Relation restricted prediction analysis

Ivo Düntsch*
School of Information and Software Engineering
University of Ulster
Newtownabbey, BT 37 0QB, N.Ireland
I.Duentsch@ulst.ac.uk

Günther Gediga*
FB Psychologie / Methodenlehre
Universität Osnabrück
49069 Osnabrück, Germany
gg@Luce.Psycho.Uni-Osnabrueck.DE

and

Institut für semantische Informationsverarbeitung
Universität Osnabrück

Abstract

For the description of dependencies between a set of independent attributes Q and a dependent attribute p, a soft computing approach such as Rough Set Data Analysis (RSDA) uses only a very simple data representation model: The set of equivalence classes of feature vectors determined by Q, and p respectively. Although this model is satisfactory for many applications, there are sometimes problems to interpret the results, because this type of prediction does not take into account relational information within the attributes, for example, orderings. We consider the problem what form prediction should take in the "nominal attributes predict an ordinal attribute" situation ((n, o)-prediction), as well as in the (o, o)-situation. We show how to define rough (n, o)- and (o, o)-prediction and approximation in terms of relational compatibility, which respects the granularity information given by the attributes. A running example is presented to demonstrate the result of the three types of data analysis.

1 Introduction

Even though most data analyzed in psychological investigations are ordinal, a succinct methodology to cope with prediction in case of multiple ordinal variables is still missing, cf. Cliff (1). The frequently used *interval – scale assumption* supposes that a theoretical construct T (e.g. intelligence) and its measurement M (e.g. results in an intelligence test) are linearly related. It is astonishing that almost data analysis procedures (linear regression, factor analysis, analysis of variance etc.) assume that even "established" psychometric measurements, such as the IQ, are interval scaled. However, one can argue that making this assumption is a pragmatic decision, because of a missing methodological approach in the ordinal scaling case.

Even though nominal scaling has been around for some time – see for example Torgerson (4) – it has come into focus only in the past few years, and it can be developed into an alternative to the classical regression model. This type of data analysis uses only equivalence type information; in other words, objects are

*Equal authorship implied

Table 1: Contraception data

Country	X_1	X_2	X_3	X_4	Y
(1) Lesotho	3.9	4	73	0	6
(2) Kenya	0.9	4	108	6	9
(3) Peru	2.7	17	367	0	14
(4) Sri Lanka	3.8	20	142	12	22
(5) Indonesia	1.2	9	61	14	25
(6) Thailand	2.1	8	142	20	36
(7) Colombia	2.7	47	284	16	37
(8) Malaysia	1.6	29	313	18	38
(9) Guayana	6.1	20	318	0	42
(10) Jamaica	6.9	8	593	23	44
(11) Jordan	1.4	53	197	0	44
(12) Panama	5.3	50	570	19	59
(13) Costa Rica	4.7	18	464	21	59
(14) Fiji	3.7	15	321	22	60
(15) Korea	4.5	15	188	24	61

Table 2: Recoded data

Country	X_1	X_2	X_3	X_4	Y
Lesotho	1	0	0	0	6
Kenya	0	0	0	0	9
Peru	1	1	2	0	14
Sri Lanka	1	1	0	1	22
Indonesia	0	0	0	1	25
Thailand	1	0	0	1	36
Colombia	1	2	1	1	37
Malaysia	0	1	2	1	38
Guayana	2	1	2	0	42
Jamaica	2	0	2	2	44
Jordan	0	2	1	0	44
Costa Rica	2	1	2	2	59
Panama	2	2	2	1	59
Fiji	1	1	2	2	60
Korea	2	1	1	2	61

discernible only up to a classification (see Gigerenzer (2), pp 133ff, and Pawlak (3)). This *nominal scale approach* is very liberal, because any function between a theoretical construct and its measurement which preserves the classification structure is admissible. We propose in this paper a simple prediction analysis approach with order – structured data, using the terminology of RSDA developed by Pawlak (3).

