Approximation quality for sorting rules [1]

Günther Gediga [2]

Institut für Evaluation und Marktanalysen,
Brinkstr. 19,
49143 Jeggen, Germany

Ivo Düntsch [2]

Department of Computer Science,
Brock University,
St Catherines, Ontario, L2S 3AI, Canada ??

Abstract

A method is presented for the approximation quality for sorting rules, which is based on Rough Set Data Analysis. Several types of rules are discussed: The "nominal → nominal" case (NN) (the classical Rough Set approach), "nominal → ordinal" (NO) rules, and "ordinal → ordinal" (OO) rules, and its generalisation to "(nominal, ordinal) → ordinal" rules (NO-O). We provide a significance test for the overall approximation quality, and a test for partial influence of attributes based on the bootstrap technology.

For the bivariate case, the relationship of $\mathcal{U}$ – statistics and the proposed approximation quality measure is discussed. It can be shown that in this case a simple linear transformation of the Kendall tau correlation forms an upper bound for the approximation quality of sorting rules.

A competing model is also studied. Whereas this method is a promising tool in case of searching for global consistency, we demonstrate that in case of local perturbations in the data set the method may offer questionable results, and that it is dissociated from the theory it claims to support.

In the final Section, an example illustrates the introduced concepts and procedures.

Key words: Decision support systems, sorting rules, approximation quality, ordinal prediction

Table 1
A simple bivariate order information table

U	x_1	x_2	x_3	x_4	x_5	x_6
q	2	1	4	3	6	5
d	1	2	3	4	5	6

1 Introduction

In multi-criteria sorting problems one is often faced with statements of the form

If someone is male and at least 30 years of age,
then he will spend at least £10 a month on magazines.

Even though real life cannot be totally explained by such a rules, it is nevertheless worthwhile to approximate the prediction quality of the set of condition criteria taking into account all rules of the above form, and aggregate the values into a single measurement. As an example, consider the information system given in Table 1. There, U is a set of objects, q is a condition criterion, and d a decision criterion. With respect to the orderings $\leq$, one can observe, among others, the following rules:

$$(\forall x) 3 \leq f_q(x) \Rightarrow 3 \leq f_d(x), \tag{1}$$
$$(\forall x) 5 \leq f_q(x) \Rightarrow 5 \leq f_d(x). \tag{2}$$

In these implications, $f_q(x)$, resp. $f_d(x)$ is the value of object x under attribute q, resp. d. We refer to the rules (1), (2) as *ordinal-ordinal* (OO)-rules (Düntsch and Gediga, 1997a), because each side of the rule addresses an order relation.

The question arises, how one can measure the overall prediction success based on the instances of the observable rules in such a way that each rule contributes to the measure. This problem is, of course, not new, and solutions have been offered e.g. by Greco et al. (1998a) within the framework of rough sets. However, it will be shown in Section 4 that the approximation quality given there does not fulfil all that is claimed, and that a more refined measurement is needed.

* Corresponding author
Email addresses: gediga@eval-institut.de (Günther Gediga), duentsch@cosc.brocku.ca (Ivo Düntsch).
[1] Co-operation was supported by EU COST Action 274 "Theory and Applications of Relational Structures as Knowledge Instruments" (TARSKI).
[2] Equal authorship is implied.

We shall investigate not only pure sortings, but “nominal” attributes are also considered – an approach which is sometimes called multi-criteria and multi-attribute classification. Including nominal attributes we may have rules such as

$$(\forall x)\, f_{q_1}(x) = M \Rightarrow f_d(x) = 1000, \tag{3}$$
$$(\forall x)\, f_{q_1}(x) = M \Rightarrow f_d(x) \geq 1000, \tag{4}$$
$$(\forall x)\, f_{q_1}(x) = F \wedge f_{q_2}(x) \geq 3 \Rightarrow f_d(x) \geq 500, \tag{5}$$

Rule (3) is a *nominal - nominal* (NN) - rule, (4) is a *nominal-ordinal* (NO) - rule, and (5) is a mixed case (NO-O) - rule.

In the present paper we are not concerned with construction of optimal rules, but with the evaluation of deterministic rules which can be extracted from the data. For the construction of (in some sense) optimal rules we refer the reader to works on Boolean reasoning, e.g. Skowron and Polkowski (1997) or Wegener (1987).

It turns out that solving this restricted problem is quite intricate, and that there is a need for rather technical considerations. For example, the formalism of pre-image relations of Järvinen (1998) is used to describe ordinal prediction on a set of observations. Because of the necessary complexity of the presentation we supply instructive examples in every Section, and a comprehensive example after the introduction of the concepts.

Every rule based method faces the problem that the discovered rules may be due to chance; therefore, a machinery must be provided which tests the significance. In Section 8 we suggest ways how this can be done.

All computations were performed with the program *NOO* (Gediga and Düntsch, 2000).

2 Definitions and notation

Suppose that S is a binary relation on the set U. The *converse of* S is defined as

$$S^{\smile} = \{\langle y,x\rangle : xSy\}.$$

For each $x \in U$, we let $S(x) = \{t \in U : xSt\}$ be the *range of* x *in* S. The universal relation $U \times U$ on U is denoted by $\mathbf{1}_U$ or just $\mathbf{1}$, if no confusion can arise. We will denote the identity relation by $1'_U$ or just $1'$, and the diversity relation by $0'_U$ or $0'$, so that

$$1' = \{\langle x,x\rangle : x \in U\},$$
$$0' = \{\langle x,y\rangle : x,y \in U, x \neq y\}.$$

Suppose that $f : U \to V$ is a mapping, and that T is a binary relation on V. We define the *pre-image relation of T on U with respect to f* by

$$x\,\mathrm{Pre}(T, f)y \iff f(x)T f(y).$$

The following properties of pre-image relations have been shown by Järvinen (1998):

Lemma 2.1 *Suppose that P is one on the properties "reflexive", "symmetric", "transitive". If T has property P, then so has* $\mathrm{Pre}(T, f)$.

In particular, the pre-image of an equivalence relation is an equivalence relation. Furthermore, the pre-image of a partial order is a dominance, i.e. reflexive and transitive; the pre-image R of a linear order is a complete dominance, so that $R \cup R^{\smile} = \mathbf{1}_U$.

Of particular interest is the relation $\mathrm{Pre}(=, f)$, which we will also denote by θ_f. It is an equivalence relation on U, sometimes called the *kernel of f*, and

$$x\theta_f y \iff f(x) = f(y).$$

Suppose that $\{\langle V_q, S_q\rangle : q \in \Omega\}$ is a family of relational structures, where each S_q is a binary relation on V_q. The *product relation S* on $V = \prod_{q\in\Omega} V_q$ is defined by

$$\langle x_q\rangle_{q\in\Omega} S \langle y_q\rangle_{q\in\Omega} \iff (\forall q \in \Omega) x_q S_q y_q.$$

For our knowledge representation, we have a finite set U of objects, a set of independent attributes (or variables) $\Omega = \{q_0, \ldots, q_m\}$ with value sets V_q, $q \in \Omega$, and a decision attribute d with value set V_d. To avoid trivialities, we suppose throughout that V_d has at least two elements. For each $Q \subseteq \Omega$ we set $V_Q = \prod_{q\in Q} V_q$. A *decision table I* is a subset of $U \times V_\Omega \times V_d$. Each $\langle x, s_0, \ldots, s_m, t\rangle \in I$ is interpreted as

"Object x has q_i-value s_i for each $i \le m$, and d-value t".

We let $f_{q_i} : U \to V_{q_i}$ and $f_d : U \to V_d$ be the projections to the attributes; to avoid awkward notation, we assume throughout that all these projections are onto functions. The kernels of these functions are denoted by θ_{q_i} and θ_d, respectively. If $Q \subseteq \Omega$, we let $f_Q : U \to \prod V_Q$ be the product mapping of the functions f_q, $q \in Q$.

The following result will be useful later on:

Lemma 2.2 *Suppose that for each $q \in \Omega$, S_q is a binary relation on V_q, $P \subseteq Q \subseteq \Omega$, and S_P, resp. S_Q are the product relations on V_P, resp. V_Q. Then,* $\mathrm{Pre}(S_Q, f_Q) \subseteq \mathrm{Pre}(S_P, f_P)$.

PROOF. Let $x\,\mathrm{Pre}(S_Q, f_Q)y$. Then, $f_q(x)S_q f_q(y)$ for all $q \in Q$. Since $P \subseteq Q$, we also have $f_q(x)S_q f_q(y)$ for all $q \in P$, and thus, $x\,\mathrm{Pre}(S_P, f_P)y$. □

3 Sorting rules

Suppose that on each V_{q_i} we have a binary relation S_{q_i}, and T is a binary relation on V_d. An $(S_0,\ldots,S_m|T)$*-rule* is a statement of the form

$$(\forall x)\left[\bigwedge_{i\le m} s_i S_{q_i} f_{q_i}(x) \Rightarrow tT f_d(x)\right] \tag{6}$$

The standard situation of multi-criteria investigation has each S_i as a linear order on V_{q_i}, and T a linear order on V_d. Since the left side of (6) is a conjunction, we are in fact looking at the product order S on $\prod_{q\in Q} V_q$ defined by

$$\langle x_0,\ldots,x_m\rangle S\langle y_0,\ldots y_m\rangle \iff x_i S_{q_i} y_i \text{ for all } i\le m.$$

Since the product relation of partial orders is again a partial order, it suffices to consider $Q=\{q\}$, and a partial order relation S on V_Q. However, we want to point out, that, in doing this, the researcher has to make a definite choice for each V_{q_i} which of $\le$ or $\ge$ is of interest in conjunction with the choices for the other attributes q_j, and disregard all other possibilities. Below, for example, rules of the form $s_1 \le_{q_1} f_{q_1}(x) \wedge s_2 \le_{q_2} f_{q_2}(x) \Rightarrow t \le_d f_d(x)$ are compared with rules of the type $s_1 \ge_{q_1} f_{q_1}(x) \wedge s_2 \le_{q_2} f_{q_2}(x) \Rightarrow t \le_d f_d(x)$, and we develop indices that show whether the $(\le_{q_1},\le_{q_2})$ rules or the $(\ge_{q_1},\le_{q_2})$ rules have more predicting power. But there are other types of rules, which may have even higher prediction success, such as rules of the type $(\ge_{q_1},\ge_{q_2})$. Simple combinatorics show that in general we have to consider $2^{|Q|}$ different types of rules, and investigating all possible types even with a moderate number of attributes is a demanding task. Therefore, in the sequel we will assume that the researcher is aware that the orientation of the attributes within Q is essential for rule generation and evaluation, and has made a choice which ones are to be investigated.

