On Diagnostic Categories of Archetypal Psychiatric Patients

Ivo Düntsch*
Department of Computer Science
Brock University
St. Catherines, Ontario, Canada, L2S 3AI
duentsch@cosc.brocku.ca

Günther Gediga*
Institut für Evaluation und Marktanalysen
Brinkstr. 19
49143 Jeggen, Germany
gediga@eval-institut.de

Joseph Rogner*
Fachbereich Psychologie/Diagnostik
Universität Osnabrück
49069 Osnabrück, Germany
jupprogn@mail.rz.Uni-Osnabrueck.DE

Abstract

The archetypal diagnostic approach is based on defining features which not all members of the class must possess [3]. In this study, we re-analyse data of the diagnostic categories of archetypal psychiatric patients presented by Mezzich & Solomon [22] with the method of enhanced rough set data analysis (RSDA) [11, 29].

RSDA has recently come to the fore as a means of data analysis in the medical sciences [e.g. 2, 7, 32, 33, 37]; it is a non-invasive method which uses only internal data information and does not rely on external parameters as most other taxonomic methods do.

Even though the data set has been used in connection with DSM–II, our results show that the findings are compatible with the outcome of more recent studies. We also show that some of our results are significant in terms of highlighting dependencies among diagnostic categories which have not been noticed before.

1 Introduction

The prototype (or archetypal) approach was introduced into psychiatry by Cantor et al. [3]. In contrast to the classical view which is based on defining features – i.e. a set of necessary and sufficient attributes – and perfect nesting, it assumes only necessary features and additionally some attributes which are correlated to the concept to a more or lesser degree, and which not all members of the class must possess. Furthermore, the features are not necessarily hereditary with respect to subclasses.

*The order of authors is alphabetical, and equal authorship is implied

In the study, 13 experienced psychiatrists were asked to list the "clinical features that characterize the prototypal patient" (p. 186) for nine (partially nested) diagnostic categories of DSM-II. The authors report a high distinctiveness of schizophrenic and affective disorders (having only one common feature), and they find a high consensus, for example, for

- Manic – depressive illness, depressed: **Flight of ideas, euphoria (elation)**, grandiosity,
- Manic – depressive illness, manic: **Depression**, hopelessness and despair, **psychomotoric retardation**,
- Paranoid schizophrenia: **Delusions of persecution, hostility**, suspicion.

The bold features are defining characteristics for this type of disorder in DSM-II.

DSM-III – introduced in 1980 – was a fundamental revision of DSM-II into the direction of the prototype view, with a significantly narrower definition of schizophrenic disorders [19], and a radical change in the classification of depressive disorders [15]. Since the current study is based on an investigation centered on DSM–II, its results cannot be transferred to DSM-III and its successor DSM-IV. However, having said this, we believe that our method, being a general taxonomic data analysis procedure, is applicable to the newer standards as well.

In a related setup, Mezzich & Solomon [22], Chapter 4, pp. 59–83, present a study of diagnostic categories of archetypal psychiatric patients, on which the present work is based. 22 experienced psychiatrists were asked to describe, one at a time, a typical patient for each of four diagnostic categories from the DSM-II nomenclature, namely,

- Manic – depressive illness, depressed (MDD)
- Manic – depressive illness, manic (MDM).
- Simple schizophrenia (SS).
- Paranoid schizophrenia (PS).

Each archetypal patient was described on a scale 0 (not present) to 6 (extremely severe) of ratings of 17 symptoms from the Brief Psychiatric Rating Scale [BPRS, 28] given in Table 1. The current 18th symptom, *Disorientation*, had not been included in the BPRS at the time.

The BPRS is a "workhorse" of psychiatric research [36], in particular for schizophrenic disorders. In connection with several scaling models, it has been useful in investigating the dimensionality of schizophrenic psychopathology; the most common models have three to five dimensions [6, 17, 18, 21, 35]. For example, Lindenmayer et al. [18] find the following five features which have been replicated in four other analyses:

• Negative • Excitement • Cognitive • Positive • Depression.

Table 1: BPRS symptoms

1.	Somatic concern (SC)	10.	Hostility (HO)
2.	Anxiety (AN)	11.	Suspiciousness (SU)
3.	Emotional withdrawal (EW)	12.	Hallucinatory behaviour (HB)
4.	Conceptual disorganization (CD)	13.	Motor retardation (MR)
5.	Guilt feelings (GF)	14.	Uncooperativeness (UN)
6.	Tension (TE)	15.	Unusual thought content (UT)
7.	Mannerism and posturing (MP)	16.	Blunted effect (BE)
8.	Grandiosity (GR)	17.	Excitement (EX)
9.	Depressive mood (DM)	(18.)	Disorientation

The aim of [22] is to find out how well various cluster analytic procedures are able to approximate or even replicate the categories set by the experts. To this end, the authors analyze the data with 18 different quantitative taxonomic methods and rank them according to specified criteria.

> "Their variety may resemble the variety observed among human judges with different theoretical backgrounds and cognitive styles, undertaking the task of developing a 'clinically useful' classification of psychiatric patients" (p. 81).

