j}\sum_{k=1}^{N_j}\frac{|X_k|}{|U|}+\frac{|Y_q|}{|U|}\frac{1}{N_q}\left(\sum_{k=1,k\neq p}^{N_q}\frac{|X_k|}{|U|}+\frac{|X_p|}{|U|}\right)\end{aligned}$$

$$> \sum_{j=1,j\neq q}^{n} \frac{|Y_j|}{|U|}\frac{1}{N_j}\sum_{k=1}^{N_j}\frac{|X_k|}{|U|} + \frac{|Y_q|}{|U|}\frac{1}{N_q+s-1}\left(\sum_{k=1,k\neq p}^{N_q}\frac{|X_k|}{|U|} + \sum_{l=1}^{s}\frac{|X_p^l|}{|U|}\right)$$
$$= \sum_{j=1,j\neq q}^{n} \frac{|Y_j|}{|U|}\frac{1}{N_j}\sum_{k=1}^{N_j}\frac{|X_k|}{|U|} + \frac{|Y_q|}{|U|}\frac{1}{N_q+s-1}\sum_{k=1}^{N_q+s-1}\frac{|X_k|}{|U|} = \gamma(S_1),$$

that is $\gamma(S_1) < \gamma(S_2)$.

This completes the proof. □

Theorem 11 shows that the support measure γ of an incomplete decision table decreases with the maximal consistent blocks in its condition part becoming finer.

Theorem 12. *Let $S_1 = (U, C_1 \cup D_1)$ and $S_2 = (U, C_2 \cup D_2)$ be two conversely consistent incomplete decision tables. If $MC_{D_1} = MC_{D_2}$, then $\gamma(S_1) = \gamma(S_2)$.*

Proof. From the converse consistencies of S_1 and S_2, one has that $X_i(S_1) \cap Y_j(S_1) = Y_j(S_1)$ and $X_i(S_2) \cap Y_j(S_2) = Y_j(S_2)$, where $X_i(S_1) \in MC_{C_1}, Y_j(S_1) \in MC_{D_1}$, $X_i(S_2) \in MC_{C_2}$ and $Y_j(S_2) \in MC_{D_2}$. Since $MC_{D_1} = MC_{D_2}$, thus $Y_j(S_1) = Y_j(S_2)$ for $Y_j(S_1) \in MC_{D_1}$ and $Y_j(S_2) \in MC_{D_2}$. Hence, one can obtain that

$$\gamma(S_1) = \sum_{j=1}^{n}\frac{|Y_j(S_1)|}{N_j|U|}\sum_{k=1}^{N_j}\frac{|X_k(S_1)\cap Y_j(S_1)|}{|U|} = \sum_{j=1}^{n}\frac{|Y_j(S_1)|}{N_j|U|}\sum_{k=1}^{N_j}\frac{|Y_j(S_1)|}{|U|} = \sum_{j=1}^{n}\frac{|Y_j(S_2)|}{N_j|U|}\sum_{k=1}^{N_j}\frac{|Y_j(S_2)|}{|U|}$$
$$= \sum_{j=1}^{n}\frac{|Y_j(S_2)|}{N_j|U|}\sum_{k=1}^{N_j}\frac{|X_k(S_2)\cap Y_j(S_2)|}{|U|} = \gamma(S_2).$$

This completes the proof. □

Theorem 13. *Let $S_1 = (U, C_1 \cup D_1)$ and $S_2 = (U, C_2 \cup D_2)$ be two conversely consistent incomplete decision tables. If $MC_{C_1} = MC_{C_2}$ and $MC_{D_1} \prec' MC_{D_2}$, then $\gamma(S_1) < \gamma(S_2)$.*