Since we also want to consider the nominal case where S_q is the identity relation, and the mixed cases of nominal and ordinal criteria, we will look at the following situations:

(1) $Q=\{q\}$ and S is the identity on V_Q.
(2) $Q=\{q\}$ and S is a partial order on V_Q, written as $\le_Q$.

We will also consider a special case of 2., namely,

(3) $Q=\{q_0,q_1\}$, S_0 is a partial order on V_{q_0}, and S_1 is the identity on V_{q_1}.

For the right hand sides of (6), we will consider the identity, a linear ordering T, written as $\le_d$, and its converse $\ge_d$. We will drop subscripts on the orderings when no confusion can arise.

The rules which we want to investigate are deterministic, and have the form

$$(\forall x)[sS\,f_Q(x) \Rightarrow tT\,f_d(x)], \tag{7}$$

where $S \in \{=_{V_Q}, \leq_Q\}$ and $T \in \{=_{V_d}, \leq_d\}$. More concretely, we look at the rules

$(\forall x)[s = f_Q(x) \Rightarrow t = f_d(x)]$	nominal-nominal (NN),
$(\forall x)[s = f_Q(x) \Rightarrow t \leq f_d(x)]$	nominal-ordinal (NO),
$(\forall x)[s \leq f_Q(x) \Rightarrow t \leq f_d(x)]$	ordinal-ordinal (OO).

In the sequel, by a rule we will always mean a deterministic rule such as (7), unless stated otherwise.

In order to gauge the predictive power of S with respect to T we consider the following sets of objects:

$$\varphi(s) = \varphi(s, \mathbf{1}) = \{x \in U : sS\,f_Q(x)\} = f_Q^{-1}(S(s)),$$

containing those objects which satisfy the conditions of (7),

$$\varphi(t) = \varphi(\mathbf{1}, t) = \{x \in U : tT\,f_d(x)\} = f_d^{-1}(T(t)),$$

containing those objects for which the consequent of (7) holds,

$$\varphi(s,t) = \{x \in U : sS\,f_Q(x) \wedge tT\,f_d(x)\} = f_Q^{-1}(S(s)) \cap f_d^{-1}(T(t)),$$

which denotes the set of objects which support (7). If it is necessary to make it clear which relations S or T are under consideration, we will write $\varphi^S(s)$, $\varphi^T(t)$ and $\varphi^{S|T}(s,t)$. We shall also use the value

$$\pi(s|t) = \frac{|\varphi(s,t)|}{|\varphi(t)|},$$

called the *covering degree* (of s with respect to t). It is the relative number of objects related to t which are explained by the range of s in S, and it estimates how well a rule covers the relational (e.g. sorting) properties of T for a fixed t.

We will frequently use *intersection tables* which are matrices indexed by $V_Q \times V_d$, whose entries are the sets $\varphi(s,t)$. These are related to the well known contingency tables which measure the joint occurrence of two or more phenomena: If we replace $\varphi(s,t)$ by its cardinality, we obtain a contingency table in the classical sense.

The set of those $s \in V_Q$ leading to a deterministic $(S|T)$-rule of the form (7) for fixed t is defined as

$$\begin{aligned}\det{}^{S|T}(t) &= \{s \in V_Q : (f_Q^{-1}(S(s)) \neq \emptyset) \wedge (\forall x)[sS\,f_Q(x) \Rightarrow tT\,f_d(x)\} \\ &= \{s \in V_Q : \emptyset \neq f_Q^{-1}(S(s)) \subseteq f_d^{-1}(T(t))\}.\end{aligned}$$

If S and T are understood, we will omit the superscript.

Since we always choose $S \in \{=_Q, \leq_Q\}$, it is transitive, and we can use the following result:

Lemma 3.1 *If S is transitive, then*

$$\bigcup_{s \in \det(t)} f_Q^{-1}(S(s)) = \bigcup_{s \in \det(t)} f_Q^{-1}(s).$$

PROOF. It is enough to show that $s \in \det(t)$ and sSs' imply $s' \in \det(t)$. Thus, suppose that the hypotheses hold, and let $s' S f_Q(x)$. Then, $sSs'S f_Q(x)$, and the transitivity of S implies $sS f_Q(x)$. It follows from $s \in \det(t)$ that $tT f_d(x)$, and therefore, $s' \in \det(t)$. $\square$

It should also be noted that for all s with $f_Q^{-1}(S(s)) \neq \emptyset$,

$$s \in \det(t) \iff f_Q^{-1}(S(s)) \subseteq f_d^{-1}(T(t)) \iff \varphi(s,t) = \varphi(s). \tag{8}$$

We say that $t \in V_d$ *splits a class K of* θ_Q, if there are $x, y \in K$ such that

$$tT f_d(x) \text{ and not } tT f_d(y). \tag{9}$$

The tasks which we want to consider are as follows:

(1) For each $t \in V_d$ find a scoring function $\gamma(S|t)$ which aggregates the covering degrees $\pi(s|t)$, $s \in V_Q$. This index should offer an evaluation how well the elements of S cover the relational properties of T for a fixed value $t \in V_d$.
(2) Find a scoring function $\gamma(S|T)$ which aggregates the scoring functions $\gamma(S|t)$ over all $t \in V_d$, and which enables an evaluation how well the relation S explains the relational properties of T.

Since we only want to count the $x \in U$ which contribute to a rule of the form (6), those $s \in V_Q$ for which $f_Q^{-1}(S(s)) \not\subseteq f_d^{-1}(T(t))$ are not of interest to us. Therefore, the simplest way of defining $\gamma(S|t)$ is to take the cardinality of the union of the sets $\varphi(s,t)$, $s \in \det(t)$ relative to the cardinality of $\varphi(t)$:

$$\gamma(S|t) = \frac{|\bigcup_{s \in \det(t)} \varphi(s,t)|}{|\varphi(t)|}. \tag{10}$$

Notwithstanding Occam's razor, the question arises, whether the simplest way is necessarily the best. We shall see below that in the NN and NO cases, the numerator of (10) is the cardinality of the sum of classes of θ_Q which agrees with the approximation quality used in rough set theory (Pawlak, 1991). In the OO or NO-O case the intersections of the sets $\{\varphi(s,t) : s \in \det(t)\}$ are not necessarily empty, which means that the sum of the cardinalities of $\varphi(s,t)$ is not a useful statistic. Because $\bigcup_{s\in\det(t)} \varphi(s,t) \subseteq \varphi(t)$, the index $\gamma(S|t)$ is bounded by the interval $[0,1]$.

In a second step, we aim to find suitable indices $w(t)$ such that

$$\gamma(S|T) = \sum_t w(t) \cdot \gamma(S,t).$$

The simplest way – a linear function – of combining the rule based measurements $\gamma(S|t)$ to the aggregated measurement $\gamma(S|T)$ is chosen once again. We shall see that this aggregation scheme is quite natural in the NN and NO case, and that it is applicable in the OO or NO-O case as well. In the OO case we will show that $\gamma(S|T)$ is connected to other well known indices of ordinal data analysis.

The reasoning behind our choice of the parameters is that we want to construct aggregation indices which measure the quality of sorting in terms of valid deterministic rules. There is, of course, an infinite number of possible schemes; on the basis of Occam's razor, we choose simple ones which seem to do what we want them to do. This is similar to the fact that the classical γ is only one of infinitely many measures which could be used, depending on the circumstances (Gediga and Düntsch, 2001).

4 A proposal for multi-criteria sorting quality

In order to make clearer where the problems lie, we recall the γ measure suggested by Greco et al. (1998a) for the approximation quality of rules of the form (6). Let us consider a partial order S on V_Q and a linear order T on V_d. We will sometimes write $\leq_Q$ for S and $\leq_d$ for T.

For each $t \in V_d$ let

$$\begin{aligned} C_t &= f_d^{-1}(t), \\ C_t^{\geq} &= \bigcup\{\mathrm{Pre}(T, f_d)\breve{}(x) : x \in C_t\} = \{x : t \geq f_d(x)\}, \\ C_t^{\leq} &= \bigcup\{\mathrm{Pre}(T, f_d)(x) : x \in C_t\} = \{x : t \leq_d f_d(x)\}, \end{aligned}$$

and for each $x \in U$

$$\begin{aligned} D^+(x) &= \mathrm{Pre}(S, f_Q)\breve{}(x) = \{y \in U : f_Q(y) S f_Q(x)\}, \\ D^-(x) &= \mathrm{Pre}(S, f_Q)(x) = \{y \in U : f_Q(x) S f_Q(y)\}. \end{aligned}$$

For the definitions above, note that the sets $C_t^{\geq}$ and $C_t^{\leq}$ have already appeared above as $\varphi^{\geq}(t)$ and $\varphi^{\leq}(t)$; the rules we consider are based on the relation $\leq_Q$ on U, while the dominance relation D used by Greco et al. (1998a) is in fact the converse of the pre-image of $\leq_Q$. In order to keep consistent with their scenario we have therefore "turned around" the direction of $C_t^{\geq}$ and $C_t^{\leq}$.

