Available online at www.sciencedirect.com
ScienceDirect
Information Sciences 178 (2008) 181–202
INFORMATION SCIENCES
AN INTERNATIONAL JOURNAL
www.elsevier.com/locate/ins

Measures for evaluating the decision performance of a decision table in rough set theory

Yuhua Qian [a,b,c], Jiye Liang [a,b,*], Deyu Li [a,b], Haiyun Zhang [a,b], Chuangyin Dang [c]

[a] Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Taiyuan 030006, China
[b] School of Computer and Information Technology, Shanxi University, Shanxi, Taiyuan 030006, China
[c] Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, Hong Kong

Received 25 December 2006; received in revised form 1 August 2007; accepted 1 August 2007

Abstract

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

Keywords: Rough set theory; Decision table; Knowledge granulation; Decision evaluation

1. Introduction

Recently, rough set theory developed by Pawlak [17] has become a popular mathematical framework for pattern recognition, image processing, feature selection, neuro computing, conflict analysis, decision support, data mining and knowledge discovery process from large data sets [1,16,20–23]. As applications of rough set theory in decision problems, a number of reduct techniques have been proposed in the last 20 years for information systems and decision tables [2,3,8,9,13–15,18,19,27,30–33,36,37]. As follows, for our further development, we briefly review some of these techniques. β-Reduct proposed by Ziarko [37] provides a kind of

* Corresponding author. Address: Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Taiyuan 030006, China.

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

0020-0255/$ - see front matter
doi:10.1016/j.ins.2007.08.010

attribute-reduction methods in the variable precision rough set model. α-Reduct and α-relative reduct that allow the occurrence of additional inconsistency were proposed in [15] 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 [30,31]. Five kinds of attribute reducts and their relationships in inconsistent systems were investigated by Kryszkiewicz [7], Li [8] and Mi [14], respectively. By eliminating some rigorous conditions required by the distribution reduct, a maximum distribution reduct was introduced by Mi [14]. Unlike the possible reduct in [14], the maximum distribution reduct can derive decision rules that are compatible with the original system.

A set of decision rules can be generated from a decision table by adopting any kind of reduction method mentioned above [6,29,35]. In recent years, how to evaluate the decision performance of a decision rule has become a very important issue in rough set theory. In [3], based on information entropy, Düntsch suggested some uncertainty measures of a decision rule and proposed three criteria for model selection. Moreover, several other measures such as certainty measure and support measure are often used to evaluate a decision rule [5,10,33]. However, all of these measures are only defined for a single decision rule and are not suitable for measuring the decision performance of a rule set. There are two more kinds of measures in the literature [17,19], 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 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 decision table. This shows that the approximation accuracy and consistency degree of a decision table cannot give elaborate depictions of the certainty and consistency for a rule set. To overcome the shortcomings of the existing measures, this paper aims to find some measures for evaluating the decision performance of a set of decision rules. Three new measures are proposed for this objective, which are certainty measure (α), consistency measure (β), and support measure (γ).

The rest of this paper is organized as follows. Some preliminary concepts such as indiscernibility relation, partition, partial relation of knowledge and decision tables 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 decision tables into three types. In Section 4, through some examples, the limitations of the two classical measures are revealed. In Section 5, three new measures (α, β and γ) are introduced for evaluating the decision performance of a set of rules, it is analyzed how each of these three measures depends on the condition granulation and decision granulation of each of the three types of decision tables, and experimental analyses of each of the three measures are performed on three practical data sets. Section 6 concludes this paper with some remarks and discussions.

2. Some basic concepts

In this section, we review some basic concepts such as indiscernibility relation, partition, partial relation of knowledge and decision tables.

An information system (sometimes called a data table, an attribute-value system, a knowledge representation system, etc.), as a basic concept in rough set theory, provides a convenient framework for the representation of objects in terms of their attribute values. An information system S is a pair (U, A), where U is a non-empty, finite set of objects and is called the universe and A is a non-empty, finite set of attributes. For each $a \in A$, a mapping a: $U \to V_a$ is determined by a given decision table, where V_a is the domain of a.

Each non-empty subset $B \subseteq A$ determines an indiscernibility relation in the following way:

$$R_B = \{(x, y) \in U \times U \mid a(x) = a(y), \forall a \in B\}.$$

The relation R_B partitions U into some equivalence classes given by

$$U/R_B = \{[x]_B | x \in U\}, \quad \text{just } U/B,$$

where $[x]_B$ denotes the equivalence class determined by x with respect to B, i.e.,

$$[x]_B = \{y \in U | (x, y) \in R_B\}.$$

Table 1
A decision table about diagnosing rheum

Patients	Headache	Muscle pain	Animal heat	Rheum
e_1	Yes	Yes	Normal	No
e_2	Yes	Yes	High	Yes
e_3	Yes	Yes	Higher	Yes
e_4	No	Yes	Normal	No
e_5	No	No	High	No
e_6	No	Yes	Higher	Yes
e_7	No	No	High	Yes
e_8	No	Yes	Higher	No

We define a partial relation $\preceq$ on the family $\{U/B|\ B \subseteq A\}$ as follows: $U/P \preceq U/Q$ (or $U/Q \succeq U/P$) if and only if, for every $P_i \in U/P$, there exists $Q_j \in U/Q$ such that $P_i \subseteq Q_j$, where $U/P = \{P_1, P_2, \ldots, P_m\}$ and $U/Q = \{Q_1, Q_2, \ldots, Q_n\}$ are partitions induced by $P,Q \subseteq A$, respectively. In this case, we say that Q is coarser than P, or P is finer than Q. If $U/P \preceq U/Q$ and $U/P \neq U/Q$, we say Q is strictly coarser than P (or P is strictly finer than Q), denoted by $U/P \prec U/Q$ (or $U/Q \succ U/P$).

It is clear that $U/P \prec U/Q$ if and only if, for every $X \in U/P$, there exists $Y \in U/Q$ such that $X \subseteq Y$, and there exist $X_0 \in U/P$, $Y_0 \in U/Q$ such that $X_0 \subset Y_0$.

A decision table is an information system $S = (U, C \cup D)$ with $C \cap D = \emptyset$, where an element of C is called a condition attribute, C is called a condition attribute set, an element of D is called a decision attribute, and D is called a decision attribute set. If $U/C \preceq U/D$, then $S = (U, C \cup D)$ is said to be consistent, otherwise it is said to be inconsistent. One can extract certain decision rules from a consistent decision table and uncertain decision rules from an inconsistent decision table. For example, a decision table about diagnosing rheum is given by Table 1.

In Table 1, $U = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8\}$ is the universe, $C = \{c_1, c_2, c_3\} = \{Headache, Muscle\ pain, Animal\ heat\}$ is the condition attribute set, and $D = \{d\} = \{Rheum\}$ is the decision attribute set.

3. Decision rule and knowledge granulation in decision tables

In the first part of this section, we briefly recall the notions of decision rules and certainty measure and support measure of a decision rule in rough set theory.

Definition 1 ([10,33]). Let $S = (U, C \cup D)$ be a decision table, $X_i \in U/C$, $Y_j \in U/D$ and $X_i \cap Y_j \neq \emptyset$. By $\mathrm{des}(X_i)$ and $\mathrm{des}(Y_j)$, we denote the descriptions of the equivalence classes X_i and Y_j in the decision table S. A decision rule is formally defined as

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

Certainty measure and support measure of a decision rule Z_{ij} are defined as follows [10,33]:

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

where $|\cdot|$ is the cardinality of a set. It is clear that the value of each of $\mu(Z_{ij})$ and $s(Z_{ij})$ of a decision rule Z_{ij} falls into the interval $\left[\frac{1}{|U|}, 1\right]$. In subsequent discussions, we denote the cardinality of the set $X_i \cap Y_j$ by $|Z_{ij}|$, which is called the support number of the rule Z_{ij}.

In rough set theory, we can extract some decision rules from a given 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 decision table. In this situation, the lower approximation of the target decision is equal to an empty set in this decision table. To characterize this type of decision tables, in the following, decision tables are classified into three types according to their consistency, which are consistent decision tables, conversely consistent decision tables and mixed decision tables.

As follows, we introduce several new concepts and notations, which will be applied in our further developments. For convenience, by $a(x)$ and $d(x)$, we denote the values of an object x under a condition attribute $a \in C$ and a decision attribute $d \in D$, respectively.

Definition 2. Let $S=(U,C\cup D)$ be a decision table, $U/C=\{X_1,X_2,\ldots,X_m\}$, and $U/D=\{Y_1,Y_2,\ldots,Y_n\}$. A condition class $X_i\in U/C$ is said to be consistent if $d(x)=d(y)$, $\forall x,y\in X_i$ and $\forall d\in D$; a decision class $Y_j\in U/D$ is said to be conversely consistent if $a(x)=a(y)$, $\forall x,y\in Y_j$ and $\forall a\in C$.

It is easy to see that a decision table $S=(U,C\cup D)$ is consistent if every condition class $X_i\in U/C$ is consistent.

Definition 3. Let $S=(U,C\cup D)$ be a decision table, $U/C=\{X_1,X_2,\ldots,X_m\}$, and $U/D=\{Y_1,Y_2,\ldots,Y_n\}$. S is said to be conversely consistent if every decision class $Y_j\in U/D$ is conversely consistent, i.e., $U/D\preceq U/C$. A decision table is called a mixed decision table if it is neither consistent nor conversely consistent.