Proof. Since S_1 and S_2 are all conversely consistent, from $MC_{D_1} \prec' MC_{D_2}$, we have that $Y_q \subseteq X_p$ and $Y_q = \bigcup_{l=1}^{s} Y_q^l$ $(s > 1)$, where $X_p \in MC_{C_2}$, $Y_q \in MC_{D_2}$ and $Y_q^l \in MC_{D_1}$. In other words, the rule Z_{pq} in S_2 can be decomposed into a family of rules $Z_{pq}^1, Z_{pq}^2, \ldots, Z_{pq}^s$ in S_1. Therefore,

$$\gamma(S_2) = \sum_{j=1}^{n}\frac{|Y_j|}{N_j|U|}\sum_{k=1}^{N_j}\frac{|X_k\cap Y_j|}{|U|} = \sum_{j=1}^{n}\frac{|Y_j|}{|U|}\frac{1}{N_j}\sum_{k=1}^{N_j}\frac{|Y_j|}{|U|} = \sum_{j=1,j\neq q}^{n}\frac{|Y_j|}{|U|}\frac{1}{N_j}\sum_{k=1}^{N_j}\frac{|X_k|}{|U|} + \frac{|Y_q|}{|U|}\frac{1}{N_q}\sum_{k=1}^{N_q}\frac{|Y_q|}{|U|}$$
$$= \sum_{j=1,j\neq q}^{n}\frac{|Y_j|}{|U|}\frac{1}{N_j}\sum_{k=1}^{N_j}\frac{|X_k|}{|U|} + \frac{|Y_q|}{|U|}\frac{1}{N_q}\sum_{k=1}^{N_q}\frac{|Y_q|}{|U|} = \sum_{j=1,j\neq q}^{n}\frac{|Y_j|}{|U|}\frac{1}{N_j}\sum_{k=1}^{N_j}\frac{|X_k|}{|U|} + \frac{|Y_q|^2}{|U|^2}$$
$$= \sum_{j=1,j\neq q}^{n}\frac{|Y_j|}{|U|}\frac{1}{N_j}\sum_{k=1}^{N_j}\frac{|X_k|}{|U|} + \frac{(|Y_q^1| + |Y_q^2| + \cdots + |Y_q^s|)^2}{|U|^2}$$
$$> \sum_{j=1,j\neq q}^{n}\frac{|Y_j|}{|U|}\frac{1}{N_j}\sum_{k=1}^{N_j}\frac{|X_k|}{|U|} + \frac{|Y_q^1|^2 + |Y_q^2|^2 + \cdots + |Y_q^s|^2}{|U|^2}$$
$$= \sum_{j=1,j\neq q}^{n}\frac{|Y_j|}{|U|}\frac{1}{N_j}\sum_{k=1}^{N_j}\frac{|X_k|}{|U|} + \sum_{l=1}^{s}\frac{|Y_q^l|}{|U|}\frac{1}{N_q^l}\sum_{k=1}^{N_q^l}\frac{|Y_q^l|}{|U|} = \sum_{j=1}^{n+s-1}\frac{|Y_j|}{|U|}\frac{1}{N_j}\sum_{k=1}^{N_j}\frac{|X_k|}{|U|} = \gamma(S_1),$$

that is $\gamma(S_1) < \gamma(S_2)$. This completes the proof. □

Theorem 13 states that the support measure γ of an incomplete decision table decreases with its decision classes becoming finer.

Table 9
γ with different numbers of features in the data set soybean-large

Measure	Features										
	1	3	6	9	12	15	18	20	25	30	35
γ	0.0199	0.0102	0.0043	0.0036	0.0035	0.0035	0.0035	0.0035	0.0033	0.0033	0.0033

Table 10
γ with different numbers of features in the data set mushroom

Measure	Features										
	1	3	4	7	9	11	13	15	17	19	22
γ	0.2503	0.1001	0.0556	0.0263	0.0099	0.0026	0.0009	0.0007	0.0004	0.0004	0.0002

Table 11
γ with different numbers of features in the data set nursery

Measure	Features							
	1	2	3	4	5	6	7	8
γ	0.1059	0.0245	0.0061	0.0015	0.0006	0.0003	0.0001	0.00007

Finally, we investigate the variation of the values of the support measure γ with the numbers of features in the three practical data sets in Table 2. The values of the measure with the numbers of features in these three data sets are shown in Tables 9–11 and Figs. 7–9.