We see that

$$D^{+}(x) \subseteq C_t^{\geq} \iff (\forall y)[f_Q(y) \leq_Q f_Q(x) \Rightarrow f_d(y) \leq_d t],$$

and

$$D^{-}(x) \subseteq C_t^{\leq} \iff (\forall y)[f_Q(x) \leq_Q f_Q(y) \Rightarrow t \leq_d f_d(y)].$$

For each $t \in V_d$ the *lower approximation of* $C_t^{\geq}$*, resp.* $C_t^{\leq}$ is now defined as

$$\begin{aligned} \underline{C_t^{\geq}} &= \{x \in U : D^{+}(x) \subseteq C_t^{\geq}\}, \\ \underline{C_t^{\leq}} &= \{x \in U : D^{-}(x) \subseteq C_t^{\leq}\}. \end{aligned}$$

Suppose that $f_Q(x) = s$. If $x \in \underline{C_t^{\geq}}$, we obtain the rule

$$(\forall y)[f_Q(y) \leq_Q s \Rightarrow f_d(y) \leq_d t].$$

Similarly, if $x \in \underline{C_t^{\leq}}$, we have

$$(\forall y)[s \leq_Q f_Q(y) \Rightarrow t \leq_d f_d(y)].$$

The *upper approximations* are defined as

$$\overline{C_t^{\geq}} = \bigcup_{x \in C_t^{\geq}} D_Q^{+}(x), \ \overline{C_t^{\leq}} = \bigcup_{x \in C_t^{\leq}} D_Q^{-}(x).$$

The *boundaries* are the sets

$$\partial(C_t^{\geq}) = \overline{C_t^{\geq}} \setminus \underline{C_t^{\geq}}, \ \partial(C_t^{\leq}) = \overline{C_t^{\leq}} \setminus \underline{C_t^{\leq}}.$$

We set

$$\partial(U,q,d) = \bigcup_{t \in T} \partial(C_t^{\geq}) \cup \bigcup_{t \in T} \partial(C_t^{\leq}).$$

Proposition 4.1

$$\begin{aligned} \partial(U,q,d) = \{x \in U : (\exists y \in U)[(f_Q(x) \leq_Q f_Q(y) \text{ and } f_d(y) \not\leq_d f_d(x))] \\ \text{or } (\exists z \in U)[(f_Q(z) \leq_Q f_Q(x) \text{ and } f_d(x) \not\leq_d f_d(z))]\} \end{aligned} \tag{11}$$

Table 2
Some sorting values for Table 1

$D^+(x_1) = \{x_1, x_2\}$	$D^+(x_3) = \{x_1, \dots, x_4\}$	$D^+(x_5) = U$
$D^+(x_2) = \{x_2\}$	$D^+(x_4) = \{x_1, x_2, x_4\}$	$D^+(x_6) = \{x_1, \dots, x_4, x_6\}$
$C_k^{\geq} = \{x_i : i \leq k\}, \quad C_k^{\leq} = \{x_i : k \leq i\}$		
$\underline{C_1^{\geq}} = \emptyset$	$\overline{C_1^{\geq}} = \{x_1, x_2\}$	$\partial(C_1^{\geq}) = \{x_1, x_2\}$
$\underline{C_3^{\geq}} = \{x_1, x_2\}$	$\overline{C_3^{\geq}} = \{x_1, \dots, x_4\}$	$\partial(C_3^{\geq}) = \{x_3, x_4\}$
$\underline{C_5^{\geq}} = \{x_1, \dots, x_4\}$	$\overline{C_5^{\geq}} = U$	$\partial(C_5^{\geq}) = \{x_5, x_6\}$

PROOF. First, note that for each t,

$$x \in \partial(C_t^{\geq}) \text{ and } \iff (\exists y, z \in U)[f_Q(z) \leq_Q f_Q(x) \leq_Q f_Q(y) \text{ and } f_d(y) \leq_d t \lneq f_d(z)], \quad (12)$$

and

$$x \in \partial(C_t^{\leq}) \iff (\exists y, z \in U)[f_Q(z) \leq_Q f_Q(x) \leq_Q f_Q(y) \text{ and } f_d(y) \lneq_d t \leq f_d(z)]. \quad (13)$$

"$\subseteq$": Let $x \in \partial(U, q, d)$, and suppose that $x \in \partial(C_t^{\geq})$. By (12), there are y, z such that $f_Q(z) \leq_Q f_Q(x) \leq f_Q(y)$ and $f_d(y) \leq_d t \lneq f_d(z)$. If $f_d(x) \leq_d t$, then $f_Q(z) \leq_Q f_Q(x)$ and $f_d(x) \leq t \lneq f_d(z)$ witness that x is in the right hand side of (11). If $t \lneq_d f_d(x)$, then $f_Q(x) \leq f_Q(y)$ and $f_d(y) \leq_d t \lneq f_d(x)$ show that x x is in the right hand side of (11). The case $x \in \partial(C_t^{\leq})$ is analogous.

"$\supseteq$": Suppose that $f_Q(x) \leq_Q f_Q(y)$ and $f_d(y) \lneq_d f_d(x)$. Set $t = f_d(y)$, $z = x$, and use (12). If $f_Q(z) \leq_Q f_Q(x)$ and $f_d(x) \lneq_d f_d(z)$, set $t = f_d(z)$, $y = x$, and use (13). $\square$

Furthermore, for $x \in \partial(U, q, d)$ and $f_Q(x) = s$ we have $f_Q^{-1}(s) \subseteq \partial(U, q, d)$.

The *quality of approximation of the partition* $\{C_t\}$ *by means of* q or the *quality of sorting* is now defined by

$$\gamma_{GMS} = \frac{|U \setminus \partial(U, q, d)|}{|U|}$$

The values for the data in Table 1 are shown in Table 2.

It is claimed by Greco et al. (1998a) that

> "On the basis of the approximations obtained by means of the dominance relations it is possible to induce a generalised description of the preferential information contained in the decision table, in terms of decision rules".

It is, however, not made clear, how these rules, which depend on the constants from the sets V_Q and V_d, are related to the approximations described above and the quality of sorting. Indeed, there does not seem to be a necessary connection, and the existence of non-trivial (deterministic) rules and γ_{GMS} seem dissociated. As an example, consider the information system given in Table 1. The existing rules (1) and (2) are certainly informative. For example, (1) can be interpreted as "If $f_Q(x)$ is not less than the midpoint of the V_Q-scale, then $f_d(x)$ is not less than the midpoint of the V_d-scale", which is quite helpful, if e.g. T is a ranking how well a firm pays back a credit, and S is a ranking of creditability computed by some simple information about the firms.

It can be seen from Table 2 that the γ_{GMS} value of the system is zero, despite the fact that there are non-trivial rules. This seems to contradict the quotation above, and shows the dissociation of γ_{GMS} from the theory it claims to reflect. Indeed, the example can be generalised to any set with cardinality $2 \cdot n$, and therefore, one can have arbitrarily many non-trivial rules, but $\gamma_{GMS} = 0$. The reason for this seems to be the overly restricted definition of boundary. One consequence of this is the sensitivity of the measurement against local perturbations: In the example of Table 1, the assignments f_Q and f_d are the same up to the transposition of adjacent elements.

Classical rough set approximation quality γ_C counts the relative number of elements which can be captured by deterministic rules, while γ_{GMS} does not. Hence, another approach is required if one wants to capture the existence of (deterministic) rules which can be discovered in the system.

We would like to stress that it is not the purpose of the present paper to suggest that the γ_{GMS} index has no value *per se*, and we would like to quote Banzhaf, who had similar concerns in the context of weighted voting:

> "As with any mathematical analysis, its intent is only to explain the effects which necessarily follow once the mathematical model and the rules of its operation are established ... whatever else may be said for or against weighted voting as a practical solution to a practical problem, it does not even theoretically produce the effects which have been claimed to justify it." (Banzhaf, 1965, p. 319)

5 The NN case

Let us start with the case that S and T are the identity relations. Consider as an example the information system shown in Table 3. The intersection table and the con-

Table 3
Information system II

U	A	B	C	D	E	F	G
q	1	1	2	2	3	3	4
d	1	1	1	2	2	3	3

Table 4
Intersection table $\varphi(s,t)$

	d = t		
q = s	1	2	3
1	$\{\mathbf{A},\mathbf{B}\}$	$\emptyset$	$\emptyset$
2	$\{C\}$	$\{D\}$	$\emptyset$
3	$\emptyset$	$\{E\}$	$\{F\}$
4	$\emptyset$	$\emptyset$	$\{\mathbf{G}\}$

Table 5
Conditional coverage $\pi(s|t)$

	d = t		
q = s	1	2	3
1	**0.67**	0.00	0.00
2	0.33	0.50	0.00
3	0.00	0.50	0.50
4	0.00	0.00	**0.50**

ditional coverage are given in Tables 4 and 5, respectively. There, as in the sequel, a bold faced entry corresponds to one of the three equivalent conditions (8). Because of the special structure of the identity relation, we have $f_d^{-1}(T(t)) = f_d^{-1}(t)$. Independence gives us $\det(t_1) \cap \det(t_2) = \emptyset$ for $t_1 \neq t_2$, and, using the principle of indifference, we let the weights $w(t)$ be proportional to the size of the class $f_d^{-1}(t)$, i.e.

$$w(t) = \frac{|f_d^{-1}(t)|}{|U|} = \frac{|\varphi(t)|}{|U|}.$$

Since $\varphi(t) = f_d^{-1}(t)$ and $\varphi(s,t) = f_Q^{-1}(s)$ for $s \in \det(t)$, we have

$$\gamma_{NN}(=|=) = \sum_{t \in V_d} \left(\frac{|f_d^{-1}(t)|}{|U|} \cdot \sum_{s \in \det(t)} \frac{|f_Q^{-1}(s)|}{|f_d^{-1}(t)|} \right) = \sum_{t \in V_d} \sum_{s \in \det(t)} \frac{|f_Q^{-1}(s)|}{|U|},$$

and thus, $\gamma_{NN}(=|=)$ agrees with the standard approximation quality of rough set theory (Pawlak, 1991).

For the example of Table 5 we have

$$\gamma_{NN}(=|=) = \frac{3}{7} \cdot 0.67 + \frac{2}{7} \cdot 0 + \frac{2}{7} \cdot 0.5 = 0.429.$$

6 The NO case

Our second case deals with the situation that S is the identity, but T is a linear order $\leq$ on V_d; thus, we face a nominal-ordinal (NO) situation. If we take into account the converse $\geq$ of T, there are several rules of type (6) which can occur:

$$(\forall x)[f_Q(x) = s \Rightarrow t \leq f_d(x)], \tag{14}$$
$$(\forall x)[f_Q(x) = s \Rightarrow t \geq f_d(x)], \tag{15}$$
$$(\forall x)[f_Q(x) = s \Rightarrow t \leq f_d(x)] \text{ or } (\forall x)[f_Q(x) = s \Rightarrow t \geq f_d(x)]. \tag{16}$$

Let us consider (14). Similar to the NN case, we arrive at

$$\gamma_{NO}(S|t) = \sum_{s \in \det(t)} \pi(s|t) = \frac{1}{|f_d^{-1}(T(t))|} \cdot \sum_{s \in \det(t)} |f_Q^{-1}(s)|.$$

Observe that

$$\gamma_{NO}(S|t) = 1 \Longleftrightarrow \sum_{s \in \det(t)} |f_Q^{-1}(s)| = |f_d^{-1}(T(t))| \Longleftrightarrow \bigcup_{s \in \det(t)} f_Q^{-1}(s) = f_d^{-1}(T(t)).$$

In other words, $\gamma(S|t) = 1$ if and only if t does not split any class of θ_Q in the sense of (9).

The determination of the weightings $w(t)$ for the aggregation of $\gamma_{NO}(S|t)$ needs a little more attention. Let $t_{\min} = \min V_d$; then, for each $s \in V_Q$ we obviously have the rule

$$(\forall x)[f_Q(x) = s \Rightarrow t_{\min} \leq f_d(x)],$$

which does not carry useful information. Therefore, we will set $w_{t_{\min}} = 0$.