To demonstrate the procedures, we will use a data set published by Cliff (1) (Table 1, Table 2). This data set was used to discuss statistical approaches to ordinal data analysis.
It is aimed to predict the ordinal relation of the countries

- % ever practicing contraception (Y)

from the characteristics

- Average years of education (X_1),
- Percent urbanized (X_2),
- Gross national product per capita (X_3),
- Expenditures on family planning (X_4),

As the basis for knowledge representation we use the OBJECT → ATTRIBUTE relationship, in which each object can be represented by a vector where each coordinate represents an attribute. More formally, an *information system*

$$\mathcal{I} = \langle U, \Omega, V_q, f_q \rangle_{q \in \Omega}$$

consists of

1. A finite set U of objects,
2. A finite set Ω of attributes (features),
3. For each $q \in \Omega$

(a) A set V_q of attribute values,

(b) An information function $f_q : U \to V_q$,

cf Pawlak (3).

If $Q \subseteq \Omega$, then we define an equivalence $\equiv_Q$ on U by

$$x \equiv_Q \stackrel{\text{def}}{\Longleftrightarrow} (\forall q \in Q)(f_q(x) = f_q(y))$$

If no confusion can arise, we usually identify singletons with the element that they contain.

2 The nominal – nominal case

The simplest dependency is based on the observation that objects in U can only be distinguished up to their feature vectors. The appropriate classifications are given by the equivalence relations $\equiv_Q$: Intuitively, $x \equiv_Q y$ if the objects x and y are indiscernible with respect to the values of their attributes from Q, and we regard these equivalence relations as nominal scales. A class A of $\equiv_Q$ is *compatible* (with $\equiv_p$) if it is a subset of a class of $\equiv_p$. A deterministic rule in this setting takes the form

$$(\forall z \in U)\left[\bigwedge_{i \leq m} f_{q_i}(z) = t_{q_i} \Rightarrow f_p(z) = t_Q\right],$$

The union M of all compatible sets can serve as a statistic of the (nominal) prediction success using deterministic rules. The normalized statistic

$$\gamma_{n,n}(Q \to p) = \frac{|M|}{|U|}$$

is called the *(n,n) – approximation quality of* p *by* Q *with respect to the nominal scales* $\equiv_Q$ *and* $\equiv_p$ (nominal, nominal). If $\gamma_{n,n}(Q \to p) = 1$, the prediction success is perfect, and in this case we call p *(n,n) – dependent on* Q; this is the usual rough set definition. The $\gamma_{n,n}$ values for our example are given in Table 3.

3 The nominal – ordinal case

Now, suppose that we have a linear order relation $\leq_p$ on V_p. If $X, Y \subseteq V_p$, we let

$$X \leq_p^+ Y \text{ if and only if } (\forall x \in X, y \in Y)(x \leq_p y).$$

The question that we ask is the following:

- If A, B are different classes of $\equiv_Q$, how are their images $f_p(A), f_p(B)$ related in $\leq_p^+$?

We call $A, B \in K_Q$ *compatible*, if either

1. $f_p(A) \leq^+ f_p(B)$, or
2. $f_p(A) \geq^+ f_p(B)$.

Table 3: The approximation values of Table 2

Prediction Set	$\gamma_{n,n}(Q \to p)$	$\gamma_{n,o}(Q \to p)$	Predicted Intervals
X_1, X_2, X_3, X_4	1.00	1.00	all elements
X_1, X_2, X_3	0.47	0.60	$\{[9], [22], [25], [37], [38], [42, 59], [59], [61]\}$
X_1, X_2, X_4	0.87	0.93	all elements, but 60, and one interval $[59, 61]$
X_1, X_3, X_4	0.73	0.93	all elements, but 25, and two intervals $[22, 36]$, $[59, 61]$
X_2, X_3, X_4	0.47	0.87	$\{[6, 9], [22], [25, 36], [37], [38], [44], [44], [59], [59, 60], [61]\}$
X_1, X_2	0.33	0.47	$\{[6, 36], [37], [38], [42, 61]\}$
X_1, X_3	0.27	0.60	$\{[6, 36], [37], [38], [42, 59], [61]\}$
X_1, X_4	0.20	0.60	$\{[6, 14], [22, 37], [42], [44, 61]\}$
X_2, X_3	0.27	0.53	$\{[6, 36], [37, 44], [59], [61]\}$
X_2, X_4	0.13	0.73	$\{[6, 9], [25, 36], [37, 59], [59, 61]\}$
X_3, X_4	0.20	0.73	$\{[6, 9], [22, 36], [37], [44], [44, 60], [61]\}$
X_1	0.00	0.40	$\{[6, 60]\}$
X_2	0.00	0.47	$\{[14, 61]\}$
X_3	0.00	0.53	$\{[6, 36], [37, 61]\}$
X_4	0.00	0.40	$\{[22, 59]\}$