In addition to the above concepts and notations, we say that $S=(U,C\cup D)$ is strictly consistent (or strictly and conversely consistent) if $U/C\prec U/D$ (or $U/D\prec U/C$).

A strictly and conversely consistent decision table has some practical implications. A strictly and conversely consistent decision table is inconsistent. In a strictly and conversely consistent decision table, there does not exist any condition class $X\in U/C$ and any decision class $Y\in U/D$ such that $X\subseteq Y$. In other words, one can not extract any certain decision rule from a strictly and conversely consistent decision table. Furthermore, when a decision table is strictly and conversely consistent, two well-known classical evaluation measures, approximation accuracy and consistency degree, can not be applied to measure its certainty and consistency. In the remaining part of this paper, one can see that the introduction of the conversely consistency will play an important role in revealing the limitations of the two classical measures and verifying the validity of the evaluation measures proposed in this paper.

In a mixed decision table, from Definition 3, one can see that there exist $X\in U/C$ and $Y\in U/D$ such that X is consistent and Y is conversely consistent. We thus obtain the following results.

- A decision table is strictly consistent iff there does not exist any $Y\in U/D$ such that Y is conversely consistent, and
- a decision table is strictly and conversely consistent iff there does not exist any $X\in U/C$ such that X is consistent.

This implies that a mixed decision table can be transformed into a conversely consistent decision table (a consistent decision table) by deleting its strictly consistent part (by deleting its strictly and conversely consistent part). Hence, in a broad sense, a mixed decision table is a combination of a consistent decision table and a conversely consistent decision table. For this reason, we only focus on the properties of a consistent decision table and a conversely consistent decision table in this paper. For general decision tables, we investigate their characters by practical experimental analyses.

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 [4,11,12,24–26,28,34]. The performance of a decision rule depends directly on the condition granularity and decision granularity of a decision table. In general, the changes of granulation of a decision table can be realized through two ways as follows: (1) refining/coarsening the domain of attributes and (2) adding/reducing attributes. The first approach is mainly used to deal with the case that the attribute values of some elements are imprecise in a decision table. For example, in Table 1, the value of decision attribute Rheum of each element in the universe is either Yes or No. Hence, Rheum degree can not be further analyzed, i.e., the decision values are imprecise. Obviously, decision rules extracting from this kind of decision tables are lack of practicability and pertinence. In the second approach, the certainty measure of a decision rule may be changed through adding or reducing some condition attributes or decision attributes in a decision table. For instance, in Table 1, the certainty measures of some decision rules can be increased by adding new condition attributes.

In general, knowledge granulation is employed to measure the discernibility ability of a knowledge in rough set theory. The smaller knowledge granulation of a knowledge is, the stronger its discernibility ability is. In [10,12], Liang introduced a knowledge granulation $G(A)$ to measure the discernibility ability of a knowledge in an information system, which is given in the following definition.

Definition 4 [10,12]. Let $S = (U, A)$ be an information system and $U/A = \{R_1, R_2, \ldots, R_m\}$. Knowledge granulation of A is defined as

$$G(A) = \frac{1}{|U|^2} \sum_{i=1}^{m} |R_i|^2. \tag{1}$$

Following this definition similarly, for a decision table $S = (U, C \cup D)$, we can 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., $U/C \prec U/D$. Then, there exists at least one decision class in U/D such that it can be represented as the union of more than one condition classes in U/C.*

Proof. Let $U/C = \{X_1, X_2, \ldots, X_m\}$ and $U/D = \{Y_1, Y_2, \ldots, Y_n\}$. By the consistency of S, a decision class $Y \in U/D$ is the union of some condition classes $X \in U/C$. Furthermore, since S is strictly consistent, there exist $X_0 \in U/C$ and $Y_0 \in U/D$ such that $X_0 \subset Y_0$. This indicates that Y_0 is equal to the union of more than one condition classes in U/C. This completes the proof. □

Lemma 2. *Let $S = (U, C \cup D)$ be a strictly and conversely consistent decision table, i.e., $U/D \prec U/C$. Then, there exists at least one condition class in U/C such that it can be represented as the union of more than one decision classes in U/D.*

Proof. The proof is similar to that of Lemma 1. By Lemmas 1 and 2, one can easily obtain the following theorem. □

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

(1) *If S is strictly consistent, then $G(C) < G(D)$; and*
(2) *if S is strictly and conversely consistent, then $G(C) > G(D)$.*

It should be noted that the inverse propositions of Lemmas 1, 2 and Theorem 1 need not be true.

4. Limitations of classical measures for decision tables

In this section, through several illustrative examples, we reveal the limitations of existing classical measures for evaluating the decision performance of a decision table.

In [18], several measures for evaluating a decision rule Z_{ij}: $\mathrm{des}(X_i) \rightarrow \mathrm{des}(Y_j)$ have been introduced, which are certainty measure $\mu(X_i, Y_j) = |X_i \cap Y_j|/|X_i|$ and support measure $s(X_i, Y_j) = |X_i \cap Y_j|/|U|$. However, $\mu(X_i, Y_j)$ and $s(X_i, Y_j)$ are only defined for a single decision rule and are not suitable for measuring the decision performance of a decision-rule set.

In [18], approximation accuracy of a classification was introduced by Pawlak. Let $F = \{Y_1, Y_2, \ldots, Y_n\}$ be a classification of the universe U, and C a condition attribute set. Then, C-lower and C-upper approximations of F are given by $\underline{C}F = \{\underline{C}Y_1, \underline{C}Y_2, \ldots, \underline{C}Y_n\}$ and $\overline{C}F = \{\overline{C}Y_1, \overline{C}Y_2, \ldots, \overline{C}Y_n\}$, respectively, where

$$\underline{C}Y_i = \bigcup\{x \in U | [x]_C \subseteq Y_i \in F\}, \quad 1 \leqslant i \leqslant n,$$

and

$$\overline{C}Y_i = \bigcup\{x \in U | [x]_C \cap Y_i \neq \emptyset, Y_i \in F\}, \quad 1 \leqslant i \leqslant n.$$

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

Example 1. Let $S_1 = (U, C \cup D_1)$ and $S_2 = (U, C \cup D_2)$ be two decision tables with the same universe U. Suppose that

$$U/C = \{\{e_1, e_2\}, \{e_3, e_4, e_5\}, \{e_6, e_7\}\},$$

$$U/D_1 = \{\{e_1\}, \{e_2, e_3, e_4\}, \{e_5, e_6\}, \{e_7\}\},$$

and

$$U/D_2 = \{\{e_1, e_3\}, \{e_2, e_4, e_6\}, \{e_5\}, \{e_7\}\}.$$

Then, six decision rules extracted from S_1 and their certainty measures and support measures corresponding to each individual rule are given by

$r_1 : \mathrm{des}(\{e_1, e_2\}) \to \mathrm{des}(\{e_1\}),\ \mu(r_1) = \frac{1}{2},\ s(r_1) = \frac{1}{7};$
$r_2 : \mathrm{des}(\{e_1, e_2\}) \to \mathrm{des}(\{e_2\}),\ \mu(r_2) = \frac{1}{2},\ s(r_2) = \frac{1}{7};$
$r_3 : \mathrm{des}(\{e_3, e_4, e_5\}) \to \mathrm{des}(\{e_3, e_4\}),\ \mu(r_3) = \frac{2}{3},\ s(r_3) = \frac{2}{7};$
$r_4 : \mathrm{des}(\{e_3, e_4, e_5\}) \to \mathrm{des}(\{e_5\}),\ \mu(r_4) = \frac{1}{3},\ s(r_4) = \frac{1}{7};$
$r_5 : \mathrm{des}(\{e_6, e_7\}) \to \mathrm{des}(\{e_6\}),\ \mu(r_5) = \frac{1}{2},\ s(r_5) = \frac{1}{7};$
$r_6 : \mathrm{des}(\{e_6, e_7\}) \to \mathrm{des}(\{e_7\}),\ \mu(r_6) = \frac{1}{2},\ s(r_6) = \frac{1}{7}.$

Furthermore, seven decision rules extracted from S_2 and their certainty measures and support measures corresponding to each individual rule are given by

$r'_1 : \mathrm{des}(\{e_1, e_2\}) \to \mathrm{des}(\{e_1\}),\ \mu(r'_1) = \frac{1}{2},\ s(r'_1) = \frac{1}{7};$
$r_{2'} : \mathrm{des}(\{e_3, e_4, e_5\}) \to \mathrm{des}(\{e_3\}),\ \mu(r'_2) = \frac{1}{3},\ s(r'_2) = \frac{1}{7};$
$r'_3 : \mathrm{des}(\{e_1, e_2\}) \to \mathrm{des}(\{e_2\}),\ \mu(r'_3) = \frac{1}{2},\ s(r'_3) = \frac{1}{7};$
$r'_4 : \mathrm{des}(\{e_3, e_4, e_5\}) \to \mathrm{des}(\{e_4\}),\ \mu(r'_4) = \frac{1}{3},\ s(r'_4) = \frac{1}{7};$
$r'_5 : \mathrm{des}(\{e_6, e_7\}) \to \mathrm{des}(\{e_6\}),\ \mu(r'_5) = \frac{1}{2},\ s(r'_5) = \frac{1}{7};$
$r'_6 : \mathrm{des}(\{e_3, e_4, e_5\}) \to \mathrm{des}(\{e_5\}),\ \mu(r'_6) = \frac{1}{3},\ s(r'_6) = \frac{1}{7};$
$r'_7 : \mathrm{des}(\{e_6, e_7\}) \to \mathrm{des}(\{e_7\}),\ \mu(r'_7) = \frac{1}{2},\ s(r'_7) = \frac{1}{7}.$

By formula (2), we have that

$$a_C(U/D_1) = \frac{\sum_{Y_i \in U/D_1} |\underline{C}Y_i|}{\sum_{Y_i \in U/D_1} |\overline{C}Y_i|} = \frac{0}{2+5+5+2} = 0,$$

$$a_C(U/D_2) = \frac{\sum_{Y_i \in U/D_2} |\underline{C}Y_i|}{\sum_{Y_i \in U/D_2} |\overline{C}Y_i|} = \frac{0}{5+7+3+2} = 0.$$

That is to say $a_C(U/D_1) = a_C(U/D_2) = 0$.