In all other cases $t \gneqq t_{\min}$, we choose the simplest way to compute $w(t)$ by taking the cardinality of $\varphi(t)$ relative to the sum of all $|\varphi(t)|$, $t \neq t_{\min}$, to obtain

$$w(t) = \begin{cases} 0, & \text{if } t = t_{\min}, \\ \frac{|\varphi(t)|}{\sum_{t \gneqq t_{\min}} |\varphi(t)|} & \text{otherwise.} \end{cases}$$

Other weighting schemes are of course possible. The chosen one has an analogous structure to the NN approach: The numerator is the number of all elements which fulfil a “local” condition, whereas the denominator is simply the sum of all possible numerators. Thus, we arrive at

$$\begin{aligned}
\gamma_{NO}(= \mid \leq) &= \sum_{t \in V_d} w(t) \cdot \gamma_{NO}(S|t) \\
&= \sum_{t \gtrsim t_{\min}} \frac{|f_d^{-1}(T(t))|}{\sum_{t \gtrsim t_{\min}} |f_d^{-1}(T(t))|} \cdot \frac{1}{|f_d^{-1}(T(t))|} \cdot \sum_{s \in \det(t)} |f_Q^{-1}(s)| \\
&= \frac{1}{\sum_{t \gtrsim t_{\min}} |f_d^{-1}(T(t))|} \cdot \sum_{t \gtrsim t_{\min}} \sum_{s \in \det(t)} |f_Q^{-1}(s)|.
\end{aligned}$$

It is easy to see that

$$0 \leq \gamma_{NO}(= \mid \leq) \leq 1,$$

because the functions $\gamma_{NO}(= |t)$ satisfy this condition, and the weightings $w(t)$ are non-negative and sum up to 1. Furthermore,

Proposition 6.1 *If $P \subseteq Q$ then $\gamma_{NO}(=_{V_P} \mid \leq) \leq \gamma_{NO}(=_{V_Q} \mid \leq)$.*

PROOF. By the definition of γ_{NO} it is enough to show that

$$\sum_{s \in \det^{S_P|T}(t)} |f_P^{-1}(s)| \leq \sum_{s \in \det^{S_Q|T}(t)} |f_Q^{-1}(s)|$$

for all $t \in T$. Suppose that $s \in \det^{S_P|T}(t)$; if $f_P^{-1}(s) = \emptyset$, then s does not contribute to the left hand sum. Otherwise, let $f_P(y) = s$, and $f_Q(y) \leq_Q f_Q(x)$. Since $P \subseteq Q$, we have $s = f_P(y) \leq_P f_P(x)$ by Lemma 2.2, and therefore, $f_Q(y) \in \det^{S_Q|T}(t)$. This implies that $y \in f_Q^{-1}(s')$ for some $s' \in \det^{S_Q|T}(t)$. Since the sets $f_Q^{-1}(s')$, $s' \in \det^{S_Q|T}(t)$ are pairwise disjoint, the conclusion follows. □

A characterisation of the extremal situations for $\gamma_{NO}(= \mid \leq)$ is given in the following

Proposition 6.2 *(1) $\gamma_{NO}(= \mid \leq) = 0 \iff$ for each class K of θ_Q there is some $x \in K$ such that $f_d(x) = t_{\min}$.*

(2) $\gamma_{NO}(= \mid \leq) = 1 \iff$ for each class K of θ_Q there is some $t \in V_d$ such that $f_d(x) = t$ for all $x \in K$.

PROOF. 1. Since $0 \leq \gamma_{NO}(= |t)$ for each $t \in V_d$, and $\gamma_{NO}(= \mid \leq)$ is the positively weighted sum of $\gamma_{NO}(=|t)$ for $t \gneq t_{\min}$, we have

$$\gamma_{NO}(=|\leq) = 0 \Longleftrightarrow (\forall t \gneq t_{\min})\gamma_{NO}(=|t) = 0, \tag{17}$$
$$\Longleftrightarrow (\forall t \gneq t_{\min})\,\det(t) = \emptyset, \tag{18}$$
$$\Longleftrightarrow (\forall t \gneq t_{\min})(\forall s \in V_Q)\, f_Q^{-1}(s) \not\subseteq f_d^{-1}(t), \tag{19}$$
$$\Longleftrightarrow (\forall t \gneq t_{\min})(\forall s \in V_Q)(\exists x \in U)[f_Q(x) = s \text{ and } f_d(x) \lneq t]. \tag{20}$$

Let t^+ be the unique element of V_d covering $t_{\min}$, and suppose that $\gamma_{NO}(=|\leq) = 0$. By (20), each class of θ_Q contains some x such that $t_{\min} = f_d(x) \lneq t^+$. Conversely, if each class of θ_Q contains some x such that $t_{\min} = f_d(x)$, then clearly (20) is satisfied.

2. By definition,

$$\gamma_{NO}(=|\leq) = 1 \Longleftrightarrow \sum_{t \gneq t_{\min}} \sum_{s \in \det(t)} |f_Q^{-1}(s)| = \sum_{t \gneq t_{\min}} |f_d^{-1}(T(t))|.$$

Since $f_Q^{-1}(s) \subseteq f_d^{-1}(T(t))$ for each $s \in \det(t)$, and the sets $f_Q^{-1}(s)$ are pairwise disjoint, we have $\sum_{s \in \det(t)} |f_Q^{-1}(s)| \leq |f_d^{-1}(T(t))|$ for each $t \in V_d$. Therefore,

$$\gamma_{NO}(=|\leq) = 1 \Longleftrightarrow (\forall t \gneq t_{\min}) \sum_{s \in \det(t)} |f_Q^{-1}(s)| = |f_d^{-1}(T(t))| \tag{21}$$
$$\Longleftrightarrow (\forall t \gneq t_{\min}) \bigcup_{s \in \det(t)} f_Q^{-1}(s) = f_d^{-1}(T(t)) \tag{22}$$
$$\Longleftrightarrow (\forall t \gneq t_{\min})(\forall x \in U)(\forall y \in \theta_Q(x))[t \leq f_d(x) \Rightarrow t \leq f_d(y)] \tag{23}$$

Suppose that $\gamma_{NO}(=|\leq) = 1$, and assume there is some class K of θ_Q and $x, y \in K$ such that $f_d(x) \neq f_d(y)$. Let w.l.o.g. $f_d(y) \lneq f_d(x)$, and set $t = f_d(x)$. Then, $f_d(y) \lneq t \leq f_d(x)$ and $f_Q(x) = f_Q(y)$, contradicting (23).

Conversely, suppose that for each class K of θ_Q there is some $t \in V_d$ such that $f_d(x) = t$ for all $x \in K$. This is the case exactly when $\theta_Q \subseteq \theta_d$, and therefore, each $f_d^{-1}(t)$ is the union of all sets $f_Q^{-1}(s)$ it contains. This implies that $\sum_{s \in \det(t)} |f_Q^{-1}(s)| = \sum_{t \gneq t_{\min}} |f_d^{-1}(T(t))|$, and therefore, $\gamma_{NO}(=|\leq) = 1$ by (21) □

Corollary 6.3 $\gamma_{NO}(=|\leq) = 1 \Longleftrightarrow \gamma_{NN}(=|=) = 1 \Longleftrightarrow \gamma_{NO}(=|\geq) = 1.$

This tells us, along with (6.1), that if θ_Q forms a finer (or equal) partition than θ_d this result in a perfect approximation of all three relations $=_d$, $\leq_d$ and $\geq_d$ by the classes of Q. This may be astonishing at first glance; however, if a class K of θ_Q is split by some t, none of the three relations can be approximated without error. Consequently, if the granularity of Q is high – i.e. the classes of θ_Q are relatively small –, we can expect a high approximation quality in γ_{NN} as well as in γ_{NO} – this

Table 6
Intersection table $\varphi^{(=|\leq)}(s,t)$ for Table 3

	d		
q	1	2	3
1	$\{\mathbf{A},\mathbf{B}\}$	$\emptyset$	$\emptyset$
2	$\{\mathbf{C},\mathbf{D}\}$	$\{D\}$	$\emptyset$
3	$\{\mathbf{E},\mathbf{F}\}$	$\{\mathbf{E},\mathbf{F}\}$	$\{F\}$
4	$\{\mathbf{G}\}$	$\{\mathbf{G}\}$	$\{\mathbf{G}\}$

Table 7
Intersection table $\varphi^{(=|\geq)}(s,t)$ for Table 3

	d		
q	1	2	3
1	$\{\mathbf{A},\mathbf{B}\}$	$\{\mathbf{A},\mathbf{B}\}$	$\{\mathbf{A},\mathbf{B}\}$
2	$\{C\}$	$\{\mathbf{C},\mathbf{D}\}$	$\{\mathbf{C},\mathbf{D}\}$
3	$\emptyset$	$\{E\}$	$\{\mathbf{E},\mathbf{F}\}$
4	$\emptyset$	$\emptyset$	$\{\mathbf{G}\}$

is the problem of "casual rules", which means that every rule is generated by only a few number of examples; we will address this issue in Section 8.