From these tables and figures, one can see that the value of the support measure γ decreases with the number of condition features becoming bigger in the same data set. Note that one may extract more decision rules through adding the number of condition features in general. In fact, the bigger the number of decision rules is, the smaller the value of the support measure is in the same data set. Therefore, the measure γ is able to effectively evaluate the support of all decision-rules extracted from a given decision table.

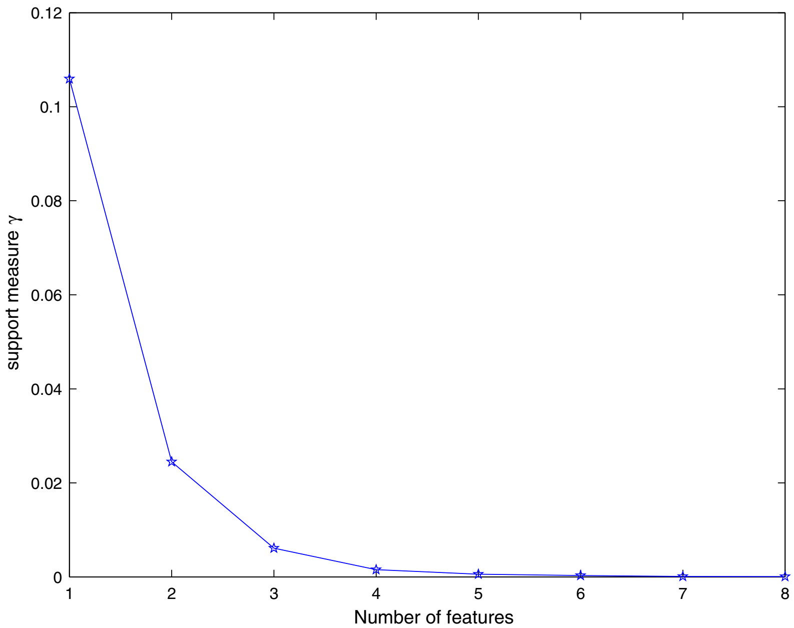

Fig. 9. Variation of the support measure γ with the number of features (data set nursery).

It is deserved to point out that the values of the three new measures (α, β and γ), in some sense, are dependent on the number of missing information in the condition part of an incomplete decision table, i.e., the scale of the cover induced by the maximal consistent blocks in the condition part. In the following, we introduce another measure ϑ to measure the scale of the cover in the condition part of an incomplete decision table.

Definition 10. Let $S=(U,C\cup D)$ be an incomplete decision table and $\mathrm{RULE}=\{Z_{ij} \mid Z_{ij}: des(X_i)\rightarrow des(Y_j), X_i\in MC_C, Y_j\in MC_D\}$. Cover measure ϑ of S is defined as

$$\vartheta(S)=\frac{1}{|U|}\sum_{i=1}^{m}\frac{|X_i|}{|U|}, \tag{10}$$

where $\frac{1}{|U|}\leqslant\vartheta(S)\leqslant 1$.

The ϑ is a measure for the scale of the cover of the universe determined by the maximal consistent blocks in the condition part of an incomplete decision table.