All methods used in [22] need parameters outside the given data, e.g. measures of similarity (correlation), relationship metrics (e.g. Euclidean distance), or assumption about the distributional representation of the data (e.g. ISODATA, NORMIX). Therefore, the study is concerned with applying "hard computing" methods to a soft data base. By "hard computing" we understand, in particular, external model assumptions which are needed in order to apply the statistical methods. Each such assumption, however, is a subjective, interpretative decision of the researcher which may influence the outcome of the experiment [see e.g. 13]. Hence, a method, which uses less external model assumptions and achieves at least the same reliability and validity as the "hard methods", has the advantage of being more objective, and hence, more credible.

In the present paper, we will re-analyze the data using a "soft", non-invasive, method, namely, "Rough information analysis", which does not require any external assumptions or parameters, and works only with the given operationalisation of the data.

Our intentions are

- To demonstrate the applicability of a non-invasive method of data analysis to psychiatric diagnosis,
- To show that the opinions of the experts can be described by an abundance of possible rule based algorithms which rely on only a few features each, and to show that it is possible to discover significant criteria of the experts' judgements,

- To show, last but by no means least, that some of the results are significant in terms of highlighting differences and dependencies among diagnostic categories which seem not to have been noticed before, thus providing the clinician with a more broadly based range of choice.

We should like to point out that our investigation has different aims than those of the study of Mezzich & Solomon [22]. While their aim was to find the most suitable method for unsupervised learning of the diagnostic classes, our aim is to apply one supervised method (i.e. with given decision classes) and find out which symptoms are significant to which degree for the experts' diagnoses. Thus, our results should not be compared with those in [22]. We shall briefly touch unsupervised learning in rough set data analysis in Section 4.2.

2 Rough information analysis

Suppose we have a set of features (attributes) Ω, and a set of objects U whose elements are are described by feature vectors $\{\bar{x}^{\Omega} : x \in U\}$, and a decision attribute d with, say, n classes (values). In our case, U is the set of experts' opinions, Ω is the chosen set of symptoms from the BPRS, and d is the archetypal diagnostic class which takes the four values MDD, MDM, SS, and PS. A tabular representation of the data as an *information system* [29], can be found in Table 2. There, for example, the expert ratings 1.1. – 1.4 are the values which expert 1 associates with the respective archetypes.

Table 2: An information system

Expert rating	Severity of BPRS symptoms																	Type
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
1.1	4	3	3	0	4	3	0	0	6	3	2	0	5	2	2	2	1	MDD
1.2	2	2	1	2	0	3	1	6	2	3	3	2	1	4	4	0	6	MDM
1.3	3	2	5	2	0	2	2	1	2	1	2	0	1	2	2	4	0	SS
1.4	2	4	3	5	0	3	1	4	2	5	6	5	0	5	6	3	3	PS
2.1	5	5	6	2	6	1	0	0	6	1	0	1	6	4	1	4	0	MDD
2.2	0	0	0	4	1	5	0	6	0	5	4	4	0	5	5	0	6	MDM
2.3	4	4	5	4	3	3	1	0	4	2	3	0	3	2	4	5	0	SS
2.4	2	4	1	1	0	3	1	6	0	6	6	4	0	6	5	0	4	PS
80 other values																		

Each subset Q of Ω and the diagnostic class d determine equivalence relations θ_Q, θ_d on U, where two elements $x, y \in U$ are identified in θ_Q (θ_d), if $\bar{x}^Q = \bar{y}^Q$ ($\bar{x}^d = \bar{y}^d$). Each class of θ_Q gives us a rule in the following way: Suppose for simplicity that $Q = \{p, q\}$. Each combination of values which p and q can take leads to a class M of θ_Q, which collects all those elements of U whose values on the features p and q are the same. Each such class intersects one or more classes $T_1, \dots, T_k$ of θ_d; in other words, each $x \in U$ which is in class M, takes one of the values of d associated with the classes $T_1, \dots, T_k$. For an example, some of the deterministic rules associated with the attribute set $Q = \{EW, DM\}$ are given in Table 3.

Table 3: Example of rules

EW	DM	d	No.	EW	DM	d	No.
6		MDM	22	1	2	MDM	2
0		MDM	18	1	0	MDM	2
5	2	SS	6		4	2	2

The first line of Table 3 consists of the two rules

If EW = 6, then d = MDM in 22 cases,
If EW = 1 and DM = 2, then d = MDM in 2 cases.

We shall denote the collection of rules generated by Q with respect to d by $Q \implies d$, and, with some abuse of language, call $Q \implies d$ itself a rule.

In the remainder of the section we shall outline the steps of our non – invasive data analysis

- Feature reduction,
- Significance analysis,
- Data filtering, and
- Uncertainty measuring.

We shall see that none of these involves any extraneous parameters or outside model assumptions, according to our objective

LET THE DATA SPEAK FOR THEMSELVES.

