Rough set data analysis

1.0

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

Guenther@Gediga.de

June 28, 1999

1 Introduction

Rough set data analysis (RSDA), developed by Z. Pawlak and his co–workers in the early 1980s [47] has become a recognised and widely researched method with over 1100 publication to date [54, Appendix 1]. The conceptual foundation of the RSDA model is the consideration that all perception is subject to granularity, and that the ability to classify is at the root of human intelligence:

> "Our claim is that knowledge is deep–seated in the classificatory abilities of human beings and other species. For example, knowledge about the environment is primarily manifested as an ability to classify a variety of situations from the point of view of survival in the real world . . . Classification on more abstract levels, seems to be a key issue in reasoning, learning and decision making, not to mention that in science classification it is of primary importance too." [48]

The main thrust in current applications of rough set theory are

- Attribute reduction,

- Rule generation,

- Prediction.

---

*The ordering of authors is alphabetical, and equal authorship is implied.

Table 1: The position of RSDA in Soft Computing [35]

	Microscopic, primarily numeric	Macroscopic, descriptive and numeric
Deductive	Chaos theory	Fuzzy methods
Inductive	Neural networks, genetic algorithms	RSDA

Many examples of applications of RSDA to process control, economics, medical diagnosis, biochemistry, environmental science, biology, chemistry psychology, conflict analysis and other fields can be found in [30, 54, 69]. For further information of developments up to 1999, we recommend the list given in [25].

RSDA is generally regarded as part of the "Soft Computing" paradigm, and in a recent book [35], it is treated as one of the five key "non traditional AI areas" (Table 1). However, while other soft methods require additional model assumptions such as the representability of the collected sample, prior probabilities, fuzzy functions, or degrees of belief, RSDA is unique in the sense that it is "non – invasive" i.e. that it uses only the information given by the operationalised data, and does not rely on other model assumptions. In other words, instead of using external numbers or other additional parameters, rough set analysis utilises solely the structure of the given data:

> "The numerical value of imprecision is not pre-assumed, as it is in probability theory of fuzzy sets - but is calculated on the basis of approximations which are the fundamental concepts used to express imprecision of knowledge ... As a result we do not require that an agent assigns precise numerical values to express imprecision of his knowledge, but instead imprecision is expressed by quantitative concepts (approximations). [48]

In this sense, RSDA, as a measurement process, can be regarded as "nominal scaling" [74]. In order to position RSDA in the data modelling process, we follow [19] in assuming that a data model consists of

1. The "acting subject".

2. A domain $\mathcal{D}$ of interest chosen by the agent.

3. An empirical system $\mathcal{E}$, which consists of a body of data and relations among the data, and a mapping $e : \mathcal{D} \to \mathcal{E}$, called *operationalisation*.

4. A numerical model $\mathcal{M}$, and a mapping $m : \mathcal{E} \to \mathcal{M}$, called *representation*.

As we can see in Fig. 1, the acting subject with her/his objectives is the central part of the modelling process, and the first, and, possibly, major factor of subjectivity: Agents choose an operationalisation according to their objectives and their subjective view of the world. This is a fact which is encountered by any method of data analysis: Operationalisation is necessary to be able to talk about a domain, but it

Figure 1: The data modelling process
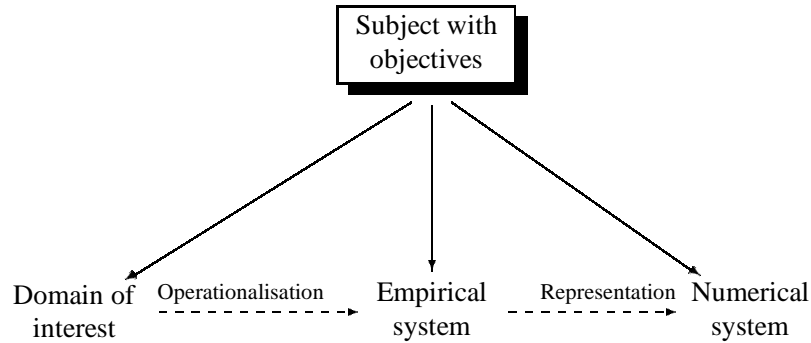
raises two questions; first, whether the elements and relations of the result of the operationalisation, i.e. the empirical model $\mathcal{E}$, are representative for the domain $\mathcal{D}$ (consistency problem), another whether the choice of features covers the relevant aspects of $\mathcal{D}$ (completeness problem).