If $\gamma_{NO}(=|\leq)=0$ and $\gamma_{NO}(=|\geq)=0$, then each class of θ_Q must contain $t_{\min}$ and $t_{\max}$ (which are different by our global assumptions). Therefore,

Corollary 6.4 *$\gamma_{NO}(=|\leq)=0$ and $\gamma_{NO}(=|\geq)=0$ imply $\gamma_{NN}(=|=)=0$.*

For the example of Table 3 we obtain the intersection tables shown in Table 6, 7. As above, the cells $\langle s,t\rangle$ with $s\in\det(t)$ are printed in boldface. Therefore, for the example,

$$\gamma_{NO}(=|\leq)=\frac{1}{6}\cdot[(2+1)+1]=\frac{2}{3},$$
$$\gamma_{NO}(=|\geq)=\frac{1}{8}\cdot[(2+2)+2]=\frac{3}{4}.$$

The definition of γ_{NO} was based on rules of type (14), where we supposed that $T=\leq$. An analogous construction can be made for $T=\geq$. By similar arguments, we define the approximation quality $\gamma_{NO}(S|\leq,\,\geq)$ for rules of type (16) by

$$\gamma_{NO}(=|\leq,\,\geq)=\frac{\sum_{t\gneq t_{\min}}|\bigcup_{s\in\det(t)}\varphi^{=|\leq}(s,t)|+\sum_{t\lneq t_{\max}}|\bigcup_{s\in\det(t)}\varphi^{=|\geq}(s,t)|}{\sum_{t\gneq t_{\min}}|\varphi^{\leq}(t)|+\sum_{t\lneq t_{\max}}|\varphi^{\geq}(t)|} \tag{24}$$
$$=\frac{\sum_{t\gneq t_{\min}}|\bigcup_{s\in\det^{\leq}(t)}f_Q^{-1}(s)|+\sum_{t\lneq t_{\max}}|\bigcup_{s\in\det^{\geq}(t)}f_Q^{-1}(s)|}{\sum_{t\gneq t_{\min}}|f_d^{-1}(T(t))|+\sum_{t\lneq t_{\max}}|f_d^{-1}(T^{\smile}(t))|} \tag{25}$$

The statistic $\gamma_{NO}(=|\leq,\,\geq)$ forms an average of the statistics $\gamma_{NO}(=|\leq)$ and $\gamma_{NO}(=|\geq)$. Note that $\gamma_{NO}(=|\leq)$ does not take into account $\det(t_{\min})$, and $\gamma_{NO}(=|\geq)$ does not use $\det(t_{\max})$, because it is a tautology that any element in the sample is not smaller (greater) than the smallest (greatest) element. Counting such rules results in counting trivialities, hence, such γ^*_{NO} would be positively biased and would result in a value which is never zero. For example, if one uses the $\det(t_{\min})$ as well, one will result in $\gamma^*_{NO}=\gamma_{NO}+CONST$ for any condition attribute. Thus, not using $\det(t_{\min})$ and $\det(t_{\max})$ is just a matter of convenience for the usage of γ_{NO}. Obviously, there is no tautological rule in the NN case, therefore no value should be omitted. In all other categories the counters of the examples for the nominator and for the denominator are identical for both statistics. Therefore, the three statistics $\gamma_{NO}(= \mid \leq, \geq), \gamma_{NO}(= \mid \leq)$ and $\gamma_{NO}(= \mid \geq)$ might differ in those cases in which the cardinalities of the $t_{\min}$ and/or the $t_{\max}$ categories are substantial. In this case, the value of $\gamma_{NO}(= \mid \leq, \geq)$ is a compromise of the two elementary statistics, assuming that the researcher is interested in $\leq$-rules as well as $\geq$-rules.

The evaluation of the example shows

$$\gamma_{NO}(= \mid \leq, \geq) = \frac{(2+1+1)+(2+2+2)}{(4+2)+(3+5)} = \frac{5}{7}.$$

Finally we observe:

Corollary 6.5 *(1)* $\gamma_{NO}(= \mid \leq, \geq) = 1 \Longleftrightarrow \gamma_{NN}(= \mid =) = 1.$
(2) $\gamma_{NO}(= \mid \leq, \geq) = 0 \Rightarrow \gamma_{NN}(= \mid =) = 0.$
(3) If $P \subseteq Q$ *then* $\gamma_{NO}(=_Q \mid \leq_Q, \geq_Q) \leq \gamma_{NO}(=_P \mid \leq_P, \geq_P).$

7 The OO case and mixed NO-O predictions

The most interesting case for multi-criteria decision making considers rules of the form

$$(\forall x)[s \leq_Q f_Q(x) \Rightarrow t \leq_d f_d(x)].$$

The arguments for the choice of the parameters $w(t)$ given in the NO case apply here as well, so that in the NO case we find formally the same expression as in the NN case, but using different sets $\det^{S|T}$:

$$\gamma_{OO}(\leq \mid \leq) = \frac{\sum_{t \gneq t_{\min}} |\bigcup_{s \in \det(t)} f_Q^{-1}(s)|}{\sum_{t \gneq t_{\min}} |f_d^{-1}(T(t))|}. \tag{26}$$

The OO-intersection tables for the system of Table 3 are given in Table 8, 9, and we compute

$$\gamma_{OO}(\leq \mid \leq) = \frac{3+1}{4+2} = \frac{2}{3},$$

and, as another example of the same evaluations strategy,

$$\gamma_{OO}(\leq \mid \geq) = \frac{0}{5+3} = 0.$$

The behaviour of γ_{OO} is similar to γ_{NO} as Proposition 7.1 shows.

Table 8
Intersection table $\varphi^{(\leq|\leq)}(s,t)$ for Table 3

	d		
q	1	2	3
1	$\mathbf{U}$	$\{D,E,F,G\}$	$\{F,G\}$
2	$\{\mathbf{C},\mathbf{D},\mathbf{E},\mathbf{F},\mathbf{G}\}$	$\{D,E,F,G\}$	$\{F,G\}$
3	$\{\mathbf{E},\mathbf{F},\mathbf{G}\}$	$\{\mathbf{E},\mathbf{F},\mathbf{G}\}$	$\{F,G\}$
4	$\{\mathbf{G}\}$	$\{\mathbf{G}\}$	$\{\mathbf{G}\}$

Table 9
Intersection table $\varphi^{(\leq|\geq)}(s,t)$ for Table 3

	d		
q	1	2	3
1	$\{A,B,C\}$	$\{A,B,C,D,E\}$	$\mathbf{U}$
2	$\{C\}$	$\{C,D,E,\}$	$\{\mathbf{C},\mathbf{D},\mathbf{E},\mathbf{F},\mathbf{G}\}$
3	$\emptyset$	$\{E\}$	$\{\mathbf{E},\mathbf{F},\mathbf{G}\}$
4	$\emptyset$	$\emptyset$	$\{\mathbf{G}\}$

Proposition 7.1 *(1) If $P \subseteq Q$ then $\gamma_{OO}(\leq_P \mid \leq) \leq \gamma_{OO}(\leq_Q \mid \leq)$.*
(2) Let M be the set of maximal elements of $\langle V_Q, \leq \rangle$. Then,

$$\gamma_{OO}(\leq \mid \leq) = 0 \Longleftrightarrow (\forall s \in M)(\exists x \in U)[f_Q(x) = s \text{ and } f_d(x) = t_{\min}].$$

(3) $\gamma_{OO}(\leq \mid \leq) = 1 \Longleftrightarrow (\forall x,y)[f_Q(x) \leq f_Q(y) \Rightarrow f_d(x) \leq f_d(y)]$.

PROOF. 1. is analogous to Proposition 6.1.

2. is analogous to Proposition 6.2.1.

3. Suppose that $\gamma_{OO}(\leq \mid \leq) = 1$. This holds, if and only if

$$\sum_{t \gneq t_{\min}} \Big| \bigcup_{s \in \det(t)} f_Q^{-1}(s) \Big| = \sum_{t \gneq t_{\min}} |f_d^{-1}(T(t))|. \tag{27}$$

Since for each t we have $\bigcup_{s \in \det(t)} f_Q^{-1}(s) \subseteq f_d^{-1}(T(t))$, (27) holds if and only if

$$(\forall t \gneq t_{\min})[\bigcup_{s \in \det(t)} f_Q^{-1}(s) = f_d^{-1}(T(t))]. \tag{28}$$

Observing that $s \in \det(t)$ and $s \leq s'$ imply $s' \in \det(t)$, we obtain for $t \gneq t_{\min}$

$$t = f_d(x) \Rightarrow [(\exists s \in \det(t))[s \leq f_Q(x) \Rightarrow f_Q(x) \subseteq \det(t)]],$$

i.e. $f_Q(x) \in \det(f_d(x))$. Observe that this holds for all $t \in V_d$, so that

$$(\forall x)(\forall y)[f_Q(x) \leq f_Q(y) \Rightarrow f_d(x) \leq f_d(y)], \tag{29}$$

which is our desired result.

Now, suppose that (29) holds, and let $t \gneq t_{\min}$ and $t \leq f_d(x)$. We need to show that $x \in \bigcup_{s \in \det(t)} f_Q^{-1}(s)$. Assume that $f_Q(x) \notin \det(t)$. Then, there is some $y \in U$ such that $f_Q(x) \leq f_Q(y)$ and $t \not\leq f_d(y)$. Since $t \leq f_d(x)$, this contradicts (29). □

The following is immediate from the preceding result:

Corollary 7.2 *(1) If $\langle V_Q, \leq_Q\rangle$ is order isomorphic to $\langle V_d, \geq_d\rangle$, then $\gamma_{OO}(\leq,\leq) =$ 0.*

(2) Suppose that $Q = \{q,p\}$, $|U| = |V_q| = |V_p|$, and $\leq_p = \geq_q$. Then $\gamma_{OO}(\leq_{\{q,p\}} \mid \leq_d) = 1$ holds for any decision attribute d.

At first glance, this results seems to be a drawback of our approach, because $\{q,p\}$ can predict every ordered attribute d without any error. Nevertheless, there are some arguments that this result is a very natural one, and that it must be observed, if the method is a valid regression type procedure. When we look at what happens in Corollary 7.2, we see that every class of θ_Q consists of one element, and therefore, no two classes of θ_Q are comparable. If the classes are not comparable, the OO-approach collapses to the NO-approach. This is another instance of the casuality problem mentioned above.

In order to see that a $\gamma_{OO}(\leq \mid \leq, \geq)$ approximation in analogy to (24) is not feasible, let us look again at the exemplary case that V_Q and V_d are linearly ordered by S, respectively, by T, and that $|U| = |V_Q| = |V_d| = N$. In this case, f_Q and f_d are bijective, and without loss of generality we suppose that

$$U = V_Q = V_d,\ \mathrm{Pre}(S, f_Q) = S,\ \mathrm{Pre}(T, f_d) = T,\ f_Q = f_d = 1'_U. \tag{30}$$

Then,

$$M \stackrel{\text{def}}{=} \frac{N \cdot (N-1)}{2} = |S \cap 0'_U| = |S\breve{} \cap 0'_U| = |T \cap 0'_U| = |T\breve{} \cap 0'_U|.$$

Statistics offers a framework for measuring monotone relationships, the *$\mathcal{U}$-statistic* (Puri and Sen, 1971), to which our measures can be related. Set $C \stackrel{\text{def}}{=} |S \cap T \cap 0'_U|$, $D \stackrel{\text{def}}{=} |S \cap T\breve{} \cap 0'_U|$; we define the following two $\mathcal{U}$-statistics

$$\mathcal{U}(\leq \mid \leq) = \frac{C}{M}, \quad \mathcal{U}(\leq \mid \geq) = \frac{D}{M}.$$

$M = C + D$ implies that $\mathcal{U}(\leq \mid \geq) = \frac{M-C}{M}$, and therefore, the measure

$$\mathcal{U}(\leq \mid \leq,\geq) = \frac{D+C}{2M} = \frac{M}{2M} = \frac{1}{2},$$

which is analogous to (24), is constant, irrespective of the possible connections between the orders on V_Q and V_d.