2.1 Feature reduction

The reduction of features without changing the structure of the system is one of the major features of rough set data analysis [29]. The reduction problem is also mentioned in [22]:

> "This problem – of finding a grid of fewer orthogonal axes to replace the grid of correlated axes . . . is fundamental to multivariate data analysis. If the number of dimensions can be reduced to two or three without too much loss of information, some ease is achieved since elements can be grouped by eye" (p. 6).

Rough set data analysis presents simple methods to achieve such reduction, if it is at all possible, without changing the internal structure of the data.

First, we look for attributes $q \in \Omega$ which are exactly a combination of others, i.e. if there exist some $P \subseteq \Omega$ with $\theta_q = \theta_P$. Such attributes are unnecessary overhead, and removing them from Ω does not change the dependency structure of the whole system.

Next, we check to what degree the features given by some $Q \subseteq \Omega$ are sufficient to explain d in the following way: A class of θ_Q is called *d – deterministic*, if it is contained in a class of θ_d, otherwise we call it *d – indeterministic*. The *approximation quality of Q with respect to d* is now defined as

$$\gamma(Q \implies d) \stackrel{\text{def}}{=} \frac{|\bigcup\{X : X \text{ is a } d - \text{deterministic class of } \theta_Q\}|}{|U|}.$$

If $\gamma(Q \implies d) = 1$, then each class of θ_Q is d – deterministic, i.e. each class of θ_d can be completely and uniquely determined by by the values its elements take on the feature set Q. In this case, we call *d dependent on Q*, written as $Q \Rightarrow d$. If Q is a proper subset of Ω, then we have achieved a feature reduction.

The next step is to try recognize the dependencies among the features and reduce the number of attributes as much as possible. We call a set $Q \subseteq \Omega$ a *reduct with respect to d* (or just a *reduct* if d is understood) if it is minimal with respect to the property

$$\gamma(Q \implies d) = 1.$$

It might be worth to point out that, in general, reducts are not unique; many reducts indicate a high substitution rate among the features.

2.2 Significance testing

If a rule is based on a few observations only, the granularity of the system is too high, and the rule may be due to chance. In order to test the significance of rules, one can use randomization methods to find the probability of the rule, given that the null hypothesis is true [8]. If this probability is low, traditionally not more than 5%, we reject the null hypothesis, and call the rule *significant*. Failure to reject the null hypothesis does not mean that it is true, and thus, such randomization tests are a necessary condition for significance [for a discussion, see 5]. Statistical significance tests must accompany any rule based method of data analysis in order to give credible results.

Randomization is a statistical technique which does not require a representative sampling from a population which is a theoretical generalization of the sample under study, because the randomization procedure uses only information within the given sample, well in accord with our stated objective. This aspect is in contrast to most other statistical methods.

Details of the use of randomisation techniques for significance testiong in RSDA can be found in [8].

2.3 Data filtering

If we find that the significance of a rule is low we can use internal data filtering techniques described in [9] to possibly increase their significance. These techniques collect values of a feature into a single value by taking a union of deterministic equivalence classes which are wholly contained in a class of the decision attribute; therefore, the underlying statistical basis of the rule may be enlarged. The important feature of this procedure is that the internal dependency structure of the system is kept intact, and that we do not need additional parameters. In other words, this step can be regarded as a part of the operationalization procedure; it can be implemented as a cheap standard algorithm if the decision attribute d is fixed.

As an example, we observe in our system that for attribute 13 (Motor retardation, MR)

- If MR = 2 or MR = 3 or MR = 4, then d = SS.

Thus, we can replace the three values 2,3,4 by one value, say, 3, without changing the structure of the system.

2.4 Uncertainty measures

Although the randomization technique introduced in Section 2.2 is quite useful,it is rather expensive in resources, and it is only applicable as a conditional testing scheme. Though it tells us when a rule may be due to chance, it does not provide us with a metric for the comparison of two different rules $Q \implies d$, $R \implies d$, let alone for different models of uncertainty.

Düntsch & Gediga [10] suggest three different approaches to describe the amount of uncertainty of rules in the spirit of the *minimum description length* approach of [31]. Again, only internal information is used, and the main approach is based on the maximum entropy principle: Given a class Y of θ_d, any indeterministic observation y is a result of a random process whose characteristics are totally unknown to the researcher. In other words, our given data is the partition obtained from Q, and we know the world only up to the equivalence classes of θ_Q. Given this assumption, no information within our data set will help us to classify the element y, and we conclude that each such y requires a rule (or class) of its own. This results in a partition of the object set, whose entropy is called the *entropy of deterministic rough prediction with respect to* $Q \implies d$. If $Q \Rightarrow d$, then the entropy of the rule reduces to the entropy of the probability distribution induced by θ_Q.

In analogy to γ, this entropy measure is normalized to the interval [0,1] in such a way that, similarly to γ, a high value is more favourable, and signifies low uncertainty and vice versa, called *Normalized Rough Entropy* (NRE).