The next step, the numerical models, are a reduction of the empirical model, and thus of the domain of interest; going from the empirical to the numerical model, often results in further decontextualisation. We invite the reader to consult [83, Section 2] for further comments on the relationship of RSDA to numerical models.

RSDA takes place on the level of the empirical model, and the main motto of RSDA is

(1.1)   LET THE DATA SPEAK FOR THEMSELVES.

An operationalisation of the type

$$\text{OBJECT} \mapsto \text{ATTRIBUTE(S)}$$

relationship in form of a data table is supposed to be given (with all its underlying assumptions), and subsequent analysis uses only the information contained in this table. This is in contrast to most statistical (or other) analysis methods. Even the bootstrap technique, which is discussed in the rough set context in [75] needs additional model assumptions, because one has to suppose that the percentages of the observed equivalence classes are representative estimators of the latent probabilities of the equivalence classes in the population.

We shall see below that dependencies and decision rules are always extracted (or learned) from existing systems, and they are not part of the design process as is the case, for example, with relational databases.

Since the field of RSDA has grown rapidly within the past five years, and there are many ramifications and generalisations, it is impossible to describe in detail every thread in the area, and we will have to be content to outline the major issues and developments.

The paper is structured as follows: We will start by describing the basic RSDA model, its mathematical foundations, and its knowledge representation in some detail. We will then describe how RSDA tackles major issues of data analysis:

3

- Attribute reduction and rule generation,

- Significance testing and rule validation,

- Model selection,

- Data discretisation, including grouping of qualitative values.

This is followed by a brief description of further development in RSDA. We conclude with a summary and an outlook to further development.

2   The basic model of RSDA

The mathematical machinery of RSDA is derived from the assumption that granularity can be expressed by partitions and their associated equivalence relations on the set of objects, also called *indiscernability relations* in this context.

Suppose that U is a nonempty set. A *partition* or *classification* of U is a family $\mathcal{P}$ of nonempty subsets of U such that

(2.1)   Each element of U is contained in exactly one element of $\mathcal{P}$.

Recall that an *equivalence relation* θ on a set U is a reflexive, symmetric, and transitive binary relation on U, i.e. for all $x, y, z \in U$, we have

$$x\theta x, \qquad\qquad\qquad\qquad \text{Reflexivity}$$
$$x\theta y \text{ implies } y\theta x, \qquad\qquad\qquad \text{Symmetry}$$
$$x\theta y \text{ and } y\theta z \text{ imply } x\theta z. \qquad\qquad \text{Transitivity}$$

Each partition $\mathcal{P}$ induces an equivalence relation θ on U by setting

(2.2) $x\theta y \iff x$ and y are in the same class of $\mathcal{P}$.

Conversely, each equivalence relation θ on U induces a partition $\mathcal{P}$ of U whose classes have the form

$$\theta x = \{y \in U : x\theta y\}.$$

By some abuse of language, we also speak of the classes of an equivalence relation when we mean the classes of its associated partition, and call θx *the class of x modulo θ*. The interpretation in rough set theory is that our knowledge of the objects in U extends only up to membership in the classes of θ, and our knowledge about a subset X of U is limited to the classes of θ and their unions. This leads to the following definition: For $X \subseteq U$, we say that

(2.3) $\underline{X} \overset{\text{def}}{=} \bigcup \{\theta x : \theta x \subseteq X\}$

Figure 2: Rough approximation
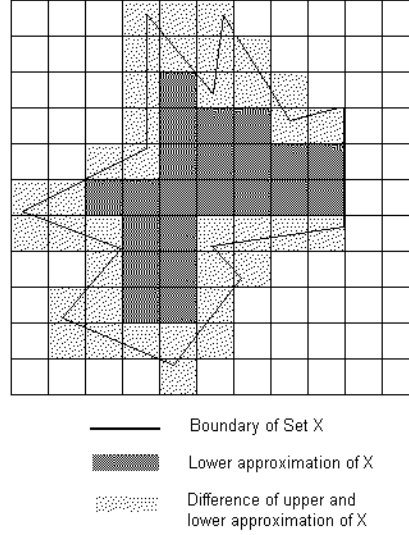
is the *lower approximation* or *positive region of X*, and

(2.4) $$\overline{X} \stackrel{\text{def}}{=} \bigcup \{\theta x : x \in X\}$$

is the *upper approximation* or *possible region* of X.