Example 9 (*Continued from Example* 2). By computing the measure ϑ, it follows that

$$MC_P=\{\{u_1\},\{u_2,u_6\},\{u_3\},\{u_4,u_5\},\{u_5,u_6\}\}$$

and

$$\vartheta(S)=\frac{1}{|U|}\sum\nolimits_{i=1}^{m}\frac{|X_i|}{|U|}=\frac{1}{6}\left(\frac{1}{6}+\frac{2}{6}+\frac{1}{6}+\frac{2}{6}+\frac{2}{6}\right)=\frac{2}{9}.$$

Theorem 14 (Minimum). *Let $S=(U,C\cup D)$ be a complete decision table, then the measure ϑ achieves its minimum value $\frac{1}{|U|}$.*

Proof. Let $MC_C=\{X_1,X_2,\ldots,X_m\}$. If $S=(U,C\cup D)$ be a complete decision table, we have the partition $U/C=MC_C=\{X_1,X_2,\ldots,X_m\}$. Thus, $\bigcup_{i=1}^{m}X_i=U$ and $X_i\cap X_j=\varnothing$ $(i\neq j)$, $i,j\leqslant m$, i.e., $\sum_{i=1}^{m}|X_i|=|U|$. Hence, we have that $\vartheta(S)=\frac{1}{|U|}\sum_{i=1}^{m}\frac{|X_i|}{|U|}=\frac{1}{|U|}\frac{|U|}{|U|}=\frac{1}{|U|}$. This completes the proof. □

Corollary 3. *The measure ϑ achieves its maximum value 1 if and only if $|U|=1$.*

In this case, $\max(\vartheta(S))=\min(\vartheta(S))=1$.

Corollary 4. *Let $S_1=(U,C_1\cup D_1)$ and $S_2=(U,C_2\cup D_2)$ be two incomplete decision tables. If $C_1\subseteq C_2$, then $\vartheta(S_1)\geqslant\vartheta(S_2)$.*

Proof. It is straightforward.

Since the measure ϑ is very simple, its experimental analysis is omitted in this paper. □

Remark. As we know, the maximal consistent blocks MC_C in the condition part can be degenerated into the equivalence classes U/C if $S=(U,C\cup D)$ is a complete decision table, and the maximal consistent block $X\in MC_C$ can be degenerated into the equivalence class. Hence, these four new measures (α, β, γ and ϑ) can also be used to measure the decision performance of a decision-rule set extracted from a complete decision table if $X\in MC_C$ is regarded as an equivalence class of U/C in the formulae 6, 8, 9 and 10. The evaluation measures proposed in this paper may be helpful for determining which of rule-extracting methods is preferred for a particular application about extracting decision rules from incomplete decision tables.

6. Conclusions

In rough set theory, several classical measures for evaluating a decision rule or a decision table, such as the certainty measure, support measure and coverage measure of a decision rule and the approximation accuracy and consistent degree of a decision table, can be extended for evaluating the decision performance of a decision rule (set) extracted from an incomplete decision table. However, these extensions are not effective for evaluating the decision performance of a decision-rule set. In this paper, the limitations of these extensions have been exemplified on incomplete decision tables. To overcome these limitations, incomplete decision tables have been classified into three types according to their consistencies and four new and more effective measures (α, β, γ and ϑ) have been introduced for evaluating the certainty, consistency, support and cover of a decision-rule set extracted from an incomplete decision table, respectively. It has been analyzed how each of these four new measures depends on the condition granulation and decision granulation of each of the three types of incomplete decision tables. The experimental analyses on three practical incomplete decision tables show that the three new measures (α, β, γ) are adequate for evaluating the decision performance of a decision-rule set extracted from an incomplete decision table in rough set theory. These four measures may be helpful for determining which of rule-extracting approaches is preferred for a practical decision problem in the context of incomplete decision tables. Another important fact we would like to point out is that the measures proposed in this paper are natural generalizations of the performance evaluation measures for complete decision tables.

Acknowledgements

The authors wish to thank the anonymous reviewers for their constructive comments on this study. This work was supported by SRG: 7002006 of City University of Hong Kong, the high technology research and development program (No. 2007AA01Z165), the national natural science foundation of China (Nos. 60773133, 70471003, 60573074), the foundation of doctoral program research of the ministry of education of China (No. 20050108004) and key project of science and technology research of the ministry of education of China.