The rough information analysis which we shall perform with the data consists of the following steps:

1. Remove unnecessary overhead, i.e., check for any dependencies between single attributes and a set of attributes.
2. Filter the independent attributes with respect to the decision attribute 'Diagnostic category".
3. Find all reducts of the system.
4. Choose the most interesting reducts in terms of size, statistical significance and normalized rough entropy.

3 Analysis of the full data set

There are no dependency relations between single attributes and sets of attributes. In other words, no attribute can be expressed exactly as a combination of others.

In the next preprocessing step, we have filtered out unnecessary data values in the underlying scaling with respect to the full data set as described in Section 2.3, and found that the simplifications shown in Table 4 do not affect the data structure. There, an expression such as $a,b,\ldots \mapsto z$ means that the attribute values $a,b,\ldots$ can be simultaneously replaced by the value z without changing the validity of a rule. We have used this filter in the subsequent analysis of the full set.

Table 4: Data filtering, full data set

Symptom	Filter
Guilt feelings	$4,5,6 \mapsto 4$
Depressive mood	$5,6 \mapsto 5,$
Motor retardation	$2,3,4 \mapsto 3$
Excitement	$4,5 \mapsto 5$

We have found that the complete system has 649 reducts with respect to d, each of which contains at most five symptoms. This signifies a huge inhomogeneity among the symptoms; in other words, the dependency among the symptoms is high, and there are many possibilities for substitution. Because of the abundance of reducts, we have concentrated on those with no more than three elements; a list of these is given in Table 5. For each reduct, we list the number of rules and the normalized rough entropy (NRE).

Of particular interest is the reduct containing only the attributes *Depressive mood* (9) and *Excitement* (17) which shows the highest NRE. Since these two symptoms suffice to classify a patient into one of the four categories – which is a considerable reduction of the 17 symptoms used in the study – we have employed the randomization techniques described in Section 2.2 to test the significance of the rule in full set, the training set, and the testing set of [22]. The astonishing result: In all three data sets the null hypothesis can be rejected by quite a large margin, so that the rules derived from the dependency

Table 5: Small reducts

Reduct	Rules	NRE	Reduct	Rules	NRE	Reduct	Rules	NRE
9 17	13	0.534	2 3 9	16	0.304	6 9 11	21	0.259
5 13 17	12	0.439	5 11 14	22	0.294	3 4 5	22	0.258
13 16 17	14	0.416	3 7 9	18	0.292	2 9 15	26	0.252
3 5 17	14	0.403	2 3 13	20	0.291	9 12 14	21	0.252
5 16 17	14	0.396	3 13 14	19	0.286	3 4 9	19	0.247
5 14 17	14	0.377	4 8 9	22	0.286	1 12 16	31	0.245
1 13 17	15	0.350	4 5 17	15	0.284	1 4 17	22	0.240
3 9 12	18	0.345	5 11 17	13	0.283	6 9 12	25	0.238
7 13 17	12	0.338	3 9 11	14	0.279	1 4 8	27	0.231
3 9 14	17	0.326	3 9 10	18	0.278	1 5 15	27	0.231
5 7 17	15	0.326	9 10 11	22	0.274	4 5 11	24	0.225
8 9 11	17	0.324	7 9 11	20	0.273	1 10 15	33	0.221
3 7 13	20	0.323	15 16 17	28	0.269	5 6 12	28	0.221
1 3 13	23	0.319	9 11 16	17	0.267	5 6 11	23	0.221
9 11 14	20	0.318	3 4 13	20	0.266	1 11 16	25	0.219
3 5 10	20	0.315	1 9 12	24	0.264	1 3 4	31	0.211
3 5 14	19	0.315	5 7 11	22	0.263	6 11 16	31	0.209
11 13 16	17	0.312	1 11 13	21	0.262	1 4 16	32	0.204
3 6 9	18	0.307	9 10 12	24	0.261	1 15 16	35	0.194
4 13 17	13	0.305						

“Depressive mood and Excitement determine the diagnostic category”

can be considered significant.

Osgood et al. [27] state that the connotative meaning of concepts can be based on the three dimensions *evaluation, potency, activity*. We can say that “Depressive mood” corresponds to “Evaluation”, and “Excitement” to “Activity”. Thus, the reduct containing these two attributes reflect two fundamental dimensions of meaning. In the clinical area, there are connections of the reduct to the “Pyramidical model of schizophrenia” [16], which specifies a triangular base, constituted from negative, positive and depressive features and a separate vertical axis made up from excitement. Goldberg et al. [14] assert that common psychiatric symptoms can be reduced to two dimensions, namely, anxiety and depression. The work of Ormel et al. [26] shows that anxiety splits into “Phobic anxiety” and “Generalized anxiety”, where the latter corresponds to the attribute “Excitement”. Similarly, Clark & Watson [4] assume in their tripartite model besides specific factors for anxiety and depression a general distress factor which also corresponds to “Excitement”.

Hence, our emphasis on “Excitement” and “Depressive mood” is supported by the fact that these are basal symptoms in other investigations as well.