A subset M of K_Q is *compatible*, if each two different elements are compatible. The (n,o) – approximation quality (nominal – ordinal) is defined by the

$$\gamma_{n,o}(Q \to p) \stackrel{\text{def}}{=} \max\{\frac{|\bigcup M|}{|U|} : M \text{ is a compatible subset of } K_Q\}.$$

A rule in this system has the form

(3.1) $$\text{If } A, B \in K_Q, A \neq B, \text{ then } f_p(A) \leq_p^+ f_p(B) \text{or } f_p(A) \leq_p^+ f_p(B).$$

Unlike the previous case (which was essentially unary), the choice of a compatible set with maximum cardinality is not necessarily unique. In this case, the researcher faces the problem that (s)he has to choose among different rule systems with the same approximation quality, and thus, other, possibly semantic, criteria have to be applied.

If $\gamma_{n,o}(Q \to p) = 1$, we call p *(n,o) – dependent on* Q. In this case, the prediction success is perfect, and the images of the classes of $\equiv_Q$ overlap at most in their extremal elements.

It is straightforward to see, that for any linear order $\leq_p$ on V_p we have $\gamma_{n,n} \leq \gamma_{n,o}$, since for each deterministic class $A \in K_Q$, $f_p(A)$ is a singleton. In the full paper we consider more general relations than linear orders, and additional structural properties.

Example: Table 3 shows the results of the (n, n) – analysis and the (n, o) – analysis applied to the data set given in Table 2. All possible combinations of the attributes X_1, X_2, X_3, X_4 are listed to predict the decision attribute $p = Y$. Because for each $\emptyset \neq Q \subseteq \{X_1, X_2, X_3, X_4\}$ there is a unique compatible subset of K_Q with maximum cardinality, we present in the last column the predicted intervals of the associated compatible set.

As an example, consider $Q = \{X_3\}$, for which we have the following information:

Class	f_{X_3}–value	$\min f_p$	$\max f_p$
A: Lesotho, Kenya, Sri Lanka, Indonesia, Thailand	0	6	36
B: Colombia, Jordan, Korea	1	37	44
C: Peru, Malaysia, Guayana, Jamaica, Costa Rica, Panama, Fiji	2	14	61

The only compatible classes are A and B, and we obtain the rule

$$\text{If } f_Q(x) = 0 \text{ and } f_Q(y) = 1, \text{ then } 6 \leq f_p(x) \leq 36 \text{ and } 37 \leq f_p(y) \leq 61.$$

Since $|A \cup B| = 8$, we obtain $\gamma_{n,o}(X_3 \to Y) = 0.53$.

The results show that X_3 alone is a fairly good candidate for (n, o) – prediction, because it discriminates between two (of three) classes; a result which cannot be achieved by using (n, n) – approximation.

If we are content with nearly perfect prediction success, both analyses offer the set X_1, X_2, X_4, but the (n, o)-prediction offers 2 additional options, namely, $\{X_1, X_3, X_4\}$ and $\{X_2, X_3, X_4\}$.

4 The ordinal – ordinal case

In this section we assume that, in addition to $\leq_p$, we are given a (not necessarily linear) order $\leq_Q$ on K_Q. Our aim is to transport the order $\leq_Q$ to $\langle \mathfrak{P}(V_p), \leq_p^+ \rangle$ while at the same time respecting the indiscernibility relations $\equiv_Q$ and $\equiv_p$.

If $\leq_Q$ arises, for example, from a product of linear orders on the sets V_Q, it is important to point out that a decision has to be made, which of the possible partial orders shall be investigated. Therefore, one may have to apply the procedure to more than one case of ordering, if the context of the research so requires. In the sequel, we suppose that the orders under consideration are fixed.