Now, let us consider the average value of the certainty measure of each of the two rule sets extracted one from S_1 and the other from S_2. Taking the average of the certainty-measure values corresponding to decision rules for each decision table, we have that

$$\frac{1}{6}\sum_{i=1}^{6} \mu(r_i) = \frac{1}{6}\left(\frac{1}{2} + \frac{1}{2} + \frac{2}{3} + \frac{1}{3} + \frac{1}{2} + \frac{1}{2}\right) = \frac{1}{2}$$

and

$$\frac{1}{7}\sum_{i=1}^{7} \mu(r'_i) = \frac{1}{7}\left(\frac{1}{2} + \frac{1}{3} + \frac{1}{2} + \frac{1}{3} + \frac{1}{2} + \frac{1}{3} + \frac{1}{2}\right) = \frac{3}{7}.$$

Obviously, $\frac{1}{2} = \frac{7}{14} > \frac{6}{14} = \frac{3}{7}$. It implies that the decision table S_1 has a higher certainty than S_2 on the average. However, this situation is not revealed by the approximation accuracy. Therefore, a more comprehensive and effective measure for evaluating the certainty of a decision table is needed.

The consistency degree of a decision table $S = (U, C \cup D)$, another classical measure proposed in [18], is defined as

$$c_C(D) = \frac{\sum_{i=1}^{n} |\underline{C}Y_i|}{|U|}. \tag{3}$$

The consistency degree expresses the percentage of objects which can be correctly classified into 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. Similar to Example 1, however, the consistency of a decision table also cannot be well characterized by the classical 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 a 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 decision table, then $\alpha_C(U/D) = 0$ and $c_C(D) = 0$.*

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

Remark. From the above analyses, it is easy to see that the shortcomings of the two classical measures are mainly caused by the condition equivalence classes that can not be included in the lower approximation of the target decision in a given decision table. As we know, in an inconsistent decision table, there must exist some condition equivalence classes that can not be included in the lower approximation of the target decision. In fact, for a strictly and conversely consistent decision table, the lower approximation of the target decision is an empty set. Hence, we can make a conclusion that the approximation accuracy and consistency degree can not be employed to effectively characterize the decision performance of an inconsistent decision table. To overcome this shortcoming of the two classical measures, the effect of the condition equivalence classes 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 decision table.

5. Evaluation of the decision performance of a rule set

To overcome the shortcomings of the two classical measures, in this section, we introduce three new measures (α, β and γ) for evaluating the decision performance of a decision table and analyze how each of these three measures depends on the condition granulation and decision granulation of each of consistent decision tables and conversely consistent decision tables. For general decision tables, by employing three decision tables from real world, we illustrate the advantage of these three measures for evaluating the decision performance of a decision rule set extracted from a decision table.

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

$$\alpha(S) = \sum_{i=1}^{m}\sum_{j=1}^{n} s(Z_{ij})\mu(Z_{ij}) = \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{|X_i \cap Y_j|^2}{|U||X_i|}, \tag{4}$$

where $s(Z_{ij})$ and $\mu(Z_{ij})$ are the certainty measure and support measure of the rule Z_{ij}, respectively.

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

(1) *For any* $Z_{ij} \in RULE$, *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. From the definitions of $\mu(Z_{ij})$ and $s(Z_{ij})$, it follows that $\frac{1}{|U|} \leqslant \mu(Z_{ij}) \leqslant 1$ and $\sum_{i=1}^{m}\sum_{j=1}^{n} s(Z_{ij}) = \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{|X_i \cap Y_j|}{|U|} = 1$.

(1) If $\mu(Z_{ij}) = 1$ for any $Z_{ij} \in \text{RULE}$, then we have that

$$\alpha(S) = \sum_{i=1}^{m}\sum_{j=1}^{n} s(Z_{ij})\mu(Z_{ij}) = \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{|X_i \cap Y_j|^2}{|U||X_i|} = \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{|X_i \cap Y_j|}{|U|} \times 1 = 1.$$

(2) If $m = 1$ and $n = |U|$, then $\mu(Z_{ij}) = \frac{1}{|U|}$ for any $Z_{ij} \in \text{RULE}$. In this case, we have that

$$\alpha(S) = \sum_{i=1}^{m}\sum_{j=1}^{n} s(Z_{ij})\mu(Z_{ij}) = \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{|X_i \cap Y_j|^2}{|U||X_i|} = \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{|X_i \cap Y_j|}{|U|} \times \frac{1}{|U|} = \frac{1}{|U|}.$$

This completes the proof. □

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., the 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, how the measure α overcomes the limitation of the classical measure $a_C(U/D)$ can be illustrated.

Example 2 (Continued from Example 1). Computing the measure α, we have that

$$\begin{aligned}\alpha(S_1) &= \sum_{i=1}^{m}\sum_{j=1}^{n} s(Z_{ij})\mu(Z_{ij}) \\ &= \frac{1}{7}\cdot\frac{1}{2} + \frac{1}{7}\cdot\frac{1}{2} + \frac{2}{7}\cdot\frac{2}{3} + \frac{1}{7}\cdot\frac{1}{3} + \frac{1}{7}\cdot\frac{1}{2} + \frac{1}{7}\cdot\frac{1}{2} = \frac{11}{21}, \\ \alpha(S_2) &= \sum_{i=1}^{m}\sum_{j=1}^{n} s(Z_{ij})\mu(Z_{ij}) \\ &= \frac{1}{7}\cdot\frac{1}{2} + \frac{1}{7}\cdot\frac{1}{2} + \frac{1}{7}\cdot\frac{1}{3} + \frac{1}{7}\cdot\frac{1}{3} + \frac{1}{7}\cdot\frac{1}{3} + \frac{1}{7}\cdot\frac{1}{2} + \frac{1}{7}\cdot\frac{1}{2} = \frac{9}{21}.\end{aligned}$$

Therefore, $\alpha(S_1) > \alpha(S_2)$.

Example 2 indicates that unlike the 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 is equal to an empty set.

Remark. From formula (2), it follows that $a_C(U/D) = 0$ if $\bigcup_{Y_i \in U/D} \underline{C}Y_i = \emptyset$. In fact, in a broad sense, $\underline{C}Y_i = \emptyset$ does not imply that the certainty of a rule concerning Y_i is equal to 0. So the measure α is much better than the approximation accuracy for measuring the certainty of a decision-rule set when a decision table is strictly and conversely consistent.

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 decision tables. If* $U/C_1 = U/C_2$ *and* $U/D_2 \prec U/D_1$, *then* $\alpha(S_1) > \alpha(S_2)$.

Proof. From $U/C_1 = U/C_2$ and the converse consistencies of S_1 and S_2, it follows that there exist $X_p \in U/C_1$ and $Y_q \in U/D_1$ such that $Y_q \subseteq X_p$. Since $U/D_2 \prec U/D_1$, there exist $Y_q^1, Y_q^2, \ldots, Y_q^s \in U/D_2$ $(s > 1)$ such that $Y_q = \bigcup_{k=1}^{s} Y_q^k$. 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 $|Z_{pq}| = \sum_{k=1}^{s}|Z_{pq}^k|$. Thus, $|Z_{pq}|^2 > \sum_{k=1}^{s}|Z_{pq}^k|^2$. Therefore, from the definition of $\alpha(S)$, it follows that $\alpha(S_1) > \alpha(S_2)$. This completes the proof.

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

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

Proof. From $U/C_2 \prec U/C_1$, there exist $X_l \in U/C_1$ and an integer $s > 1$ such that $X_l = \bigcup_{k=1}^{s} X_l^k$, where $X_l^k \in U/C_2$. It is clear that $|X_l| = \sum_{k=1}^{s}|X_l^k|$. Therefore, $\frac{1}{|X_l|} < \frac{1}{|X_l^1|} + \frac{1}{|X_l^2|} + \cdots + \frac{1}{|X_l^s|}$.

Noticing that both S_1 and S_2 are conversely consistent, we have $| Z_{lq} |=| Z_{lq}^k |$ $(k = 1,2,\cdots,s)$.