All rules of the reduct $\{9, 17\}$ using the filtered data are given in Table 6. We have normalized the rules in the sense that we chose those with a minimal number of conjunctions and no disjunctions on the left hand side. Values enclosed in $\langle \ldots \rangle$ are those which an attribute takes which is not necessary to make up the rule. For example, the first rule of Table 6 says

$$\text{If Depressive mood = 5, then d = MDD and Excitement} \in \{0, 1, 2, 3\},$$

and the fourth rule is

$$\text{If Depressive mood = 2 and Excitement = 3, then d = PS.}$$

Table 6: Dependency rules for the reduct $\{9, 17\}$

Attribute values			No of pat.	Attribute values			No of pat.
Depr. mood	Excit.	Class		Depr. mood	Excit.	Class	
5	⟨0,1,2,3⟩	MDD	22	0	0	SS	4
⟨0,1,2,3⟩	6	MDM	22	1	0	SS	2
⟨0,1,2,3⟩	4	PS	16	2	0	SS	6
2	3	PS	2	3	0	SS	2
2	2	PS	2	3	1	SS	2
1	1	PS	2	3	2	SS	4
				4	⟨0⟩	SS	2

We see that the top rating in the two symptoms determines each of the manic-depressive types. The next lower value of the *Excitement* symptom determines 16 of the 22 cases of PS. These phenomena are replicated in the training and learning set. In other words, two (filtered) values determine the two manic depressive classes, and one value of *Excitement* determines most of PS. It is therefore of no surprise that *Excitement* is the more important attribute of the two: The approximation qualities of *Depressive mood* and *Excitement* is 0.27, respectively 0.43, so that the drop is larger when *Excitement* is omitted (see also Figure 1).

We can see from the rules given in Table 6 that the rules determining SS are rather heterogeneous. This is not surprising, since its diagnosis presupposes a history of schizophrenic symptoms; historical and socialization data, however, is not covered by the BPRS. A similar result was obtained already by Blashfield [1] who could not cover SS with the Inpatient Multidimensional Psychiatric Scale [20].

One might argue that the simple structure of the results is due to the chosen data analysis. However, this is not the case. Using ordinary discriminant analysis given the diagnostic categories as groups and “Excitement” (EX) and “Depressive Mood” (DM) as dependent variables, we found a similarly simple result with about the same – impressive – reclassification rate (Table 7).

The members of the MDD and MDM groups are exactly reclassified for the same reason as in the rough analysis: Extreme DM values indicate an MDD membership whereas extreme EX values predict an MDM membership. Discriminant analysis has slight problems to discriminate between MDD

Table 7: Discriminant analysis

Reclassification	Predicted Group			
Actual Group	MDD	MDM	SS	PS
MDD	22 (100%)	0	0	0
MDM	0	22 (100%)	0	0
SS	2 (9.1%)	0	20 (90.9%)	0
PS	0	0	4 (18.2%)	18 (81.8%)

Group descriptions	Group			
Mean / Stand. Dev.D	MDD	MDM	SS	PS
Mean Depr. Mood	5.91	0.82	2.09	1.55
Stand. Dev. Depr. Mood	0.29	1.05	1.27	0.91
Mean Excitement	0.59	6.00	0.45	3.73
Stand. Dev. Excitement	0.96	0.00	0.80	1.24

and SS as well as between SS and PS. If we inspect the structure of the rules in Table 6, the reason is obvious: The ratings "DM $\in \{0,1,2,3\}$" are spread among the three categories MDD, SS, and PS, and a linear classification function will run into problems.

We have inspected the other possible combinations of symptoms with respect to their approximation quality of the diagnostic categories. Table 8 shows all combinations of two attributes with $\gamma \geq 0.8$. There, for a class T of θ_d and two symptoms p, q, α(T) is the ratio of the number of all elements of U which are certainly in the class T to the number of elements which are possibly in T, using the knowledge given by the classes of $\theta_{p,q}$. A high value indicates only a few uncertain cases for this particular class, and a value = 1 says that T is a union of classes of $\theta_{p,q}$.

Table 8: All combinations of two symptoms with $\gamma \geq 0.80$

Symptoms		γ	α(MDD)	α(MDM)	α(SS)	α(PS)	NRE
Excitement	Depr. mood	1.000	1.000	1.000	1.000	1.000	0.534
Excitement	Guilt feeling	0.920	1.000	1.000	0.720	0.833	0.556
Excitement	Motor retard.	0.909	0.667	1.000	0.714	1.000	0.553
Emotional withdr.	Depr. mood	0.841	1.000	0.833	0.615	0.533	0.396

We observe that, in combination with EX, GF has difficulties in correctly classifying the two types of schizophrenia, whereas MR cannot completely separate MDD and SS; as above, SS seems hardest to classify. The NRE of both sets is slightly lower than that of the reduct EX, DM. The combination MR, GF, EX is a reduct, however, its NRE is only 0.398.

If we substitute EX by EW, the approximation quality of 0.841 may still be acceptable; however, the combination performs poorly on the schizophrenic classes, and there is a significant drop in NRE.