If $X \subseteq U$ is given by a predicate P and $x \in U$, then

1. $x \in \underline{X}$ means that x *certainly* has property P,

2. $x \in \overline{X}$ means that x *possibly* has property P,

3. $x \in U \setminus \overline{X}$ means that x definitely *does not have* property P.

The *area of uncertainty* extends over

$$\overline{X} \setminus \underline{X},$$

and the *area of certainty* is

$$\underline{X} \cup -\overline{X}.$$

A *rough subset* of U is a pair $\langle \underline{X}, \overline{X} \rangle$, where $X \subseteq U$. A subset X of U is called *definable* if $\underline{X} = \overline{X}$. In this case, X is empty or a union of equivalence classes of θ, and the area of uncertainty is $\emptyset$. An equivalent – and, possibly, more transparent – way of describing the information given by $\langle U, \theta \rangle$ is to consider pairs $\langle \underline{X}, -\underline{X} \rangle$; this gives the area of positive certainty in the first component, and that of negative certainty in the second one, and it is closer to the machine learning situation of positive and negative examples. It also avoids certain statistical difficulties which arise when using the upper approximation.

It has been shown that the collection of all rough subsets of U can be made into a regular double Stone algebra [6, 10, 24, 56], and that these algebras can serve as semantic models for three valued

Table 2: Fisher's Iris data [18]

Object	Sepal length	Sepal width	Petal length	Petal width	Class
s_1	50	33	14	2	1
s_2	46	34	14	3	1
s_3	65	28	46	15	2
s_4	62	22	45	15	2
s_6	67	30	50	17	3
s_7	64	28	56	22	3
<143 other specimen>					

Łukasiewicz logic [9]. The connections between algebra and logic of rough set systems are explored in some detail in [45]. A different logical approach to the rough set model is given by [41] and [31]: There, the lower approximation is considered a necessity operator $\Box$, and the upper approximation a possibility operator $\Diamond$.

It may be also worthy of mention that RSDA can be interpreted in Shafer's evidence theory [60] in such a way that beliefs are obtained internally from the lower approximation of a set, and plausibility from its upper approximation [61, 63].

3 Knowledge representation

Knowledge representation in RSDA is done via information systems, which are a form of data table. More precisely, 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,

3. For each $q \in \Omega$

   - A set V_q of attribute values,
   - An information function $f_q : U \to V_q$.

Table 2 which shows part of the famous Iris data [18] is an example of such a system: U consists of 150 specimen of Iris flowers; there are five attributes, namely, *sepal length, sepal width, petal length, petal width*, and an attribute *class*. The sets V_q of attribute values for the first four attributes consist of lengths, measured in millimetres, and the attribute *class* takes its values from the set $\{1, 2, 3\}$, which

code the species *Setosa, Versicolor, Virginica*. We think of the descriptor $f_q(x)$ as the value which object x takes at attribute q; thus, for example,

$$f_{\text{sepal length}}(s_1) = 50, \ f_{\text{petal length}}(s_2) = 14, \ f_{\text{class}}(s_3) = 2.$$

What we have above is, in fact, a *decision system*: Besides four independent features, there is a dependent decision attribute *class*.

We observe that in information systems, each value $f_q(x)$ is unique, and there can be no missing values. A generalisation of the information system above are the *multi–valued information systems*, introduced in [44]. There, the information functions take on set values, i.e. for each $q \in \Omega$ and $x \in U$,

$$f_q(x) \subseteq V_q.$$

These can cater for non–deterministic information and null values. We refer the reader to [43] for a more detailed description of multi–valued information systems and their algebraic and logical properties.

4 Attribute reduction and rule generation

The discovery of data dependencies and the reduction of the number of attributes is one of the major topics in data mining. As we shall see below, RSDA offers purely structural methods to achieve both.