Thus,

$$\alpha(S_1) = \sum_{i=1}^{m}\sum_{j=1}^{n} s(Z_{ij})\mu(Z_{ij}) = \frac{1}{|U|}\sum_{i=1}^{l-1}\sum_{j=1}^{n}\frac{|Z_{ij}|^2}{|X_i|} + \frac{1}{|U|}\sum_{j=1}^{n}\frac{|Z_{lj}|^2}{|X_l|} + \frac{1}{|U|}\sum_{i=l+1}^{m}\sum_{j=1}^{n}\frac{|Z_{ij}|^2}{|X_i|}$$
$$< \frac{1}{|U|}\sum_{i=1}^{l-1}\sum_{j=1}^{n}\frac{|Z_{ij}|^2}{|X_i|} + \frac{1}{|U|}\sum_{k=1}^{s}\sum_{j=1}^{n}\frac{|Z_{lj}|^2}{|X_l^k|} + \frac{1}{|U|}\sum_{i=l+1}^{m}\sum_{j=1}^{n}\frac{|Z_{ij}|^2}{|X_i|} = \alpha(S_2).$$

This completes the proof. □

Theorem 4 states that the certainty measure α of a conversely consistent decision table increases with its condition classes becoming finer.

For a general decision table, in the following, through experimental analyses, we illustrate the validity of the measure α for assessing the decision performance of a decision-rule set extracted from the decision table. In order to verify the advantage of the measure α over the approximation accuracy $a_C(U/D)$, we have downloaded three public data sets with practical applications from UCI Repository of machine learning databases [38], which are described in Table 2. All condition attributes and decision attributes in the three data sets are discrete.

In Table 2, the data set Tic-tac-toe is the encoding of the complete set of possible board configurations at the end of tie-tac-toe games, which is used to obtain possible ways to create a "three-in-a-row"; the data set Dermatology is a decision table about diagnosing dermatosis according to some clinical features, which is used to extract general diagnosing rules; and Nursery data set is derived from a hierarchical decision model originally developed to rank applications for nursery schools.

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.

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 approximation accuracy will achieve the same value 1 if the decision table becomes consistent through adding the number of selected features. However, from Fig. 1, it is easy to see that the values of approximation accuracy equal to zero when the number of features equals 1 or 2. In this situation, the lower approximation of the target decision equals an empty set in the decision table. Hence, the approximation accuracy cannot be used to effectively characterize the certainty of the decision table when the value of approximation accuracy equals 0. But, for the same

Table 2
Data sets description

Data sets	Samples	Condition features	Decision classes
Tic-tac-toe	958	9	2
Dermatology	366	33	6
Nursery	12,960	8	5

Table 3
$a_C(D)$ and α with different numbers of features in the data set Tie-tac-toe

Measure	Features								
	1	2	3	4	5	6	7	8	9
$a_C(D)$	0.0000	0.0000	0.0668	0.0886	0.2647	0.6348	0.8933	1.0000	1.0000
α	0.5557	0.5661	0.6414	0.6650	0.7916	0.9000	0.9718	1.0000	1.0000

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

Measure	Features										
	3	6	9	12	15	18	21	24	27	30	33
$a_C(D)$	0.0010	0.2244	0.6358	0.8458	0.9625	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
α	0.3006	0.7167	0.9144	0.9688	0.9909	1.0000	1.0000	1.0000	1.0000	1.0000	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	0.3425	0.4292	0.4323	0.4437	0.4609	0.4720	0.4929	1.0000

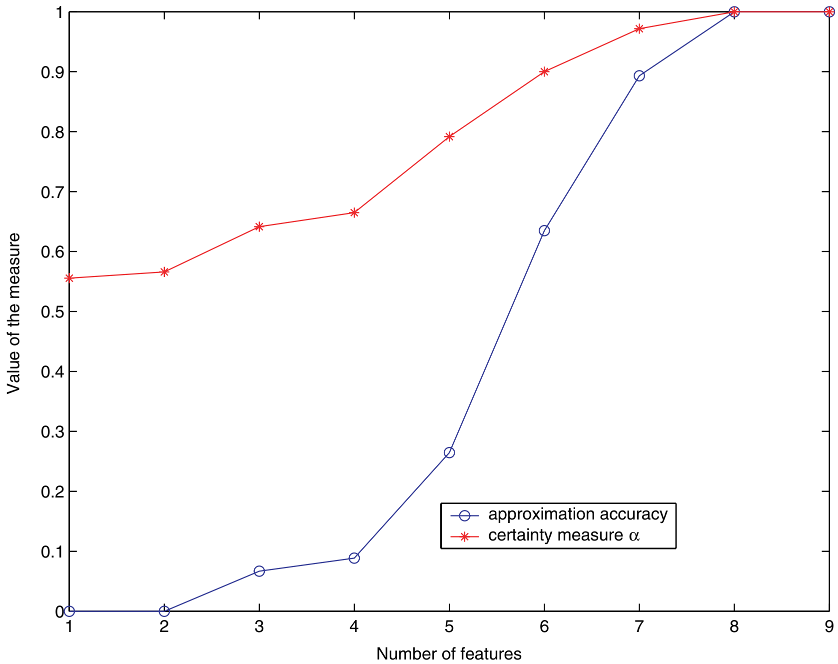

Fig. 1. Variation of the certainty measure α and the approximation accuracy with the number of features (data set Tie-tac-toe).

situation as that the numbers of features equal 1 and 2, the values of the certainty measure α equal 0.557 and 0.5661, respectively. It shows that unlike the approximation accuracy, the certainty measure of the decision table with two features is higher than that of the decision table with only one feature. Hence, the measure α is much better than the approximation accuracy for an inconsistent decision table. We can make the same

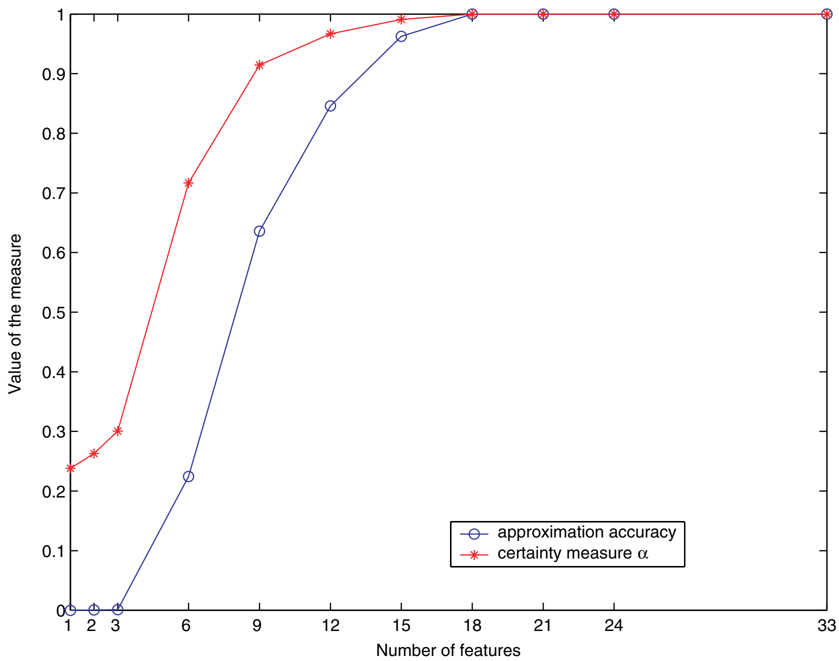

Fig. 2. Variation of the certainty measure α and the approximation accuracy with the number of features (data set Dermatology).

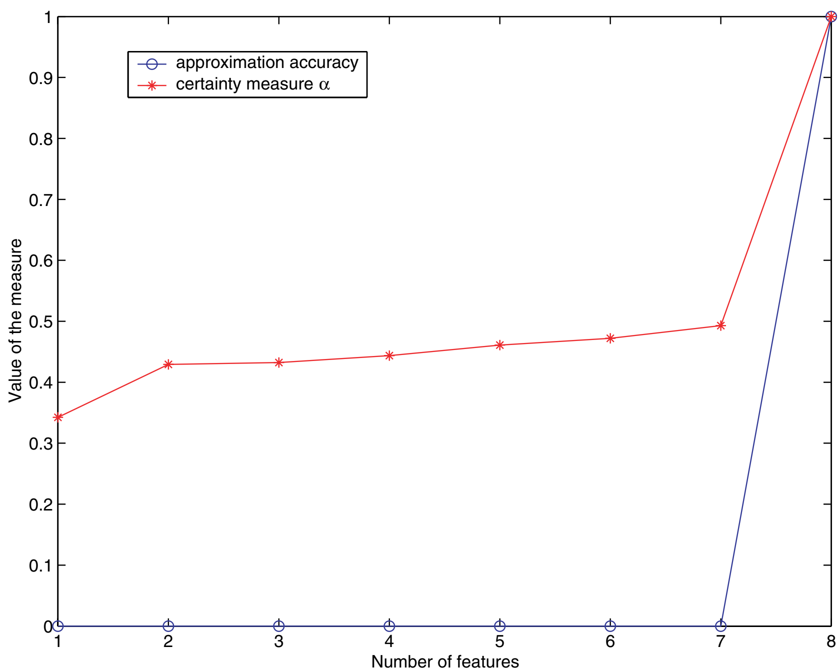

Fig. 3. Variation of the certainty measure α and the approximation accuracy with the number of features (data set Nursery).

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. Hence, the measure α may be better than the approximation accuracy for evaluating the certainty of a decision table.

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

Now, we introduce a measure β to evaluate the consistency of a set of decision rules extracted from a decision table.