The $\mathcal{U}$-statistics are related to the well known *Kendall's τ statistic* (Kendall, 1970) via

$$\tau = \mathcal{U}(\leq \mid \leq) - \mathcal{U}(\geq \mid \leq). \tag{31}$$

The value of τ is bounded by $-1 \leq \tau \leq 1$. A value of τ near $+1$ can only appear if most of the pairs of elements are collected in the $\mathcal{U}(\leq \mid \leq)$ statistic, which is interpreted as a high positive monotone correlation of V_Q with V_d. A value near -1 indicates that V_Q has a high negative monotone correlation with V_d. A value of τ near 0 is interpreted as no monotone correlation of V_Q and V_d.

The next proposition connects τ and $\gamma_{OO}(\leq \mid \leq)$ in the 2-attribute-case for which τ was designed:

Proposition 7.3 *Let U, S, T, C, D, M be as above. Then, $\gamma_{OO}(\leq \mid \leq) \leq \frac{1+\tau}{2}$.*

PROOF. We assume the simplifications given in (30). Then,

$$\begin{aligned}
\gamma_{OO}(\leq \mid \leq) &= \frac{\sum_{t \gneq t_{\min}} |\bigcup_{s \in \det(t)} f_Q^{-1}(s)|}{\sum_{t \gneq t_{\min}} |f_d^{-1}(T(t))|}, \\
&= \frac{\sum_{t \gneq t_{\min}} |\det(t)|}{\sum_{t \gneq t_{\min}} |T(t)|}, \\
&= \frac{\sum_{t \gneq t_{\min}} |\det(t)|}{\frac{(N-1)\cdot N}{2}} \\
&= \frac{\sum_{t \gneq t_{\min}} |\det(t)|}{M}.
\end{aligned}$$

Since $D = M - C$, we find

$$\tau = \frac{2C}{M} - 1$$

and thus, it suffices to show $\gamma \leq \frac{C}{M} = \mathcal{U}(\leq \mid \leq)$, which simplifies to

$$\sum_{t \gneq t_{\min}} |\det(t)| \leq |S \cap T \cap 0'_U|.$$

By our hypotheses, we can take w.l.o.g.

$$U = V_Q = V_d = \{1, 2, \ldots, N\}.$$

Let $\leq_d$ be the natural order $\leq$ on U, and $s_r \leq_Q s_k \iff r \leq k$. For $k \in U$ let $s_{t_k} = k$. The following facts are easily verified:

$$t_1 \leq t_2 \Rightarrow \det(t_2) \subseteq \det(t_1). \tag{32}$$

$$\det(t) = \emptyset \iff s_N \lneq t. \tag{33}$$

$$s \in \det(t) \Rightarrow 1 \gneq_Q s. \tag{34}$$

To each pair $\langle t, s \rangle$ where $1 \lneq t$ and $s \in \det(t)$ we want to assign an element of $S \cap T \cap 0'$, such that the assignment is injective. Thus, suppose that s and t fulfil these conditions. We then have

$$1 \lneq t \leq s.$$

There are two cases:

(1) $t \lneq_Q s$: Since $s \in \det(t)$ we have $t \leq_d s$, and thus we can assign $\langle t, s\rangle \mapsto \langle t, s\rangle$.
(2) $s \leq_Q t$: In this case, we assign $\langle t, s\rangle \mapsto \langle 1, s\rangle$. By (34) above, we have $1 \leq_Q s$, and it follows that $\langle 1, s\rangle \in S \cap T \cap 0'$. To show that the assignment is one-one, suppose that $s \leq_Q t_1$, $s \leq_Q t_2$, and $s \in \det t_1 \cap \det t_2$. Then, $s \in \det(t_1)$ and $s \leq_Q t_2$ implies $t_1 \leq t_2$, and similarly, $t_2 \leq t_1$. Therefore, $s \leq_Q t$ and $s \in \det(t)$ is possible for at most one $1 \lneq t$, and therefore, the assignment is one-one.

This completes the proof. □

It is possible that $\gamma_{OO} \lneq \frac{1+\tau}{2}$:

V_Q	4 2 3 1
V_d	1 2 3 4

In this situation, $\gamma_{OO}(\leq \mid \leq) = 0$ and $-1 \lneq \tau$, so that $0 \lneq \frac{1+\tau}{2}$.

We have shown that

$$\gamma_{OO}(\leq \mid \leq) \leq \mathcal{U}(\leq \mid \leq)$$

holds in the 2-attribute situation, and by the same argument, one can show that the other γ statistics are bounded by comparable $\mathcal{U}$ statistics as well. The interpretation is straightforward: The number of admissible relation compatible pairs of elements is an upper bound for the number of possible relation compatible rules. This motivates the definition of a modified τ statistic, which reflects the rule compatibility by

$$\tau_{OO} = \gamma_{OO}(\leq \mid \leq) - \gamma_{OO}(\leq \mid \geq).$$

The construction of τ_{OO} is similar to τ in the 2-attribute situation, with the difference that τ_{OO} is based on γ statistics and τ is based on $\mathcal{U}$ statistics. As with the original τ, the value of τ_{OO} is close to $+1$, if $\leq_Q$ explains $\leq_d$ well, and -1 if $\leq_Q$ explains $\geq_d$.

If we have more than one element in Q, the problem of choice of relevant relations arises again, and we refer the reader to the remarks on page 5. If orientation of the orders S_q is fixed, two interesting statistics remain: $\gamma_{OO}(\leq \mid \leq)$ and $\gamma_{OO}(\geq \mid \geq)$. We propose the statistic

$$\gamma_{OO} = \frac{\sum_{t \gneq t_{\min}} |\bigcup_{s \in \det^{\leq|\leq}(t)} f_Q^{-1}(s)| + \sum_{t \lneq t_{\max}} |\bigcup_{s \in \det^{\geq|\geq}(t)} f_Q^{-1}(s)|}{\sum_{t \gneq t_{\min}} |f_d^{-1}(T(t))| + \sum_{t \lneq t_{\max}} |f_d^{-1}(T^{\smile}(t))|}$$

as a mean value between $\gamma_{OO}(\leq \mid \leq)$ and $\gamma_{OO}(\geq \mid \geq)$. This is a useful statistic, if

- The user is aware that the orientation within Q is considered for the calculation of γ_{OO} , and
- The user is willing to interpret $(\leq,\leq)$ rules as well as $(\geq,\geq)$ rules.

As in the case of *NO* prediction, the computation of the statistics $\gamma_{OO}(\leq \mid \leq)$ and $\gamma_{OO}(\geq \mid \geq)$ differ only in (non-) consideration of $t_{\min}$ and $t_{\max}$. This means that both statistics disagree if the categories defined by $t_{\min}$ and $t_{\max}$ are substantial.

Consider the following example:

d	5	5	5	5	5	5	4	3	2	1
q	6	6	6	6	6	1	2	3	4	5

We see that $\gamma_{OO}(\leq,\leq) = \frac{2}{3}$, but $\gamma_{OO}(\geq,\geq) = 0$. The statistic $\gamma_{OO} = 0.5$ is the compromise of both values. In this example the value of γ_{OO} has a smaller distance to $\gamma_{OO}(\leq,\leq)$ than to $\gamma_{OO}(\geq,\geq)$. This is due the fact that there are more non-tautological rules based on $\leq_d$ than on $\geq_d$ and therefore $\gamma_{OO}(\leq,\leq)$ receives more weight in the compromise measure γ_{OO}.

Proposition 7.1 implies the monotony result

Corollary 7.4 *(1) If $P \subseteq Q$ then $\gamma_{OO}(P \Rightarrow d) \leq \gamma_{OO}(Q \Rightarrow d)$.*
(2) Let M be the set of maximal elements of $\langle V_Q, \leq\rangle$ and M' be the set of minimal elements of $\langle V_Q, \leq\rangle$. Then,

$$\gamma_{OO} = 0 \Longleftrightarrow (\forall s \in M)(\exists x \in U)[f_Q(x) = s \text{ and } f_d(x) = t_{\min}] \\ \text{and } (\forall s \in M')(\exists x \in U)[f_Q(x) = s \text{ and } f_d(x) = t_{\max}].$$

(3) $\gamma_{OO} = 1 \Longleftrightarrow (\forall x,y)[f_Q(x) \leq f_Q(y) \Rightarrow f_d(x) \leq f_d(y)]$.

Suppose we want to use nominal and ordinal attributes jointly for prediction (NO-O), e.g. gender and salary. We can easily accommodate this situation into the OO scenario by observing that the product relation of an equivalence and a partial order again is a partial order [3]. As a consequence, given a nominally scaled attribute q, an additional criterion is only of interest within the equivalence classes of θ_q. This means for the gender/salary example that the salary orderings within “male” and “female” are used to predict d, but salary order between the “male” and the “female” groups are not taken into account.

In Table 10 a simple data set is given to demonstrate the effect of an NO-O prediction. If $Q = \{q_1\}$, the prediction results in the poor approximation quality $\gamma_{NO} =$

[3] This is not a new idea in this context. One will find the same arguments in Greco et al. (1998b).

Table 10
A simple (nominal-order)-order information table

U	a	b	c	d	e	f	g	h
q_1	M	F	M	F	M	F	M	F
q_2	5	1	6	2	7	3	8	4
d	1	2	3	4	5	6	7	8

0.143, because the only non trivial rules that can be extracted are

$$M = f_{q_1}(x) \Rightarrow 7 \geq f_d(x),$$
$$F = f_{q_1}(x) \Rightarrow 2 \leq f_d(x).$$

The approximation quality of the prediction $Q = \{q_2\}$ is also not very high ($\gamma_{OO} = 0.429$), because the ordering in V_{q_2} allows only a few predictions of the d ordering (e.g. $6 \leq f_{q_2}(x) \Rightarrow 4 \leq f_d(x)$). The combined NO-O prediction $\{q_1, q_2\} \Rightarrow d$ shows a perfect approximation quality, because the orderings of q_2 and d are 1:1 within the classes "M" and "F".