We should remark that the attribute set "Excitement" and "Depressive mood" does not have the highest NRE. Among the three attribute sets with highest NRE it has, however, the highest re-classification success. In Table 9 we list the three sets with highest NRE together with the percentages of prediction success by the training-testing set (TT) method (where the data set is randomly split into a training set and a testing set 200 times, and the reported percentage is the mean), and the jackknife method where each observation is left out and the prediction of its class is done by the rest of the dataset.

Table 9: Prediction quality of high NRE sets

Symptoms		γ	NRE	TT success	Std. dev.	Jackknife success
Guilt feeling	Excitement	0.920	0.556	91.44	3.78	90.91
Motor retard.	Excitement	0.909	0.553	95.13	2.30	90.91
Depr. Mood	Excitement	1.000	0.534	97.50	2.85	100.00

Mezzich & Solomon [22] note that the symptom “Suspiciousness” seems to differentiate the class PS well from the other three groups because of its high mean value relative to the other diagnostic groups. Rough set analysis shows that the approximation quality of this symptom with respect to the classes “Paranoid schizophrenia” and ”Any of the other three" is 0.920, which suggests the same conclusion; its significance is 0.01.

A similar observation can be made of the symptom “Guilt feelings” with respect to the class MDD where the approximation quality is 0.807, and the total significance is 0.01.

At the end of the paper, the authors report that the clustering methods disagree with the experts for paranoid schizophrenia (about 20% misclassifications) and for mania (about 25% misclassifications), and they observe that

> ”One of the most conspicuous results of this study was the considerable degree of misclassification or overlapping observed between the manic and the paranoid schizophrenic groups of archetypal patients.” (p. 82)

This is not surprising. In their original paper, Overall & Gorman [28] have sampled BPRS ratings of archetypal patients (without the category ‘Excitement’). Their observed ‘scoring weights’ (p. 809) admit only a low differentiation of MDM and PS, with the exception of category 12 (Hallucinatory behavior). This corresponds to the well known fact that, in general, the differential diagnosis of manic and schizophrenic disorders is difficult and requires additionally historic data [see e.g. 34].

Gara et al. [12] perform a hierarchical classes analysis with 133 symptoms for the diagnoses specified in DSM-III-R. The BPRS contains only one (Grandiosity) of the 10 symptoms which make up the manic class, and one can infer that the manic syndrome is insufficiently represented in the BPRS; this may also explain problems in the differential diagnosis of this class.

If we are faced with a choice between MDM and PS alone, then rough information analysis gives us a clear one dimensional criterion which differentiates the two classes: In this situation, ‘Excitement’ can be filtered into two values, *low* (0–3), and *high* (4–6), and we obtain the rule

If 'Excitement' = high, then MDM, otherwise, PS.

4 Further analysis

4.1 Hierarchical Analysis

The same kind of analysis which was applied to the ”Diagnosis Category” can be applied to the attributes ”Depressive Mood” and ”Excitement” in order to find those attributes with which they can be replaced.

The dependent attribute ”Excitement” with the independent attributes 1–16 can be described by 685 reducts. The best reduct in terms of NRE=0.266 consists of the three attributes

"Emotional withdrawal", "Conceptual disorganization', and "Unusual thought content".

Each of the attributes is significant on the 0.1% level, as well as the test of the whole system. Comparing this with ordinary linear regression, we observe that UT has no significant contribution to the regression function, whereas the other two attributes are linear predictors of EX.

Each reduct of ”Depressive Mood” contains at least four features. If we allow some error in prediction (5%), we observe that the combination ”Somatic concern”, ”Conceptual disorganization”, and ”Guilt feelings” give an acceptable rough description of the data ($\gamma = 0.955$). The NRE (0.233) of this combination is higher than the NREs of any reduct of ”Depressive Mood”. Once again: The attributes as well as the test of the whole systems turned out to be significant on the 0.1% level. Comparing this with ordinary linear regression, we observe that CD has no significant contribution to the regression function (which indicates a non-linear relationship between CD and DM), whereas the other two attributes are linear predictors of ”Depressive Mood”.

Merging the results, we obtain the structure of the test as a diagram given in Fig 1. Each arc from attribute $q \in Q$ to an attribute p is labeled with a "Rough strength estimator" $R_{q,p|Q}$ which is defined as follows:

$$R_{q,p|Q} = \gamma(Q \implies p) - \gamma((Q \setminus \{q\}) \implies p).$$

$R_{q,p|Q}$ estimates the percentage of cases which need attribute q for the prediction of the outcome of attribute p, given the set of prediction attributes Q. If $R_{q,p|Q} = 0$, the attribute q is superfluous for the prediction of p.

Consider, for example, the reduct $Q = \{\text{EW, CD, UT}\}$ of EX. We find that

$$R_{\text{CD,EX}|Q} = 1 - \gamma(\{\text{EW,UT}\} \implies \text{EX}) \approx 1 - 0.568 = 0.432.$$

Figure 1: Hierarchical analysis

The leading factor of EX seems to be EW, because 60% of the cases need an EW class. CD seems to be a leading factor for DM, but CD is also necessary for 43% of the cases to describe EX.