Definition 6. Let $S=(U,C\cup D)$ be a decision table and $\mathrm{RULE}=\{Z_{ij}—Z_{ij}\text{: } \mathrm{des}(X_i)\to\mathrm{des}(Y_j),X_i\in U/C,Y_j\in U/D\}$. Consistency measure β of S is defined as

$$\beta(S)=\sum_{i=1}^{m}\frac{|X_i|}{|U|}[1-\frac{4}{|X_i|}\sum_{j=1}^{N_i}|X_i\cap Y_j|\mu(Z_{ij})(1-\mu(Z_{ij}))], \tag{5}$$

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

Theorem 5 (*Extremum*). *Let $S=(U,C\cup D)$ be a decision table and $RULE=\{Z_{ij}|Z_{ij}{:}des(X_i)\to des(Y_j),X_i\in U/C,\ Y_j\in U/D\}$.*

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

Proof. From the definition of $\mu(Z_{ij})$, it follows that $\frac{1}{|U|}\leqslant\mu(Z_{ij})\leqslant 1$.

(1) If $\mu(Z_{ij})=1$ for all $Z_{ij}\in\mathrm{RULE}$, then we have that

$$\beta(S)=\sum_{i=1}^{m}\frac{|X_i|}{|U|}[1-\frac{4}{|X_i|}\sum_{j=1}^{N_i}|X_i\cap Y_j|\mu(Z_{ij})(1-\mu(Z_{ij}))]=\sum_{i=1}^{m}\frac{|X_i|}{|U|}[1-\frac{4}{|X_i|}\sum_{j=1}^{N_i}|X_i\cap Y_j|\times 1\times(1-1)]$$
$$=\sum_{i=1}^{m}\frac{|X_i|}{|U|}=1.$$

(2) If $\mu(Z_{ij})=\frac{1}{2}$ for all $Z_{ij}\in\mathrm{RULE}$, then we have that

$$\beta(S)=\sum_{i=1}^{m}\frac{|X_i|}{|U|}\left[1-\frac{4}{|X_i|}\sum_{j=1}^{N_i}|X_i\cap Y_j|\mu(Z_{ij})(1-\mu(Z_{ij}))\right]=\sum_{i=1}^{m}\frac{|X_i|}{|U|}\left[1-\frac{4}{|X_i|}\sum_{j=1}^{N_i}|X_i\cap Y_j|\times\frac{1}{4}\right]$$
$$=\sum_{i=1}^{m}\frac{|X_i|}{|U|}\left[1-\frac{1}{|X_i|}\sum_{j=1}^{N_i}|X_i\cap Y_j|\right]=\sum_{i=1}^{m}\frac{|X_i|}{|U|}[1-1]=0.$$

This completes the proof. □

It should be noted that the measure β achieves its maximum one when $S=(U,C\cup D)$ is a consistent decision table.

Remark. Unlike the consistency degree $c_C(U/D)$, the measure β can be used to evaluate the consistency of a decision-rule set when $c_C(U/D)=0$, i.e., the lower approximation of each decision class is equal to an empty set. From formula (3), it follows that $c_C(D)=0$ if $\bigcup_{Y_i\in U/D}\underline{C}Y_i=\varnothing$. In fact, as we know, $\underline{C}Y_i=\varnothing$ does not imply that the certainty of a rule concerning Y_i is equal to 0. So the measure β is much better than the classical consistency degree for measuring the consistency of a decision-rule set when decision tables are strictly and conversely consistent.

The monotonicity of the measure β on conversely consistent decision tables can be found in the following Theorem 6 and 7.

Theorem 6. *Let $S_1=(U,C_1\cup D_1)$ and $S_2=(U,C_2\cup D_2)$ be two conversely consistent decision tables. If $U/C_1=U/C_2$ and $U/D_2\prec U/D_1$, 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 $U/C_1 = U/C_2$ and the conversely consistencies of S_1 and S_2, there exist $X_p \in U/C_1$ and $Y_q \in U/D_1$ such that $Y_q \subseteq X_p$. By $U/D_2 \prec U/D_1$, we derive that there exist $Y_q^1, Y_q^2, \ldots, Y_q^s \in U/D_2$ $(s > 1)$ such that $Y_q = \bigcup_{k=1}^{s} Y_q^k$. 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 $|Z_{pq}| = \sum_{k=1}^{s} |Z_{pq}^k|$.

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. So, we know that if $x_l \in X_i \cap\ Y_j$, then $\delta_D(Z_{il}) = \mu(Z_{ij})$ holds.

Hence, it follows that

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

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

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

Similar to this idea, $\beta(S_1) > \beta(S_2)$ when $\forall\mu(Z_{ij}) \geqslant \frac{1}{2}$ can be proved.This completes the proof. □

Theorem 6 states that the consistency measure β of a conversely consistent 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 7. *Let $S_1 = (U, C_1 \cup D_1)$ and $S_2 = (U, C_2 \cup D_2)$ be two conversely consistent decision tables. If $U/D_1 = U/D_2$ and $U/C_2 \prec U/C_1$, 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. So, we know that if $x_l \in X_i \cap Y_j$, then $\delta_C(Z_{il}) = \mu(Z_{ij})$ holds.

From $U/C_2 \prec U/C_1$, we know there exist $X_p \in U/C_1$ and an integer $s > 1$ such that $X_p = \bigcup_{k=1}^{s} X_p^k$, where $X_p^k \in U/C_2$. It is clear that $|X_l| = \sum_{k=1}^{s}|X_l^k|$, and $|X_p^k| < |X_p|$ for every $X_p^k \in U/C_2$. From the converse consistencies of S_1 and S_2, it follows that

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

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

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

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

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

Theorem 7 states that the consistency measure β of a conversely consistent decision table decreases with its condition classes becoming finer when $\forall\mu(Z_{ij}) \leqslant \frac{1}{2}$, and increases with its condition classes becoming finer when $\forall\mu(Z_{ij}) \geqslant \frac{1}{2}$.

For general 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 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 decision table becomes consistent through adding the number of selected features, the measure β and the 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 equals 1 or 2. In this situation, the lower approximation of the target decision in the decision table equals an empty set. Hence, the consistency degree cannot be used to effectively characterize the consistency of the decision table when the value of the consistency degree equals 0. But, for the same situation as that the numbers of features equal 1 and 2, the values of the consistency measure β equal 0.1114 and 0.1322, respectively. It shows that unlike the consistency degree, the consistency measure β of the decision table with two features is higher than that of the decision table with only one feature. Therefore, the measure β is much better than the consistency degree for an inconsistent decision table. Obviously, we can make 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 0. Given this advantage, we may conclude that the measure β is much better than the classical consistency degree for evaluating the consistency of a decision table.

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

Finally, we consider how to define a better support measure for evaluating a decision-rule set.

Table 6
$c_C(D)$ and β with different numbers of features in the data set Tie-tac-toe

Measure	Features								
	1	2	3	4	5	6	7	8	9
$c_C(D)$	0.0000	0.0000	0.1253	0.1628	0.4186	0.7766	0.9436	1.0000	1.0000
β	0.1114	0.1322	0.2827	0.3300	0.5832	0.8000	0.9436	1.0000	1.0000

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

Measure	Features										
	3	6	9	12	15	18	21	24	27	30	33
$c_C(D)$	0.0055	0.4372	0.8060	0.9290	0.9809	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
β	0.3101	0.5285	0.8471	0.9429	0.9818	1.0000	1.0000	1.0000	1.0000	1.0000	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

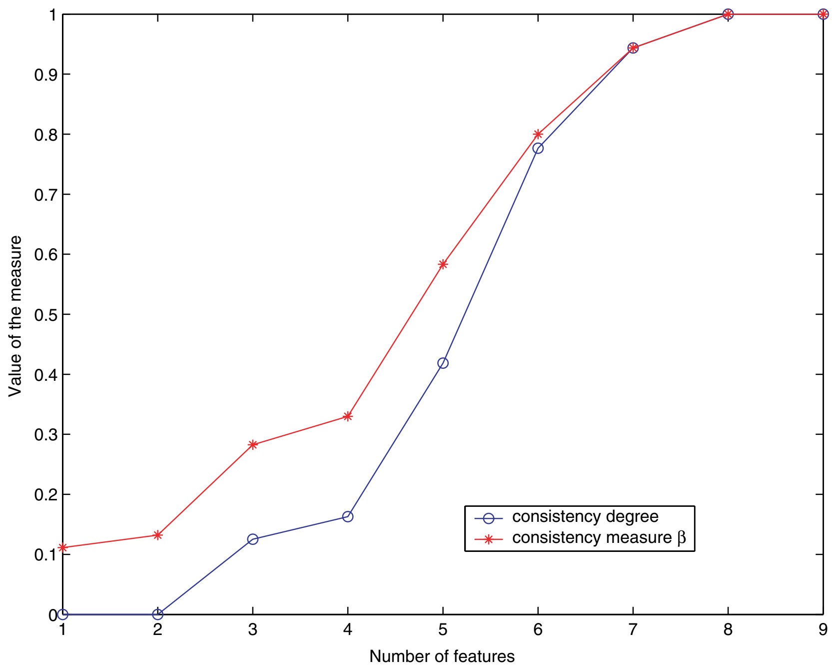

Fig. 4. Variation of the consistency measure β and the consistency degree with the number of features (data set Tie-tac-toe).