References

[1] J. Bazan, J.F. Peters, A. Skowron, HS. Nguyen, M. Szczuka, Rough set approach to pattern extraction from classifiers, Electronic Notes in Theoretical Computer Science 82 (4) (2003) 1–10.
[2] T. Beaubouef, Rules in incomplete information systems, Information Sciences 113 (1999) 271–292.
[3] T. Beaubouef, F.E. Perty, G. Arora, Information-theoretic measures of uncertainty for rough sets and rough relational databases, Information Sciences 109 (1998) 185–195.
[4] M. Beynon, Reducts within the variable precision rough sets model: a further investigation, European Journal of Operational Research 134 (3) (2001) 592–605.
[5] M.R. Chmielewski, J.W. Grzymala-Busse, N.W. Peterson, S. Than, The rule induction system LERS, a version for personal computers, Foundations of Computing and Decision Sciences 18 (3–4) (1993) 181–212.
[6] P. Clark, T. Niblett, The CN2 induction algorithm, Machine Learning 3 (1989) 261–283.
[7] W. Cohen, Fast effective rule induction, in: A. Prieditis, S. Russell (Eds.), Proceedings of the 12th International Conference on Machine Learning, Morgan Kaufman, Tahoe City, CA, 1995, pp. 115–123.
[8] I. Düntsch, G. Gediga, Uncertainty measures of rough set prediction, Artificial Intelligence 106 (1998) 109–137.
[9] Y.Y. Gan, H.K. Wang, Set-valued information systems, Information Science 176 (17) (2006) 2507–2525.
[10] S. Greco, Z. Pawlak, R. Slowinski, Can Bayesian confirmation measures be useful for rough set decision rules?, Engineering Applications of Artificial Intelligence 17 (2004) 345–361
[11] J.W. Grzymala-Busse, Characteristic relations for incomplete data: a generalization of the indiscernibility relation, Transactions on Rough Sets 4 (2005) 58–68.
[12] J.W. Grzymala-Busse, W. Gryzymala-Busse, An experimental comparison of three rough set approaches to missing attribute value, Transactions on Rough Sets 6 (2007) 31–50.
[13] J. Huysmans, B. Baesens, J. Vanthienen, A new approach for measuring rule set consistency, Data and Knowledge Engineering 63 (1) (2007) 167–182.
[14] J. Komorowski, Z. Pawlak, L. Polkowski, A. Skowron, Rough sets: a tutuorial, in: S.K. Pal, A. Skowron (Eds.), Rough Fuzzy Hubridization: A New Trend in Decision Making, Springer, Singapore, 1999, pp. 3–98.
[15] I. Kononenko, I. Bratko, E. Roskar, Experiments in automatic learning of medical diagnostic rules, Technical Report, Jozef Stefan Institute, Ljubljana, Yugoslavia, 1984.
[16] M. Kryszkiewicz, Rough set approach to incomplete information systems, in: Proceedings of Second Annual Joint Conference on Information Sciences, Fuzzy Logic, Neural Computing, Pattern Recognition, Computer Vision, Evolutionary Computing, Information Theory, Computational Intelligence, Wrightsville Beach, North Carolina, USA, 1995, pp. 194–197.
[17] M. Kryszkiewicz, Rough set approach to incomplete information systems, Information Sciences 112 (1998) 39–49.
[18] M. Kryszkiewicz, Rule in incomplete information systems, Information Sciences 113 (1999) 271–292.
[19] M. Kryszkiewicz, Comparative study of alternative type of knowledge reduction in inconsistent systems, International Journal of Intelligent Systems 16 (2001) 105–120.
[20] R. Latkowski, On decomposition for incomplete data, Fundamenta Informaticae 54 (1) (2003) 1–16.