About $\frac{3}{4}$ of all cases need the EX symptom to describe the diagnostic class, whereas more than $\frac{1}{2}$ of the sample need the DM attribute.

4.2 Unsupervised reduct analysis

In a further analysis, we have looked for the reducts of the whole system of symptoms without reference to the diagnostic categories. In other words, we have computed those minimal sets of symptoms which give us the same classification as the complete symptom set. From this list, we have chosen the small members and have computed their significance and NRE with respect to the decision attribute. Since the complete set determines the diagnostic category, so does each of its reducts; as its construction does not involve the decision attribute, one might call this method a way of "unsupervised learning".

We have found that the full system has 1022 reducts, which again indicates a huge inhomogeneity among the symptoms, and many possibilities for substitution. The 31 reducts with four symptoms are given in Table 10.

We observe that "Depressive mood" does not appear in any of these reducts; this suggests that DM is actually an aggregated symptom.

The best reduct in terms of NRE consists of "Somatic Concern", "Conceptual Disorganization", "Guilt Feelings" and "Mannerism and Posturing" with an NRE of 0.196.

Table 10: Predictive Power of the results of unsupervised reduct search

Reduct	NRE	Reduct	NRE	Reduct	NRE
1 4 5 7	0.196	1 3 4 16	0.178	3 6 11 16	0.178
1 2 11 17	0.189	1 3 15 16	0.178	4 6 7 11	0.178
1 2 15 17	0.189	1 5 11 16	0.178	4 6 7 13	0.178
2 3 5 6	0.189	1 7 11 15	0.178	6 7 13 15	0.178
5 6 7 15	0.189	1 8 15 16	0.178	6 7 15 16	0.178
5 6 12 14	0.189	2 4 6 12	0.178		
5 6 12 13	0.189	2 4 10 16	0.178		
1 4 5 16	0.183	2 5 6 12	0.178		
1 5 6 12	0.183	2 6 7 15	0.178		
1 5 12 15	0.183	2 6 11 16	0.178		
2 4 5 6	0.183	2 6 12 15	0.178		
4 5 6 7	0.183	2 7 10 16	0.178		
5 6 11 16	0.183	2 10 15 16	0.178		

One problem with this type of unsupervised procedure [which is also used by 22] is the fact that random agreements among attributes can have a strong influence on the structure of the solution. Here, different attitudes of the psychiatrists towards the patients with respect to the attributes HO, UN, TE, or MP may require a differentiation between them, even though these attributes play only a minor role for the classification with respect to diagnostic class. In this case, these "singular" cases have to be classified as well which, as demonstrated, results in a significant drop of NRE.

In the case of cluster analyses – as used by [22] – there are several experiments which show that in this situation the statistical reliability of the cluster may be rather low [23, 24, 25].

5 Discussion

The present study is aiming at different goals. The first goal was to show that a non–invasive data analysis is applicable to data of the prototype approach. We have shown that a sensible analysis of the data can been done with reasonable computational effort.

Another goal was to find a brief and precise description of what the psychiatrists reported about the archetypal psychiatric patients. To our surprise we could show that the data collected in the late 1970s are concordant to current theoretical positions. The features "Depressive mood" and "Excitement" suffice to explain the classification into any of the four categories.

We also showed that the opinions of the experts can be described by an abundance of possible rule based algorithms which rely on only a few features each, and to show that it is possible to discover significant criteria of the experts' judgements. The large number of reducts indicates a huge homogeneity among the symptoms. Our findings suggest that DM is actually an aggregated symptom, best described by "Somatic concern", "Conceptual disorganization", "Guilt feelings".

References

[1] Blashfield, R. (1973). An evaluation of the DSM-II classification of schizophrenia as a nomenclature. *Journal of Abnormal Psychology*, **82**, 382–389.

[2] Browne, C. (1997). Enhanced rough set data analysis of the Pima Indian diabetes data. In *Proc. 8th Ireland Conference on Artificial Intelligence, Derry (1997)*, 32–39.

[3] Cantor, N., Smith, E. E., French, R. S. & Mezzich, J. (1980). Psychiatric diagnoses as prototype categorization. *Journal of Abnormal Psychology*, **89**, 181–193.

[4] Clark, L. A. & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology*, **100**, 316–336.

[5] Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, **45**, 1304–1312.

[6] Dingemans, P. M., Linszen, D. H., Lenior, M. E. & Smeets, R. M. W. (1995). Component structure of the expanded Brief Psychiatric Rating Scale (BPSR-E). *Psychopharmacology*, **122**, 263–267.

[7] Düntsch, I. & Gediga, G. (1995). Rough set dependency analysis in evaluation studies: An application in the study of repeated heart attacks. *Informatics Research Reports*, **10**, 25–30.

[8] Düntsch, I. & Gediga, G. (1997). Statistical evaluation of rough set dependency analysis. *International Journal of Human–Computer Studies*, **46**, 589–604.