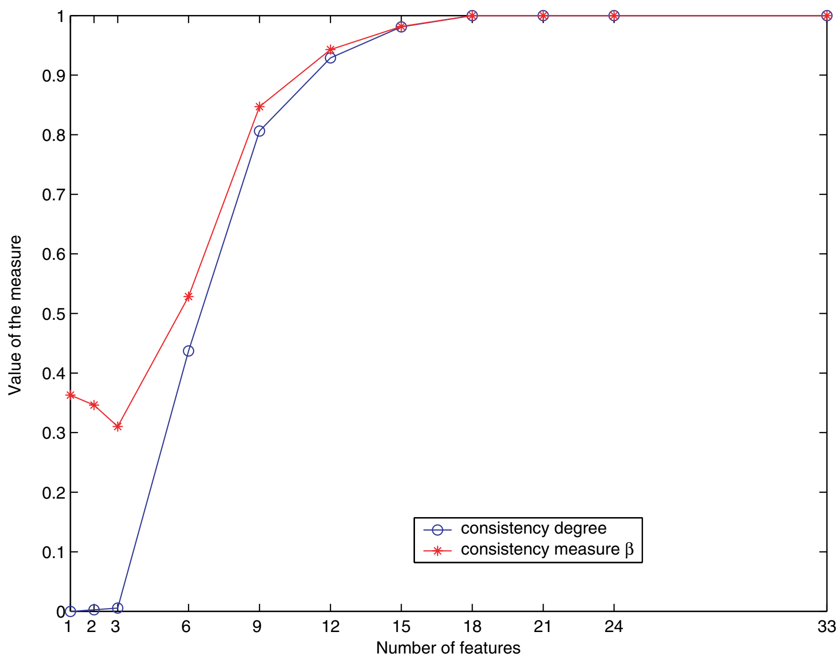

Fig. 5. Variation of the consistency measure β and the consistency degree with the number of features (data set Dermatology).

Let $S = (U, C \cup D)$ be a decision table and RULE $= \{Z_{ij}$—Z_{ij}: des$(X_i) \rightarrow$ des$(Y_j), X_i \in U/C, Y_j \in U/D\}$. Intuitively, the mean value of the support measures $S(Z_{ij})$ of all rules Z_{ij} seems to be suitable for this task. However, the following example indicates our intuition unreliable.

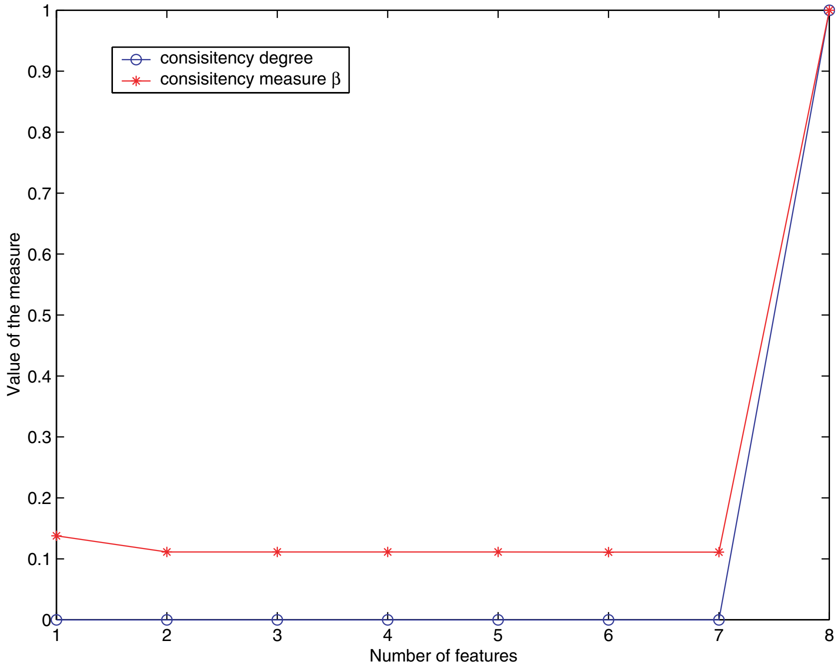

Fig. 6. Variation of the consistency measure β and the consistency degree with the number of features (data set Nursery).

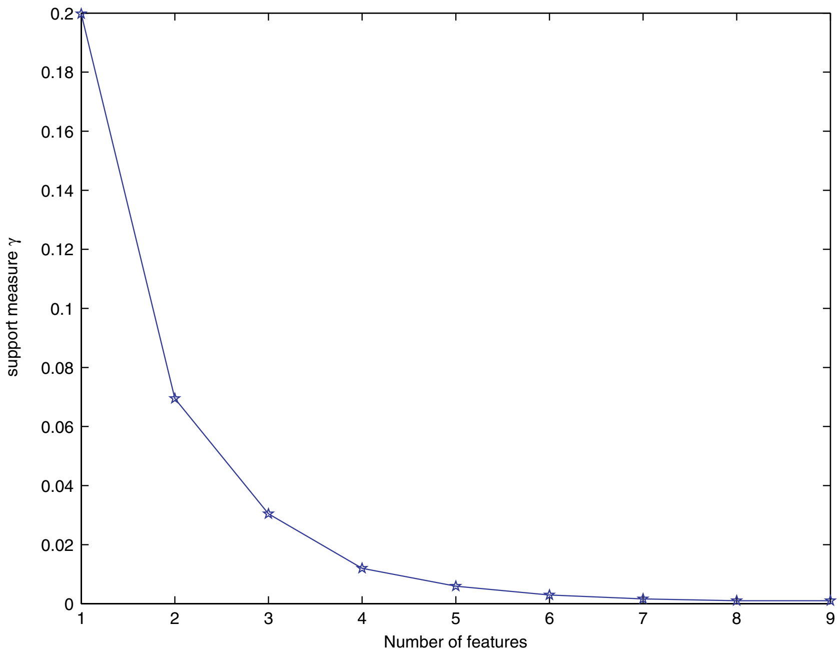

Fig. 7. Variation of the support measure γ with the number of features (data set Tie-tac-toe).

Example 3. Let $S_1 = (U, C_1 \cup D_1)$ and $S_2 = (U, C_2 \cup D_2)$ be two decision tables with the same universe U. Suppose that

$$U/C_1 = \{\{e_1, e_2\}, \{e_3, e_4\}, \{e_5, e_6, e_7, e_8, e_9\}\},$$
$$U/D_1 = \{\{e_1, e_2, e_3, e_4\}, \{e_5, e_6, e_7, e_8, e_9\}\},$$

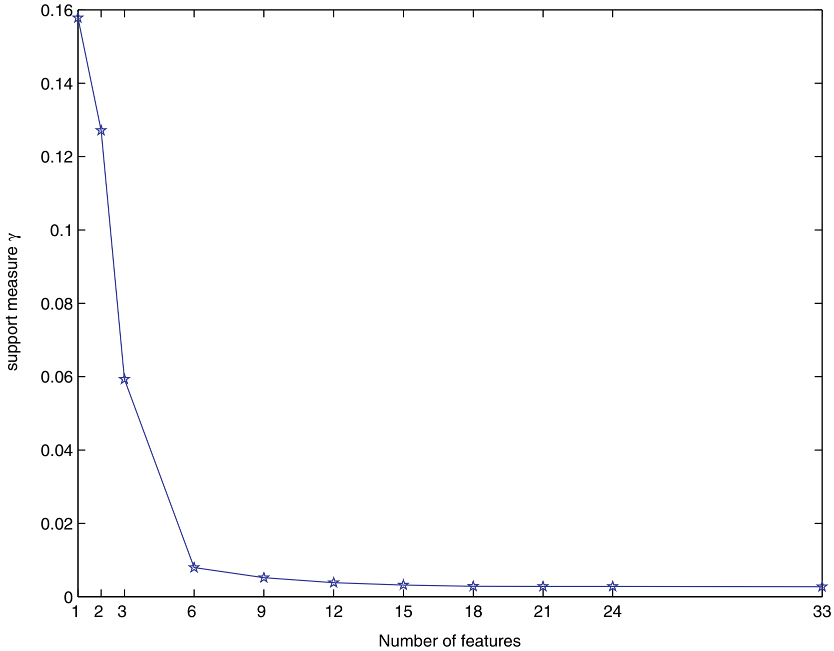

Fig. 8. Variation of the support measure γ with the number of features (data set Dermatology).

$$U/C_2 = \{\{e_1,e_2,e_3,e_4,e_5,e_6\},\{e_7,e_8,e_9\}\},$$
$$U/D_2 = \{\{e_1,e_2,e_3\},\{e_4,e_5,e_6\},\{e_7,e_8,e_9\}\}.$$

By taking the average value, it follows that

$$ws(S_1) = \frac{1}{|RULE|}\sum_{i=1}^{m}\sum_{j=1}^{n} s(Z_{ij}) = \frac{1}{3}\left(\frac{2}{9}+\frac{2}{9}+\frac{5}{9}\right) = \frac{1}{3};$$
$$ws(S_2) = \frac{1}{|RULE|}\sum_{i=1}^{m}\sum_{j=1}^{n} s(Z_{ij}) = \frac{1}{3}\left(\frac{3}{9}+\frac{3}{9}+\frac{3}{9}\right) = \frac{1}{3}.$$

Therefore, in this case, $ws(S_1) = ws(S_2)$.