In the sequel, we let $\mathcal{I} = \langle U, \Omega, V_q, f_q \rangle_{q \in \Omega}$ be a generic information system. As a running example, we shall use the decision system of credit card applications shown in Table 3. We let $\Omega = \{c_1, c_2, c_3, c_4\}$, and consider the decision attribute separately.

Table 3: Credit card applications

Applicant	Condition attributes				d
	c_1 Account	c_2 Balance	c_3 Employed	c_4 Monthly outgoing	Decision
a_1	bank	medium	yes	low	accept
a_2	bank	low	yes	high	reject
a_3	none	low	yes	medium	reject
a_4	other	high	yes	high	accept
a_5	other	medium	yes	high	reject
a_6	other	high	yes	low	accept
a_7	bank	high	no	medium	accept
a_8	none	low	no	low	reject

Attribute reduction in RSDA is achieved by comparing equivalence relations generated by sets of attributes. With each $Q \subseteq \Omega$ we associate an equivalence relation θ_Q on U by

$$x \, \theta_Q \, y \text{ if and only if } f_q(x) = f_q(y) \text{ for all } q \in Q.$$

Intuitively, $x\theta_Q y$ if the objects x and y are indiscernible with respect to the values of their attributes from Q. It is not hard to see that for $P, Q \subseteq \Omega$

$$P \subseteq Q \text{ implies } \theta_Q \subseteq \theta_P.$$

Thus, the finest equivalence relation obtained this way is θ_Ω, and $\theta_\emptyset$ is the universal relation $U \times U$.

The partition of θ_Ω in our system of Table 3 (without the decision attribute d) is the identity relation; all eight applicants are pairwise distinguishable by the given attributes.

We use these equivalence relations to express dependency among attribute sets in the following way: Suppose that $P, Q \subseteq \Omega$; P is called *dependent on Q* – written as $Q \Rightarrow P$ – if $\theta_Q \subseteq \theta_P$. In this case, every class of θ_P is a union of classes of θ_Q, so that membership in a class of θ_Q determines membership in a class of θ_P. In other words, the partition of U induced by P can be expressed by the partition induced by Q, so that the world according to P is coarser than the world according to Q. If $P = \{p\}$, we usually just write $Q \Rightarrow p$.

A set $P \subseteq Q \subseteq \Omega$ is called a *reduct of Q*, if

1. $\theta_P = \theta_Q$,

2. For each $R \subsetneq P$ we have $\theta_R \neq \theta_Q$.

In other words, $P \subseteq Q$ is a reduct of Q, if P is minimal among all subsets of Q which generate the same classification as Q; the attributes within a reduct are independent, and none of them can be omitted for the description of Q. Reducts produce deterministic rules. It is not hard to see, that each $Q \subseteq \Omega$ has a reduct, though this is usually not unique. The intersection of all reducts of Q is called the *core of Q*, written as $\mathsf{core}(Q)$, and the elements of the core of Q are called *indispensable* for Q. If $q \in Q$, and q is not an element of any reduct of Q, then we call q *redundant for Q*. If $Q = \Omega$ we just speak of *reducts of $\mathcal{I}$, core of $\mathcal{I}$* etc. The set of all reducts of $\mathcal{I}$ is denoted by $Red(\mathcal{I})$. Reducts correspond to keys of a relational database; consequently, as was pointed out in [57] the problem of finding a reduct of minimal cardinality is, in general, NP-hard, and finding all reducts has exponential complexity [66].

A transparent method to find the core and the reducts of an information system $\mathcal{I}$ via discernibility matrices was given in [66]: Define a *discernibility function* $\delta : U \times U \to 2^\Omega$ by

$$(4.1) \qquad \delta(a, b) \overset{\text{def}}{=} \{q \in \Omega : f_q(a) \neq f_q(b)\}.$$

The function δ leads to a pseudo-metric, since

$$(4.2) \qquad a = b \text{ implies } |\delta(a, b)| = 0.$$

$$(4.3) \qquad |\delta(a, b)| = |\delta(b, a)|.$$

$$(4.4) \qquad |\delta(a, c)| \leq |\delta(a, b)| + |\delta(b, c)|.$$

If θ_Ω is the identity, then the converse of (4.2) holds as well, and the assignment $\langle a, b \rangle \mapsto |\delta(a, b)|$ is a metric.