8 Statistical issues

As we have already mentioned above, a high γ value can possibly be achieved by random influence, and a low γ value need not indicate that the remaining rules do not have a substantial base within the data. This situation is identical to that in the NN prediction case, which we have discussed in Düntsch and Gediga (1997b). In case that the approximation quality will be used to recommend a decision, a test of the usefulness of the observed empirical γ values is therefore needed. One possibility are cross-validation methods; however, these use additional model assumptions which may not be fulfilled in the situation under review, and thus, procedures which do not require such assumptions are often more adequate. The method of random labelling, which we have used for NN prediction in Düntsch and Gediga (1997b) for testing the significance of the NN approximation quality can be used in the present situations as well. The idea is that the observed γ should be extreme in the distribution of all γ values which can be constructed, if we leave the V_Q values unchanged and randomise the V_d values by reordering the elements of U by a randomly chosen permutation σ. While the computation of the whole distribution of $\gamma(\sigma)$ is very costly, a simulation using about 1000 random permutations will suffice to approximate the distribution of γ assuming random labelling. If there are less than 5% values which are not lower than the empirical γ, we claim γ to be *significant*.

Even if γ is significant, we cannot conclude that the approximation quality is really substantial. A simple way to gauge the "real prediction effect" is the comparison with the expectation $\mathcal{E}_\sigma[\gamma(\sigma)]$ of γ given random labelling. If the empirical γ and the expectation of γ show a large difference, we can conclude that γ is substantial. Because the difference is restricted by the maximum of γ, the index

$$\kappa = \frac{\gamma - \mathcal{E}_\sigma[\gamma(\sigma)]}{1 - \mathcal{E}_\sigma[\gamma(\sigma)]}$$

can be used to compute the relative distance of γ to the expected value (Cohen, 1960). As a rule of thumb, $\kappa > 0.5$ indicates a high effect.

After having determined a significant and substantial rule, it may be of interest whether some of the attributes within Q are dispensable. A simple inspection of the difference

$$\Delta_i(\gamma(\leq_Q, \leq_d)) = \gamma(\leq_Q, \leq_d) - \gamma(\leq_{Q\setminus\{q_i\}}, \leq_d)$$

gives a first hint about the value of q_i as attribute within the set Q for predicting d. If $\Delta_i(\gamma)$ is close to zero, we may possibly leave out q_i without changing the approximation quality dramatically. The problem remains to gauge higher $\Delta_i(\gamma)$ values, because even high jumps do not necessarily indicate high additional prediction quality of an attribute. The interested reader is referred to the forthcoming Gediga and Düntsch (2002) for a discussion of the interpretation of differences of γ values.

A statistical treatment of the influence of an attribute starts with the idea that the standard deviation of $\Delta_i(\gamma|Q,d)$ should be much smaller then the difference itself, if we draw other samples of elements with the same dependency structure. The *Bootstrap* (Efron, 1982) can be employed to estimate $\hat{s}(\Delta_i(\gamma(\leq_Q, \leq_d))$ using simulation methods. The comparison

$$z(\Delta_i(\gamma(\leq_Q, \leq_d)) = \frac{\Delta_i(\gamma(\leq_Q, \leq_d))}{\hat{s}(\Delta_i(\gamma| \leq_{Q\setminus\{q_i\}}, \leq_d))}$$

can be used as a guide for the evaluation of the additional prediction quality of attribute q_i: If $z(\Delta_i(\gamma|Q,d)) \leq 2$, there is no good reason to assume that $\Delta_i(\gamma|Q,d)$ is substantial, because the statistical fluctuation within the sample is about as high as the difference itself. A value $z(\Delta_i(\gamma|Q,d)) \gnsim 2$ indicates that attribute q_i is not dispensable for the prediction of d.

There is a minor problem with the distribution of $\Delta_i(\gamma(\leq_Q, \leq_d))$: If the differences are very small, the distribution tends to be skewed, and the decision based on the $z \gnsim 2$ rule of thumb may be misleading. This problem can be solved by investing more simulations and estimating the probability $p[\Delta_i(\gamma|(\leq_Q, \leq_d)) \leq 0]$. If $\hat{p}[\Delta_i(\gamma|(\leq_Q, \leq_d)) \leq 0] \lnsim 5\%$, we conclude that the observed difference $\Delta_i(\gamma|(\leq_Q, \leq_d))$ is substantial. This approach needs at least 1000 bootstrap simulations in order to achieve an acceptable confidence for the value $p[\Delta_i(\gamma|(\leq_Q, \leq_d)) \leq 0]$. Estimating the z value needs at least 100 bootstrap simulations. We have tested numerous cases, and for every problem we have checked, both approaches led to same

Table 11
Contraception data

Country		q_1	q_2	q_3	q_4	d
(1)	Lesotho	1	0	0	0	6
(2)	Kenya	0	0	0	0	9
(3)	Peru	1	1	2	0	14
(4)	Sri Lanka	1	1	0	1	22
(5)	Indonesia	0	0	0	1	25
(6)	Thailand	1	0	0	1	36
(7)	Colombia	1	2	1	1	37
(8)	Malaysia	0	1	2	1	38
(9)	Guyana	2	1	2	0	42
(10)	Jamaica	2	0	2	2	44
(11)	Jordan	0	2	1	0	44
(12)	Panama	2	2	2	1	59
(13)	Costa Rica	2	1	2	2	59
(14)	Fiji	1	1	2	2	60
(15)	Korea	2	1	1	2	61

Table 12
Recoded data

Country		q_1	q_2	q_3	q_4	d
(1)	Lesotho	1	0	0	0	0
(2)	Kenya	0	0	0	0	0
(3)	Peru	1	1	2	0	0
(4)	Sri Lanka	1	1	0	1	0
(5)	Indonesia	0	0	0	1	0
(6)	Thailand	1	0	0	1	1
(7)	Colombia	1	2	1	1	1
(8)	Malaysia	0	1	2	1	1
(9)	Guyana	2	1	2	0	1
(10)	Jamaica	2	0	2	2	1
(11)	Jordan	0	2	1	0	1
(12)	Panama	2	2	2	1	2
(13)	Costa Rica	2	1	2	2	2
(14)	Fiji	1	1	2	2	2
(15)	Korea	2	1	1	2	2

conclusion.

9 Example

To demonstrate our procedures, we will use a modified [4] data set investigated by Cliff (1994) shown in Table 11, with the data recoded in Table 12. It is aimed to predict the values of the countries in the criterion

- % ever practising contraception (d)

from the following criteria, which are coded as ratings (0=low; 1=middle; 2=high) in our example:

- Average years of education (q_1),
- Percent urbanised (q_2),
- Gross national product per capita (q_3),
- Expenditures on family planning (q_4),

and to find those characteristics that are most valuable to predict d. The decision criterion is coded in two different ways; for the analysis we use the recoded data as well as the original one in order to compare the differences due to the recoding

[4] The original data set consist of measurements $q_1, \ldots, q_4$. Because γ_{NN} and γ_{NO} will offer only trivial results with the original data, we use ratings instead of measurements, which cluster the data.

Fig. 1. The $\leq_Q$ relation for Table 12

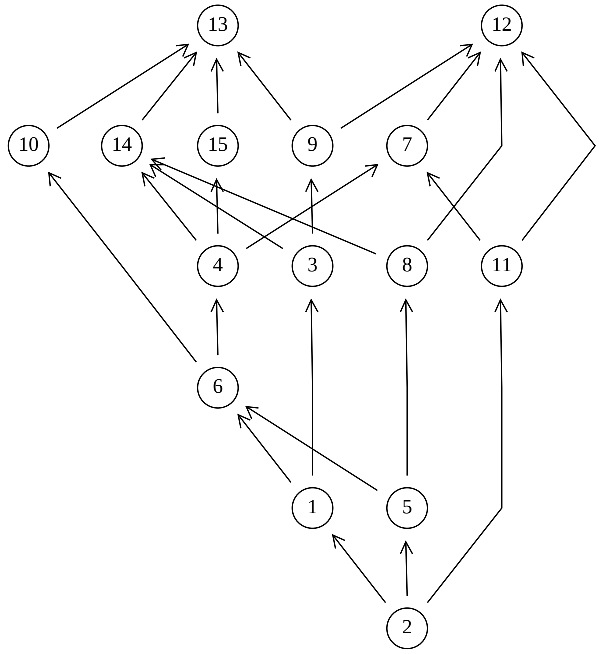

Table 13
Analysis of contraception data; * = significant at $\alpha = 0.05$; ** = significant at $\alpha = 0.01$; *** = significant at $\alpha = 0.001$

Attribute set	γ_{NN}		γ_{NO}		γ_{OO}		γ_{GMS}	
	d coded	*d* orig.	*d* coded	*d* orig.	d coded	*d* orig.	d coded	d orig.
$\{q_1,q_2,q_3,q_4\}$	1.000	1.000	1.000	1.000	0.900***	0.933***	0.333**	0.867***
$\{q_1,q_2,q_3\}$	0.600	0.467	0.789	0.733	0.639*	0.633*	0.000	0.400**
$\{q_1,q_2,q_4\}$	1.000	0.867	0.978	1.000	0.867***	0.900***	0.267**	0.800***
$\{q_1,q_3,q_4\}$	0.733	0.733	0.967	0.867	0.783***	0.700**	0.000	0.400*
$\{q_2,q_3,q_4\}$	0.733	0.467	0.900	0.867	0.817***	0.833***	0.133*	0.667***
$\{q_1,q_2\}$	0.467	0.333	0.694	0.633	0.594*	0.567**	0.000	0.333*
$\{q_1,q_3\}$	0.400	0.267	0.739*	0.633	0.544**	0.500*	0.000	0.133
$\{q_1,q_4\}$	0.333	0.200	0.756	0.667	0.561*	0.533**	0.000	0.067
$\{q_2,q_3\}$	0.400	0.267	0.606	0.533	0.422*	0.333	0.000	0.067
$\{q_2,q_4\}$	0.467	0.133	0.789*	0.733	0.750***	0.733**	0.067*	0.467***
$\{q_3,q_4\}$	0.333	0.200	0.822**	0.667	0.722***	0.600**	0.000	0.200*
$\{q_1\}$	0.000	0.000	0.344	0.300	0.322*	0.300*	0.000	0.000
$\{q_2\}$	0.000	0.000	0.294	0.267	0.261	0.267	0.000	0.000
$\{q_3\}$	0.000	0.000	0.411*	0.267	0.306*	0.167	0.000	0.000
$\{q_4\}$	0.000	0.000	0.450**	0.300	0.450**	0.300*	0.000	0.000

procedure. A graph of the relation $\leq_Q$, drawn by the RELVIEW program (Behnke et al., 1998), is shown in Figure 1.