In fact, the weight information of each decision rule has not been considered in this measure. Hence, it may not be able to effectively characterize the the support measure of a complete decision table.

In the following, we define a more effective support measure γ for evaluating the support of a decision-rule set.

Definition 7. Let $S = (U, C \cup D)$ be a decision table and RULE $= \{Z_{ij}|Z_{ij}\text{: des}(X_i) \rightarrow \text{des}(Y_j), X_i \in U/C, Y_j \in U/D\}$. Support measure γ of S is defined as

$$\gamma(S) = \sum_{i=1}^{m}\sum_{j=1}^{n} s^2(Z_{ij}) = \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{|X_i \cap Y_j|^2}{|U|^2}, \tag{6}$$

where $s(Z_{ij})$ is the support measure of the rule Z_{ij}.

Theorem 8 (*Extremum*). *Let $S = (U, C \cup D)$ be a decision table and $RULE = \{Z_{ij}—Z_{ij}{:}des(X_i) \rightarrow des(Y_j), X_i \in U/C, Y_j \in U/D\}$.*

(1) *If $m = n = 1$, then the measure γ achieves its maximum value 1, and*
(2) *if $m = |U|$ or $n = |U|$, then the measure γ achieves its minimum value $\frac{1}{|U|}$.*

Proof. From the definition of $s(Z_{ij})$, it follows that $\frac{1}{|U|} \leqslant \mu(Z_{ij}) \leqslant 1$ and $\sum_{i=1}^{m}\sum_{j=1}^{n} s(Z_{ij}) = \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{|X_i \cap Y_j|}{|U|} = 1$.

(1) If $m = n = 1$, then $s(Z_{ij}) = \frac{|X_i \cap Y_j|}{|U|} = \frac{|U|}{|U|} = 1$. Therefore, one can obtain that $\gamma(S) = \sum_{i=1}^{m}\sum_{j=1}^{n} s^2(Z_{ij}) = 1$.
(2) If $m = |U|$ or $n = |U|$, then $s(Z_{ij}) = \frac{1}{|U|}$ for all $Z_{ij} \in \mathrm{RULE}$. Hence, $\gamma(S) = \sum_{i=1}^{m}\sum_{j=1}^{n} s^2(Z_{ij}) = \sum_{i=1}^{m} \frac{1}{|U|^2} = \frac{1}{|U|}$.

This completes the proof. □

Example 4 (Continued from Example 3). From the definition of the measure γ, it follows that

$$\gamma(S_1) = \sum_{i=1}^{m}\sum_{j=1}^{n} s^2(Z_{ij}) = \left(\frac{2}{9}\right)^2 + \left(\frac{2}{9}\right)^2 + \left(\frac{5}{9}\right)^2 = \frac{34}{81} \quad \text{and}$$

$$\gamma(S_2) = \sum_{i=1}^{m}\sum_{j=1}^{n} s^2(Z_{ij}) = \left(\frac{3}{9}\right)^2 + \left(\frac{3}{9}\right)^2 + \left(\frac{3}{9}\right)^2 = \frac{27}{81}.$$

Therefore, $\gamma(S_1) > \gamma(S_2)$.

Example 4 indicates that the measure γ may be better than the measure ws used in Example 3 for evaluating a decision-rule set.

Theorem 9. *Let $S_1 = (U, C_1 \cup D_1)$ and $S_2 = (U, C_2 \cup D_2)$ be two decision tables. Then, $\gamma(S_1) < \gamma(S_2)$ if and only if $G(C_1 \cup D_1) < G(C_2 \cup D_2)$.*

Proof. Suppose that $U/(C \cup D) = \{X_i \cap Y_j | X_i \cap Y_j \neq \emptyset,\ X_i \in U/C_1, Y_j \in U/D\}$ and $\mathrm{RULE} = \{Z_{ij}—Z_{ij}: X_i \to Y_j, X_i \in U/C, Y_j \in U/D\}$. From Definition 4 and $s(Z_{ij}) = \frac{|X_i \cap Y_j|}{|U|}$, it follows that

$$G(C \cup D) = \frac{1}{|U|^2}\sum_{i=1}^{m}\sum_{j=1}^{n}|X_i \cap Y_j|^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{|X_i \cap Y_j|}{|U|}\right)^2 = \sum_{i=1}^{m}\sum_{j=1}^{n} s^2(Z_{ij}) = \gamma(S).$$

Therefore, $\gamma(S_1) < \gamma(S_2)$ if and only if $G(C_1 \cup D_1) < G(C_2 \cup D_2)$. This completes the proof. □

Theorem 9 states that the support measure γ of a decision table increases with the granulation of the decision table becoming bigger. As a direct result of Theorem 9, we obtain

Corollary 1. *Let $S_1 = (U, C_1 \cup D_1)$ and $S_2 = (U, C_2 \cup D_2)$ be two decision tables. If $U/(C_1 \cup D_1) \prec U/(C_2 \cup D_2)$, then $\gamma(S_1) < \gamma(S_2)$.*

Theorem 10. *Let $S_1 = (U, C_1 \cup D_1)$ and $S_2 = (U, C_2 \cup D_2)$ be two conversely consistent decision tables. If $U/C_1 = U/C_2$ and $U/D_1 \prec U/D_2$, then $\gamma(S_1) < \gamma(S_2)$.*

Proof. From $U/C_1 = U/C_2$ and the converse consistencies of S_1 and S_2, it follows that there exist $X_l \in U/C_1$ and $Y_{j_0} \in U/D_2$ such that $Y_{j_0} \subseteq X_l$. By $U/D_1 \prec U/D_2$, we derive that there exist $Y_{j_0}^1, Y_{j_0}^2, \cdots, Y_{j_0}^s \in U/D_1$ $(s > 1)$ such that $Y_{j_0} = \bigcup_{k=1}^{s} Y_{j_0}^k$ and $|Y_{j_0}| = \sum_{k=1}^{s}|Y_{j_0}^k|$. It is clear that $|Z_{lj_0}| = \sum_{k=1}^{s}|Z_{lj_0}^k|$. Hence,

$$\gamma(S_2) = \sum_{i=1}^{m}\sum_{j=1}^{n} s^2(Z_{ij}) = \sum_{i=1}^{l-1}\sum_{j=1}^{n} s^2(Z_{ij}) + \sum_{j=1}^{n} s^2(Z_{lj}) + \sum_{i=l+1}^{m}\sum_{j=1}^{n} s^2(Z_{ij})$$

$$= \sum_{i=1}^{l-1}\sum_{j=1}^{n} s^2(Z_{ij}) + \sum_{j=1, j\neq j_0}^{n} s^2(Z_{lj}) + s^2(Z_{lj_0}) + \sum_{i=l+1}^{m}\sum_{j=1}^{n} s^2(Z_{ij})$$

$$= \sum_{i=1}^{l-1}\sum_{j=1}^{n} s^2(Z_{ij}) + \sum_{j=1, j\neq j_0}^{n} s^2(Z_{lj}) + \frac{|X_l \cap Y_{j_0}|^2}{|U|^2} + \sum_{i=l+1}^{m}\sum_{j=1}^{n} s^2(Z_{ij})$$

$$= \sum_{i=1}^{l-1}\sum_{j=1}^{n} s^2(Z_{ij}) + \sum_{j=1, j\neq j_0}^{n} s^2(Z_{lj}) + \frac{|X_l \cap \left(\bigcup_{k=1}^{s} Y_{j_0}^k\right)|^2}{|U|^2} + \sum_{i=l+1}^{m}\sum_{j=1}^{n} s^2(Z_{ij})$$

$$> \sum_{i=1}^{l-1}\sum_{j=1}^{n} s^2(Z_{ij}) + \sum_{j=1, j\neq j_0}^{n} s^2(Z_{lj}) + \frac{\sum_{k=1}^{s}|X_l \cap Y_{j_0}^k|^2}{|U|^2} + \sum_{i=l+1}^{m}\sum_{j=1}^{n} s^2(Z_{ij})$$

$$= \sum_{i=1}^{l-1}\sum_{j=1}^{n} s^2(Z_{ij}) + \sum_{j=1, j\neq j_0}^{n} s^2(Z_{lj}) + \sum_{k=1}^{s} s^2(Z_{lj_0}^k) + \sum_{i=l+1}^{m}\sum_{j=1}^{n} s^2(Z_{ij}) = \gamma(S_1),$$

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

Theorem 10 states that the support measure γ of a decision table decreases with its decision classes becoming finer.

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

Proof. The proof is similar to that of Theorem 10.

Theorem 11 states that the support measure γ of a decision table decreases as the condition classes becomes finer. As a result of Theorem 11, we obtain the following two corollaries. square

Corollary 2. *Let $S_1 = (U, C_1 \cup D_1)$ and $S_2 = (U, C_2 \cup D_2)$ be two consistent decision tables. If $U/C_1 = U/C_2$, then $\gamma(S_1) = \gamma(S_2)$.*

Proof. Suppose $U/(C \cup D) = \{X_i \cap Y_j | X_i \cap Y_j \neq \emptyset, X_i \in U/C_1, Y_j \in U/D\}$. Since both S_1 and S_2 are consistent, we have that $U/C_1 \preceq U/D_1$ and $U/C_2 \preceq U/D_2$, i.e., $U/(C_1 \cup D_1) = U/C_1$ and $U/(C_2 \cup D_2) = U/C_2$. It follows from $U/C_1 = U/C_2$ and $s(Z_{ij}) = \frac{|X_i \cap Y_j|}{|U|}$ that $\gamma(S_1) = \gamma(S_2)$. This completes the proof. □

Corollary 3. *Let $S_1 = (U, C_1 \cup D_1)$ and $S_2 = (U, C_2 \cup D_2)$ be two conversely consistent decision tables. If $U/D_1 = U/D_2$, then $\gamma(S_1) = \gamma(S_2)$.*

Proof. The proof is similar to that of Corollary 2.