Now,

Proposition 4.1. *[66]*

1. *The core of $\mathcal{I}$ is the set*

$$\{q \in \Omega : \delta(a,b) = \{q\} \text{ for some } a,b \in U\}.$$

2. *$P \subseteq \Omega$ is a reduct of $\mathcal{I}$ if P is minimal with respect to the property*

$$P \cap \delta(a,b) \neq \emptyset$$

for all $a,b \in \Omega$, $\delta(a,b) \neq \emptyset$.

The discernibility matrix of the system of Table 3 without the decision attribute is given in Table 4.

Table 4: Discernibility matrix

	1	2	3	4	5	6	7	8
1	$--$	c_2, c_4	c_1, c_2, c_4	c_1, c_2, c_4	c_1, c_4	c_1, c_2	c_2, c_3, c_4	c_1, c_2, c_3
2	$--$	$--$	c_1, c_4	c_1, c_2	c_1, c_2	c_1, c_2, c_4	c_2, c_3, c_4	c_1, c_3, c_4
3	$--$	$--$	$--$	c_1, c_2, c_4	c_1, c_2, c_4	c_1, c_2, c_4	c_1, c_2, c_3	c_3, c_4
4	$--$	$--$	$--$	$--$	c_2	c_4	c_3, c_4	c_1, c_2, c_3, c_4
5	$--$	$--$	$--$	$--$	$--$	c_2, c_4	c_1, c_2, c_3, c_4	c_1, c_2, c_3, c_4
6	$--$	$--$	$--$	$--$	$--$	$--$	c_1, c_3, c_4	c_1, c_2, c_3
7	$--$	$--$	$--$	$--$	$--$	$--$	$--$	c_1, c_2, c_4

The entries $\langle 4, 5 \rangle$, $\langle 4, 6 \rangle$ show that the core of the system I is $\{c_2, c_4\}$, and $\langle 5, 6 \rangle$ shows that this the also the only reduct of $\mathcal{I}$.

The relations among the classes of the equivalence relations associated with attribute sets can be used to generate decision rules; for notational simplicity, we suppose that we have a set $Q = \{q_1, \ldots, q_n\}$ of independent attributes, and a single dependent attribute d; this is no restriction of generality, since we are using only the partition information of θ_d, and thus, d can be a "composite attribute", obtained from some $P \subseteq \Omega$.

We assume that the partition induced by θ_Q is $\{X_1, \ldots, X_s\}$, and the one induced by θ_d is $\{Y_1, \ldots, Y_t\}$. With each X_i we associate the set $M_i = \{Y_j : X_i \cap Y_j \neq \emptyset\}$. Since the sets $Y_1, \ldots, Y_t$ partition U, we see that

(4.5) $\qquad\qquad$ If $x \in X_i$, then $x \in Y_{j_1}$ or $\ldots$ or $x \in Y_{j_{i(j)}}$,

where $M_i = \{Y_{j_1}, \ldots Y_{j_{i(j)}}\}$. Now recall that each class X_i of θ_Q corresponds to a feature vector $(a_j)_{1 \leq i \leq n}$, where $x \in X_i$ if and only if $f_{q_1}(x) = a_1$ and $\ldots$ and $f_{q_n}(x) = a_n$; similarly, $x \in Y_j$ if

9

and only if $f_d(x) = b_j$ for some $b_j \in V_d$. If we translate condition (4.5), this leads to a rule of the form

(4.6)   If $f_{q_1}(x) = a_1$ and ... and $f_{q_n}(x) = a_n$, then $f_d(x) = b_{j_1}$ or ... or $f_d(x) = b_{j_{i(j)}}$.

It follows that RSDA rules are able to express deterministic as well as indeterministic information: If some class X_i of θ_Q intersects exactly one Y_j, then $X_i \subseteq Y_j$, and the value of d of any $x \in X_j$ is uniquely determined; otherwise, $f_d(x)$ may be in any class contained in M_i, and we have a proper disjunction on the right hand side of (4.6). A class X_j is called *deterministic*, if it is contained in some Y_j, otherwise, we call it *indeterministic*. If all classes X_j are deterministic, then, clearly, $\theta_Q \subseteq \theta_d$, and d is dependent on Q. We will write $Q \to d$ for the collection of rules (4.5), and, with some abuse of language call $Q \to d$ a rule of the information system $\mathcal{I}$.