[9] Düntsch, I. & Gediga, G. (1998a). Simple data filtering in rough set systems. *International Journal of Approximate Reasoning*, **18**, 93–106.

[10] Düntsch, I. & Gediga, G. (1998b). Uncertainty measures of rough set prediction. *Artificial Intelligence*, **106**, 77–107.

[11] Düntsch, I. & Gediga, G. (2001). Roughian – Rough Information Analysis. *International Journal of Intelligent Systems*, **16**, 121–147.

[12] Gara, M. A., Rosenberg, S. & Goldberg, L. (1992). DSM-III-R as a taxonomy. A cluster analysis of diagnoses and symptoms. *Journal of Nervous and Mental Disease*, **180**, 11–19.

[13] Gigerenzer, G. (1981). Messung und Modellbildung in der Psychologie. Basel: Birkhäuser.

[14] Goldberg, D. P., Bridges, P., Duncan-Jones, P. & Grayson, D. A. (1987). Dimensions of neurosis seen in primary care settings. *Psychological Medicine*, **17**, 461–470.

[15] Goldstein, W. N. & Anthony, R. N. (1988). The diagnosis of depression and the DSMs. *American Journal of Psychotheraphy*, **42**, 180–196.

[16] Kay, S. R. & Sevy, S. (1990). Pyramidal model of schizophrenia. *Schizophrenia Bulletin*, **16**, 537–545.

[17] Lenzenweger, M. F. & Dworkin, R. H. (1996). The dimensions of schizophrenia phenomenology. *British Journal of Psychiatry*, **168**, 432–440.

[18] Lindenmayer, J.-P., Bernstein-Hyman, R., Grochowski, S. & Bark, N. (1995). Psychophatology of schizophrenia: Initial validation of a 5-factor model. *Psychopathology*, **28**, 22–31.

[19] Loranger, A. (1990). The impact of DSM-II on diagnostic practice in a university hospital. *Archives of General Psychology*, **47**, 672–675.

[20] Lorr, M., Klett, C. & McNair, D. (1963). Syndromes of psychosis. London: Pergamon.

[21] Maß, R., Burmeister, J. & Kraus, M. (1997). Dimensionale Struktur der deutschen Version der Brief Psychiatric Rating Scale (BPRS). *Nervenarzt*, **68**, 239–244.

[22] Mezzich, J. E. & Solomon, H. (1980). Taxonomy and Behavioral Science. London: Academic Press.

[23] Milligan, G. W. (1980). An examination of the effect of six type of error pertubation on fifteen clustering algorithms. *Psychometrika*, **45**, 325–342.

[24] Milligan, G. W. (1981). A monte carlo study of thirty internal criterion measures for cluster analysis. *Psychometrika*, **46**, 187–199.

[25] Milligan, G. W. (1989). A study of the beta-flexible clustering method. *Multivariate Behavioral Research*, **24**, 163–176.

[26] Ormel, J., Oldehinkel, A. J., Goldberg, D. P., Hodiamont, P. P. G., Wilmink, F. W. & Bridges, K. (1995). The structure of common psychiatric symptoms: How many dimension of neurosis? *Psychological Medicine*, **25**, 521–530.

[27] Osgood, C. E., Suci, G. J. & Tannenbaum, P. H. (1957). The measurement of meaning. Urbana: University of Illinois Press.

[28] Overall, J. E. & Gorman, D. R. (1962). The Brief Psychiatric Rating Scale. *Psychological Reports*, **10**, 799–812.

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

[30] Polkowski, L. & Skowron, A. (Eds.) (1998). Rough sets in knowledge discovery, Vol. 2. Heidelberg: Physica–Verlag.

[31] Rissanen, J. (1978). Modeling by the shortest data description. *Automatica*, **14**, 465–471.

[32] Słowinski, K. & Stefanowski, J. (1998). Multistage rough set analysis of therapetic experience with acute pancreatitis. In [30], 272–294.

[33] Tanaka, H., Koyama, K. & Maeda, Y. (1996). Reduction method for divisions of attributes and its application to medical data. In S. Tsumoto, S. Kobayashi, T. Yokomori, H. Tanaka & A. Nakamura (Eds.), *The fourth International Workshop on Rough Sets, Fuzzy Sets, and Machine Discovery,* RS96FD*)*, 61–54.

[34] Tölle, R. (1988). Psychiatrie. Berlin: Springer.

[35] van der Doers, A. J. W., Linszen, D. H., Dingemans, P. M., Nugter, M. A. & Scholte, W. F. (1993). A dimensional and categorial approach to the symptomatology of recent-onset schizophrenia. *Journal of Nervous and Mental Disease*, **181**, 744–749.

[36] Ventura, J., Green, M. F., Shaner, A. & Liberman, R. P. (1993). Training and quality assurance with the Brief Psychiatric Rating Scale. *International Journal of Methods in Psychiatric Research*, **2**, 221–244.

[37] Woolery, L. K. & Grzymala-Busse, J. (1994). Machine learning for an expert system to predict preterm birth risk. *Journal of the American Medical Informatics Association,* **1**, 439–446.