[21] R. Latkowski, M. Mikolajczyk, Data decomposition and decision rule joining for classification of data with missing values, Transactions on Rough Sets 3 (2004) 299–320.
[22] R. Latkowski, Flexible indiscernibility relations for missing attribute values, Fundamenta Informaticae 67 (1) (2005) 131–147.
[23] Y. Leung, D.Y. Li, Maximal consistent block technique for rule acquisition in incomplete information systems, Information Sciences 153 (2003) 85–106.
[24] D.Y. Li, B. Zhang, Y. Leung, On knowledge reduction in inconsistent decision information systems, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 12 (5) (2004) 651–672.
[25] J.Y. Liang, D.Y. Li, Uncertainty and Knowledge Acquisition in Information Systems, Science, Press, Beijing, China, 2005.
[26] J.Y. Liang, K.S. Qu, Information measures of roughness of knowledge and rough sets in incomplete information systems, Journal of System Science and System Engineering 24 (5) (2001) 544–547.
[27] J.Y. Liang, Z.Z. Shi, The information entropy, rough entropy and knowledge granulation in rough set theory, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 12 (1) (2004) 37–46.
[28] J.Y. Liang, Z.Z. Shi, D.Y. Li, M.J. Wierman, The information entropy, rough entropy and knowledge granulation in incomplete information system, International Journal of General Systems 35 (6) (2006) 641–654.
[29] J.Y. Liang, Z.B. Xu, The algorithm on knowledge reduction in incomplete information systems, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 24 (1) (2002) 95–103.
[30] E. Marczewski, A general scheme of independence in mathematics, Bull. Acad. Polon. Sci. Ser. Sci. Math. Astronom. Phys. 6 (1958) 331–362.
[31] J.S. Mi, W.Z. Wu, W.X. Zhang, Comparative studies of knowledge reductions in inconsistent systems, Fuzzy Systems and Mathematics 17 (3) (2003) 54–60.
[32] J.S. Mi, W.Z. Wu, W. X Zhang, Approaches to knowledge reduction based on variable precision rough set model, Information Sciences 159 (2004) 255–272.
[33] H.S. Nguyen, D. Slezak, Approximation reducts and association rules correspondence and complexity results, In: N. Zhong, A. Skowron, S. Oshuga (Eds.), Proceeding of RSFDGrc'99, Yamaguchi, Japan, LNAI 1711, 1999, pp 137–145.
[34] S.K. Pal, W. Pedrycz, A. Skowron, R. Swiniarski, Presenting the special issue on rough-neuro computing, Neurocomputing 36 (2001) 1–3.
[35] Z. Pawlak, Rough sets, International Journal of Computer and Information Science 11 (1982) 341–356.
[36] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Dordrecht, 1991.
[37] Z. Pawlak, Rough set theory and its applications in data analysis, Cybernetics and Systems 29 (1998) 661–688.
[38] Z. Pawlak, Some remarks on conflict analysis, Information Sciences 166 (2005) 649–654.
[39] Z. Pawlak, A. Skowron, Rudiments of rough sets, Information Sciences 177 (2007) 3–27.
[40] Z. Pawlak, A. Skowron, Rough sets: some extensions, Information Sciences 177 (2007) 28–40.
[41] Z. Pawlak, A. Skowron, Rough sets and boolean reasoning, Information Sciences 177 (2007) 41–73.
[42] Z. Pawlak, J.W. Grzymala-Busse, R. Slowiski, W. Ziarko, Rough sets, Commun. ACM 38 (11) (1995) 89–95.
[43] Y.H. Qian, J.Y. Liang, Combination entropy and combination granulation in incomplete information systems, Lecture Notes in Artificial Intelligence 4062 (2006) 184–190.
[44] Y.H. Qian, J.Y. Liang, D.Y. Li, H.Y. Zhang, C.Y. Dang, Measures for evaluating the decision performance of a decision table in rough set theory, Information Sciences 178 (1) (2008) 181–202.
[45] Y.H. Qian, J.Y. Liang, C.Y. Dang, Converse approximation and rule extracting from decision tables in rough set theory, Computer and Mathematics with Applications (2007), doi:10.1016/j.camwa.2007.08.031.
[46] M. Quafatou, α-RST: a generalization of rough set theory, Information Sciences 124 (2000) 301–316.
[47] J.R. Quinlan, Induction of decision trees, in: J.W. Shavlik, T.G. Dietterich (Eds.), Readings in Machine Learning, Morgan Kaufman, Los Altes, CA, 1990, pp. 57–69.
[48] J.R. Quinlan, C45 Programs for Machine Learning, Morgan Kaufman, San Francisco, CA, USA, 1993.
[49] A. Skowron, C. Rauszer, The discernibility matrices functions in information systems, in: R. Słowiński (Ed.), Intelligent Decision Support, Handbook of Applications and Advances of the Rough Sets Theory, Kluwer Academic, Dordrecht, 1992, pp. 331–362.
[50] D. Slezak, Approximate reducts in decision tables, in: Proceeding of IPMU'96, Granada, Spain, vol. 3, 1996, pp 1159–1164.
[51] D. Slezak, Searching for dynamic reducts in inconsistent decision tables, in: Proceeding of IPMU'98, France, vol. 2, 1998, pp 1362–1369.
[52] R. Slowinski, J. Stefanowski, Rough-set reasoning about uncertain data, Foundamenta Informaticae 27 (2–3) (1996) 229–244.
[53] W.Z. Wu, M. Zhang, H.Z. Li, J.S. Mi, Knowledge reduction in random information systems via dempster-shafer theory of evidence, Information Sciences 174 (2005) 143–164.
[54] W. Ziarko, Variable precision rough set model, Journal of Computer System Science 46 (1993) 39–59.
[55] W.X. Zhang, W.Z. Wu, J.Y. Liang, D.Y. Li, Theory and Method of Rough Sets, Science Press, Beijing, China, 2001.
[56] P. Zheng, R. Germano, J. Ariën, K.Y. Qin, Y. Xu, Interpreting and extracting fuzzy decision rules from fuzzy information systems and their inference, 176 (2006) 1869–1897.
[57] W. Zhu, F.Y. Wang, Reduction and axiomization of covering generalized rough sets, Information Sciences 152 (2003) 217–230.
[58] The UCI machine learning repository, <http://mlearn.ics.uci.edu/MLRepository.html>.