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 we may extract more decision

Table 9
γ with different numbers of features in the data set Tie-tac-toe

Measure	Features								
	1	2	3	4	5	6	7	8	9
γ	0.1998	0.0695	0.0304	0.0120	0.0060	0.0030	0.0016	0.0010	0.0010

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

Measure	Features										
	3	6	9	12	15	18	21	24	27	30	33
r	0.0593	0.0080	0.0052	0.0038	0.0032	0.0029	0.0028	0.0028	0.0028	0.0027	0.0027

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

Measure	Features							
	1	2	3	4	5	6	7	8
γ	0.11418	0.02861	0.00721	0.00185	0.00064	0.00033	0.00011	0.00007

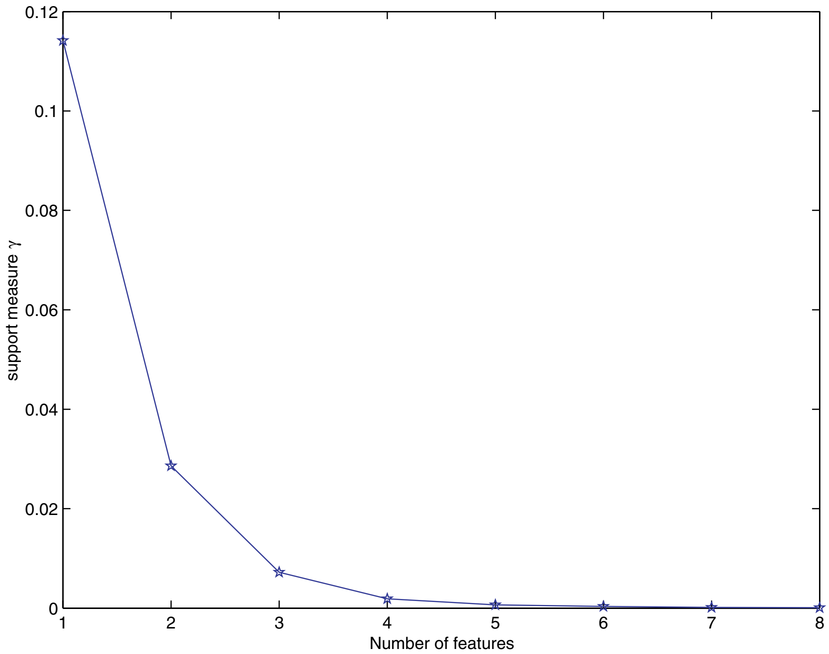

Fig. 9. Variation of the support measure γ with the number of features (data set Nursery).

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.

From the above experimental results and analyses, the proposed measures α, β and γ appear to be well suited for evaluating the decision performance of a decision table and a decision-rule set. These measures will be helpful for selecting a preferred rule-extracting method for a particular application.

6. Conclusions

In rough set theory, some classical measures for evaluating a decision rule or a decision table, such as the certainty measure and support measure of a rule and the approximation accuracy and consistency degree of a decision table have been suggested. However, these existing measures are not effective for evaluating the decision performance of a decision-rule set. In this paper, the limitations of these classical measures have been exemplified. To overcome these limitations, decision tables have been classified into three types according to their consistencies and three new more effective measures (α, β and γ) have been introduced for evaluating the certainty, consistency and support of a decision-rule set extracted from a decision table, respectively. It has been analyzed how each of these three new measures depends on the condition granulation and decision granulation of each of the three types of decision tables. The experimental analyses on three practical decision tables show that these three new measures are adequate for evaluating the decision performance of a decision-rule set extracted from a decision table in rough set theory.

Acknowledgements

The authors wish to thank the anonymous reviewers for their constructive comments. Also, we would like to thank Professor Witold Pedrycz for useful comments on this study.

This work was supported by the high technology research and development program (No. 2007 AA01Z165), the national natural science foundation of China (No. 70471003, 60573074, and 60275019), the foundation of doctoral program research of ministry of education of China (No. 20050108004), the top scholar foundation of Shanxi, China, Key Project of Science and Technology Research of the Ministry of Education of China (No. 206017) and the graduate student innovation foundation of Shanxi.

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] M. Beynon, Reducts within the variable precision rough sets model: a further investigation, European Journal of Operational Research 134 (3) (2001) 592–605.

[3] I. Düntsch, G. Gediaga, Uncertainty measures of rough set prediction, Artificial Intelligence 106 (1) (1998) 109–137.

[4] 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.

[5] V.N. Huynh, Y. Nakamori, A roughness measure for fuzzy sets, Information Sciences 173 (2005) 255–275.

[6] 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.

[7] M. Kryszkiewicz, Comparative study of alternative type of knowledge reduction in inconsistent systems, International Journal of Intelligent Systems 16 (2001) 105–120.

[8] 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.

[9] 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.

[10] J.Y. Liang, D.Y. Li, Uncertainty and Knowledge Acquisition in Information Systems, Science Press, Beijing, 2005.

[11] J.Y. Liang, Y.H. Qian, C.Y. Chu, D.Y. Li, J.H. Wang, Rough set approximation based on dynamic granulation, Lecture Notes in Artificial Intelligence 3641 (2005) 701–708.

[12] J.Y. Liang, Z.Z. Shi, D.Y. Li, 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.

[13] E. Marczewski, A general scheme of independence in mathematics, Bulletin de l Academie Polonaise des Sciences-Serie des Sciences Mathematiques Astronomiques et Physiques 6 (1958) 331–362.

[14] 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.

[15] HS. Nguyen, D. Slezak, Approximate reducts and association rules correspondence and complexity results, in: N. Zhong, A. Skowron, S. Oshuga (Eds.), Proceeding of RSFDGrc'99, Yamaguchi, Japan, LNAI, vol. 1711, 1999, pp 137–145.

[16] S.K. Pal, W. Pedrycz, A. Skowron, R. Swiniarski, Presenting the special isssue on rough-neuro computing, Neurocomputing 36 (2001) 1–3.

[17] Z. Pawlak, Rough sets, International Journal of Computer and Information Science 11 (1982) 341–356.

[18] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, Kluwer Academic Publisher, London, 1991.

[19] Z. Pawlak, Rough set theory and its applications in data analysis, Cybernetics and Systems 29 (1998) 661–688.

[20] Z. Pawlak, Some remarks on conflict analysis, Information Sciences 166 (2005) 649–654.

[21] Z. Pawlak, A. Skowron, Rudiments of rough sets, Information Sciences 177 (2007) 3–27.

[22] Z. Pawlak, A. Skowron, Rough sets: some extensions, Information Sciences 177 (2007) 28–40.

[23] Z. Pawlak, A. Skowron, Rough sets and boolean reasoning, Information Sciences 177 (2007) 41–73.

[24] W. Pedrycz, G. Vukovich, Granular worlds: representation and communication problems, International Journal of Intelligent Systems 15 (2000) 1015–1026.

[25] W. Pedrycz, Relational and directional in the construction of information granules, IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 32 (5) (2002) 605–614.

[26] Y.H. Qian, J.Y. Liang, Combination entropy and combination granulation in incomplete information system, Lecture Notes in Artificial Intelliegence 4062 (2006) 184–190.

[27] M. Quafatou, α – RST: a generalization of rough set theory, Information Sciences 124 (2000) 301–316.

[28] A. Skowron, Extracting laws from decision tables: a rough set approach, Computational Intelligence 11 (1995) 371–388.

[29] A. Skowron, C. Rauszer, The discernibility matrices and 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.

[30] D. Slezak, Approximate reducts in decision tables, in: Proceeding of IPMU'96, vol. 3, Granada, Spain, 1996, pp. 1159–1164.

[31] D. Slezak, Searching for dynamic reducts in inconsistent decision tables, in: Proceeding of IPMU'98, vol. 2, France, 1998, pp. 1362–1369.
[32] 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.
[33] W.X. Zhang, W.Z. Wu, J.Y. Liang, D.Y. Li, Theory and Method of Rough Sets, Science Press, Beijing, 2001.
[34] L. Zhang, B. Zhang, Fuzzy reasoning model under quotient space structure, Information Sciences 173 (2005) 353–364.
[35] P. Zheng, R. Germano, J. Ariën, K.Y. Qin, Y. Xu, Interpreting and extracting fuzzy decision rules from fuzzy information systems and their inference, Information Sciences 176 (2006) 1869–1897.
[36] W. Zhu, F.Y. Wang, Reduction and axiomization of covering generalized rough sets, Information Sciences 152 (2003) 217–230.
[37] W. Ziarko, Variable precision rough set model, Journal of Computer and System Science 46 (1993) 39–59.
[38] The UCI machine learning repository, <http://mlearn.ics.uci.edu/MLRepository.html>.