We call $A, B \in K_Q$ *compatible*, if $A \lneqq_Q B$ implies $f_p(A) \leq_p^+ f_p(B)$. A subset of K_Q is *compatible*, if any two different elements are compatible.

As before, the (o,o) – approximation quality is defined by the maximal cardinality of the union of compatible sets:

$$\gamma_{o,o}(Q \to p) \stackrel{\text{def}}{=} \max\{\frac{|\bigcup M|}{|U|} : M \text{ is compatible}\}.$$

We now say that p *is (o,o)–dependent on* Q (ordinal – ordinal), if $\gamma_{o,o}(Q \to p) = 1$; Table 4 shows the (o,o) – dependency values. Whereas the (n, o)-analysis as well as the (n, n)-analysis favor the selection of the attributes $\{X_1, X_2, X_4\}$, the (o, o)-analysis shows the ordinal approximation quality success will not increase from the approximation quality of the attributes $\{X_2, X_4\}$ or $\{X_3, X_4\}$; obviously, the attribute X_1 has no ordinal impact on the dependent attribute.

5 Conclusion

We have presented a methodology to analyze ordered data in the context of an underlying nominal scale assumption such as rough set data analysis. We have pointed out that there is an urgent need to develop such methods, because the current practice in data analysis uses interval – scaled strategies, even though the underpinning of interval – scales is questionable in many applications, and ordering relations seem to be better candidates in many cases.. Similar to (nominal) rough set data analysis, we define and discuss the prediction success – called approximation quality – and prediction rules.

There are some – we think minor – drawbacks of this approach. First, even in the simplest setting of one nominal and one ordinal attribute we observe a dissociation between the approximation quality and the fixing of a rule system. Further research is necessary to investigate the nature and the impact of

Table 4: Rough order analysis

Prediction Set	$\gamma_{o,o}(Q \to p)$	Intervals
1,2,3,4	0.73	$\{[6], [14], [25], [36], [37], [38], [42], [44], [59], [60], [61]\}$
1,2,3	0.53	$\{[6, 36], [38], [44], [44], [42, 59], [59]\}$
1,2,4	0.73	$\{[6], [14], [25], [36], [38], [42], [44], [44], [59], [59, 61]\}$
1,3,4	0.73	$\{[6], [14], [22, 36], [37], [42], [44], [59], [60], [61]\}$
2,3,4	0.73	$\{[6, 9], [14, 42], [25, 36], [37], [44], [59], [59, 60]\}$
1,2	0.47	$\{[9, 25], [44], [42, 61]\}$
1,3	0.53	$\{[6, 36], [37], [42, 59]\}$
1,4	0.60	$\{[6, 14], [22, 37], [42], [44, 61]\}$
2,3	0.53	$\{[14, 59], [37, 44], [59]\}$
2,4	0.73	$\{[6, 9], [14, 42], [25, 36], [44], [44], [59, 61]\}$
3,4	0.73	$\{[6, 9], [14, 42], [22, 36], [37], [44, 60]\}$
1	0.40	$\{[6, 60]\}$
2	0.47	$\{[14, 61]\}$
3	0.53	$\{[6, 36], [37, 61]\}$
4	0.40	$\{[22, 59]\}$

this dissociation. Second, in a pre–processing step, the researcher has to decide which relations are meaningful in the context. Otherwise, in case of no additional knowledge, (s)he faces a computational problem: If k ordinal relations are involved in the attribute domains of Q , then 2^{k-1} different data analyses are needed.

References

[1] Norman Cliff, *Predicting ordinal relations*, British J. Math. Statist. Psych. **47** (1994), 127–150.

[2] G. Gigerenzer, *Messung und Modellbildung in der Psychologie*, Birkhäuser, Basel, 1981.

[3] Zdzisław Pawlak, *Rough sets*, Internat. J. Comput. Inform. Sci. **11** (1982), 341–356.

[4] W. S. Torgerson, *Theory and methods of scaling*, Wiley, New York, 1958.