Yuhua Qian is a doctoral student of School of Computer and Information Technology at Shanxi University, China. His research interests include rough set theory, granular computing and artificial intelligence. He received the M.S. degree in Computers with applications at Shanxi University (2005).

Chuangyin Dang received a Ph.D. degree in operations research/economics from the University of Tilburg, The Netherlands, in 1991, a M.S. degree in applied mathematics from Xidian University, China, in 1986, and a B.S. degree in computational mathematics from Shanxi University, China, in 1983. He is Associate Professor at the City University of Hong Kong. He is best known for the development of the D1-triangulation of the Euclidean space and the simplicial method for integer programming. His current research interests include computational intelligence, optimization theory and techniques, applied general equilibrium modeling and computation. He is a senior member of IEEE and a member of INFORS and MPS.

Jiye Liang is a professor of School of Computer and Information Technology and Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education at Shanxi University. His research interests include artificial intelligence, granular computing data mining and knowledge discovery. He received the Ph.D degree in Information Science from Xi'an Jiaotong University. He also has a B.S. in computational mathematics from Xi'an Jiaotong University.

Haiyun Zhang is a research assistant of School of Computer and Information Technology at Shanxi University, China. His research interests include rough set theory and artificial intelligence. He received the M.S. degree in Computers with applications at Shanxi University (2007).

Jianmin Ma is a doctoral student of Faculty of Science at Xi'an Jiaotong University, China. His research interests include rough set theory, granular computing and concept lattice.