Table 13 shows the result of analysing the contraception data with different predictions systems. The first column displays the predicting sets of attributes. The next two columns report the rough set approximation quality using the NN - prediction approach. Note, that no prediction is significant, which means that the approximation by those variables can be explained by random assignment as well. Furthermore, given coded attributes $q_1, \ldots, q_4$, the approximation of the original decision attribute is not as high as the approximation of the recoded decision attributes. This is not astonishing, because the coded decision attribute generates a partition which is coarser than the original partition, and any approximation of the latter is an approximation of the set in the partition based on the coded decision attribute.

The next two columns show the results of the NO - prediction of the decision attribute. The results are similar to the NN - prediction, but γ_{NO} is higher for smaller number of attributes. This is due the fact that the values of d need only to cut among the equivalence classes of Q, which is easier to achieve than approximation of a d class by Q classes. The first significant dependencies can be observed here: Using the coded data, there is a significant sorting of equivalence classes in attribute q_4, q_3 and $\{q_3, q_4\}$.

The next two columns show the results of the OO - prediction of the decision attribute. First of all $\gamma_{OO} \leq \gamma_{NO}$, which is of course given by construction, because γ_{OO} is based on the d-sortable equivalence classes in Q, which are additionally ordered by $\leq$ in the Q attributes. We observe that although the numerical values of γ_{OO} were smaller than those of γ_{NO}, they are often significant. This means that the observed d-sortable equivalence classes in Q, which are additionally ordered by $\leq$ in the Q attributes can only rarely be observed under random assignment of the d attribute values to the Q attribute values. The results show that assuming a model of stability of the order of the Q classes, the set $\{q_2, q_4\}$ is a promising attribute set, because it shows a high and significant OO-prediction.

Bootstrap analysis – based on 1000 simulations – of the partial increase of γ verified this claim as the following table shows. Note, that the rule "$z(boot) \gtrsim 2$" points to the same dispensable attributes as the more expensive rule "$p(\text{diff } \leq 0) \lesssim 5\%$".

```
CODED ...
gamma      = 0.8667
Analysis of partial gamma
Attr. 1: gammadiff=0.1167 s(boot)=0.0723 z(boot)=1.6146 p(diff<=0)=0.059
Attr. 2: gammadiff=0.3056 s(boot)=0.1288 z(boot)=2.3732 p(diff<=0)=0.008
Attr. 4: gammadiff=0.2722 s(boot)=0.0983 z(boot)=2.7686 p(diff<=0)=0.002

UNCODED ...
gamma      = 0.9000
Analysis of partial gamma
Attr. 1: gammadiff=0.1667 s(boot)=0.0882 z(boot)=1.8897 p(diff<=0)=0.148
Attr. 2: gammadiff=0.3667 s(boot)=0.1281 z(boot)=2.8614 p(diff<=0)=0.026
Attr. 4: gammadiff=0.3333 s(boot)=0.1347 z(boot)=2.4751 p(diff<=0)=0.027
```

The last two columns show the results using γ_{GMS} as the measure of approximation quality. We see that γ_{GMS} and γ_{OO} show about the same results, if we look at the significance of their observed values. But due to reasons discussed above, the values of γ_{GMS} are low or even extremely low, if we use uncoded data.

10 Discussion

We have encountered three different types of ordinal dependency measurements: The classical $\mathcal{U}$ and τ statistics, the statistics γ_{GMS}, and our statistics γ_{NO} and γ_{OO} with γ_{NO-O} as a special case. The $\mathcal{U}$-based τ statistic counts relation consistent pairs of elements, which need not necessarily lead to a global rule. Therefore, τ measures a form of local consistency. The counting algorithm for γ_{GMS} is very strict: Only if there exists no counter example of the form $f_Q(x) \leq_Q f_Q(z) \wedge f_d(x) \gneq_d f_d(z)$ or $f_Q(y) \leq_Q f_Q(w) \wedge f_d(y) \gneq_d f_d(w)$, a consistent pair (x,y) votes for γ_{GMS}. Contrary to τ, γ_{GMS} measures a global consistency, which guarantees that the consistent observations are part of some rule – but global consistency may overlook interesting rules. The statistic γ_{OO} is in between both approaches, because it counts those consistent pairs which are rule consistent.

We have demonstrated that γ_{OO} can be used in analogy of an ordinary regression approach: The significance of the full model $\gamma_{OO}(\leq_Q, \leq_d)$ can be tested by randomisation methods, and the partial influence of every $q_i \in Q$ can be tested by means of the Bootstrap technology. But the method should be used with care: If we find out that one attribute q_k does not significantly increase the overall γ, we cannot conclude that there is no partial relationship between q_k and d, given the other attributes $q_1, \ldots q_{k-1}$. This is due the fact that the procedure does not check rules of the form

$$s_1 \leq f_{q_1}(x) \wedge \ldots \wedge s_{k-1} \leq f_{q_{k-1}}(x) \wedge s_k \geq f_{q_k}(x) \Rightarrow t \leq f_d(x).$$

It may happen that the evaluation of this rule shows better results. Therefore, in order to determine the influence of q_k, we need to perform (at least) two analyses. If both of these show that q_k shows no significant influence, we can conclude that there is no partial relationship between q_k and d. Because this argument can be iterated for the other $k-1$ attributes, we obtain 2^k different models which, in principle, have to be analysed; this results in an immense amount of computing time if many criteria are used. If domain knowledge poses the restriction that only one direction is of interest, the problem disappears, because the inverse rank orders of the attributes are of no interest in that case.

Although there is some safeguard against casualness using the randomisation significance test procedure, this problem should be kept in mind. A similar situation, the *collinearity problem* (Belsley et al., 1980), occurs in linear modelling: A high monotone correlation within the Q attributes may lead to bad prediction of the dependent d attribute.

Despite these problems, which our approach shares with all other methods, the given measures allow the researcher to evaluate <u>all</u> rules based on ordinal or mixed nominal-ordinal data within a multi-criteria decision making context. It is neither only locally applicable such as the traditional τ statistic – of which it can be regarded as a globalisation –, nor is it overly restrictive like the γ_{GMS} index. Therefore, we think that it is a suitable approach to the evaluation of sorting rules.

A final note concerns the relation of the proposed rough set methods to methods based on *(quasi) monotone decision trees* (e.g. Potharst and Bioch (2000)). As the name expresses, the latter methods are variants of the decision tree methodology (Breiman et al., 1984; Quinlan, 1986) and are focused on finding one good solution to the ordinal classification problem. In doing so, the relation of both approaches is quite the same as those of classical (nominal) rough set analysis and classical decision tree methodology: Both perform rule system analysis, but with a different and complementary focus. The focus of rough set analysis it to find promising attributes and offers evaluation strategies for this job, whereas decision tree methodology aims at (nearly) optimal tree representation of the data set.

Acknowledgement

We would like to thank the referees for their careful reading and constructive remarks which helped to improve the readability of the paper considerably.

References

Banzhaf, J. F., 1965. Weighted voting doesn't work: A mathematical analysis. Rutgers Law Review 19, 317–343.

Behnke, R., Berghammer, R., Meyer, E., Schneider, P., 1998. RELVIEW — A system for calculating with relations and relational programming. Lecture Notes in Computer Science 1382, 318–321.

Belsley, D., Kuh, E., Welsch, R., 1980. Regression Diagnostics: Identifying influential data and source of collinearity. Wiley, New York.

Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J., 1984. Classification and Regression Trees. Wadsworth International Group, Belmont.

Cliff, N., 1994. Predicting ordinal relations. British J. Math. Statist. Psych. 47, 127–150.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20, 37–46.

Düntsch, I., Gediga, G., 1997a. Relation restricted prediction analysis. In: Sydow, A. (Ed.), Proc. 15th IMACS World Congress, Berlin. Vol. 4. Wissenschaft und Technik Verlag, Berlin, pp. 619–624.

Düntsch, I., Gediga, G., 1997b. Statistical evaluation of rough set dependency analysis. International Journal of Human–Computer Studies 46, 589–604.

Efron, B., 1982. The Jackknife, the Bootstrap and other resampling plans. SIAM, Philadelphia.

Gediga, G., Düntsch, I., 2000. Program NOO: Computing approximation quality for sorting rules, available at http://www.roughian.co.uk/papers/noo.html.

Gediga, G., Düntsch, I., 2001. Rough approximation quality revisited. Artificial Intelligence 132, 219–234.

Gediga, G., Düntsch, I., 2002. On model evaluation, indices of importance, and interaction values in rough set analysis. In: S.K.Pal, L.Polkowski, A.Skowron (Eds.), Rough-Neuro Computing: A way for computing with words. Physica Verlag, Heidelberg, to appear.

Greco, S., Matarazzo, B., Słowinski, 1998a. Rough set handling of ambiguity. In: Zimmermann, H.-J. (Ed.), Proceedings of the 6th European Congress on Intelligent Techniques and Soft Computing (EUFIT'98). Mainz Wissenschaftsverlag, Aachen, pp. 3–14.

Greco, S., Matarazzo, B., Słowiński, R., 1998b. A new rough set approach to multicriteria and multiattribute classification. In: Polkowski, L., Skowron, A. (Eds.), Rough sets and current trends in computing. Vol. 1424 of Lecture Notes in Artificial Intelligence. Springer–Verlag, pp. 60–67.

Järvinen, J., Jun. 22–26 1998. Preimage relations and their matrices. In: Polkowski, L., Skowron, A. (Eds.), Proceedings of the 1st International Conference on Rough Sets and Current Trends in Computing (RSCTC-98). Vol. 1424 of LNAI. Springer, Berlin, pp. 139–146.

Kendall, M. G., 1970. Rank correlation methods. Charles Griffin, London.

Pawlak, Z., 1991. Rough sets: Theoretical aspects of reasoning about data. Vol. 9 of System Theory, Knowledge Engineering and Problem Solving. Kluwer, Dordrecht.

Potharst, R., Bioch, J., 2000. A decision tree algorithm for ordinal classification. In: Hand, D., Kok, J., Berthold, M. (Eds.), Advances in Intelligent Data Analysis. Vol. 1642 of Lecture Notes in Computer Science. Springer–Verlag, New York, pp. 187–199.

Puri, M., Sen, P., 1971. Non-parametric methods in multivariate analysis. Wiley, New York.

Quinlan, J. R., 1986. Induction of decision trees. Machine Learning 1, 81–106.

Skowron, A., Polkowski, L., 1997. Synthesis of decision systems from data tables. In: Lin, T. Y., Cercone, N. (Eds.), Rough sets and data mining. Kluwer, Dordrecht, pp. 259–299.

Wegener, I., 1987. The complexity of Boolean functions. Wiley, New York.