Let us consider again our example of Table 3. If $Q = \{c_2\}$ and d is the decision attribute, we have

$$X_1 = \{a_1, a_5\},\ X_2 = \{a_2, a_3, a_8\},\ X_3 = \{a_4, a_6, a_7\}$$
$$Y_1 = \{a_1, a_4, a_6, a_7\},\ Y_2 = \{a_2, a_3, a_5, a_8\}.$$

X_2 and X_3 are deterministic classes, while X_1 is not, and we obtain the following decision rules with respect to d:

(4.7)   If bank balance is medium, then accept or reject.

(4.8)   If bank balance is low, then reject.

(4.9)   If bank balance is high, then accept.

By taking into account attribute c_4 we can split X_1 in θ_{c_2,c_4}, and refine rule (4.7) by

(4.10)   If bank balance is medium and monthly expense is low then accept, otherwise, reject.

By using c_1 and c_2, we obtain another rule system:

(4.11)   If bank balance is medium and account is bank then accept, otherwise, reject.

We refer the reader to [21, 22, 62, 71, 85] for various methods to generate RSDA rules.

5   Approximation measures

Even though RSDA is a symbolic method of analysis, it uses counting information provided by the classes of the equivalence relations under consideration. The inherent statistics of an approximation space $\langle U, \theta \rangle$ are the two *approximation functions*

(5.1) $$\gamma_\theta(X) \stackrel{\text{def}}{=} \frac{|\underline{X}| + |\underline{-X}|}{|U|},$$

(5.2) $$\alpha_\theta(X) \stackrel{\text{def}}{=} \frac{|\underline{X}|}{|\overline{X}|}\ \text{(for } X \neq \emptyset),$$

see [48], p. 16ff. If θ is understood, we shall usually omit the subscripts.

$\gamma_\theta(X)$ is the percentage of objects of U which can be correctly classified with the knowledge given by θ as being in X or not, while $\alpha_\theta(X)$ expresses the degree of completeness of our knowledge of X. It was shown in [15] that the γ approximation is a manifestation of the underlying statistical principle of RSDA, namely, the *principle of indifference*: Within each equivalence class, the elements are assumed to be randomly distributed.

We can generalise the approximation functions to partitions with more than two classes in the following way: As a measure of the *quality of an approximation* of a partition $\mathcal{P}$ by a set Q of attributes we define $\gamma(Q \to \mathcal{P})$ by

$$(5.3) \qquad \gamma(Q \to \mathcal{P}) = \frac{\sum_{X \in \mathcal{P}} |\underline{X}_{\theta_Q}|}{|U|},$$

In case $\mathcal{P}$ is induced by θ_P for some $P \subseteq \Omega$, we will write $\gamma(Q \to P)$ instead of $\gamma(Q \to \mathcal{P})$. If $X \in \mathcal{P}$, then $\underline{X}_{\theta_Q}$ is the set of all elements of X that are correctly classified with respect to the attributes in Q, and $\gamma(Q \to \mathcal{P})$ is the ratio of the number of all certainly $\theta_{\mathcal{P}}$ – classified elements of U with the information provided by the attributes in Q to the total number of elements of U. Note that $Q \Rightarrow P$ implies $\gamma(Q \to P) = 1$.

Suppose that for $Q, P \subseteq \Omega$, Q is a reduct of P, so that $Q \Rightarrow P$, and $Q \setminus \{q\} \not\Rightarrow P$ for any $q \in Q$. In rough set theory, the impact of attribute q on the fact that $Q \Rightarrow P$ is usually measured by the drop of the approximation function γ from 1 to $\gamma(Q \setminus \{q\} \to P)$: The larger the difference, the more important one regards the contribution of q to the rule [48, p.58].

While the approximation quality γ measures the global classification success in terms of the equivalence classes, one can use the same principle for elements and define *rough membership functions* [49]: For each $X \subseteq U$, let $\mu_X : U \to [0,1]$ be a function defined by

$$(5.4) \qquad \mu_X(x) = \frac{|\theta x \cap X|}{|\theta x|}.$$

It is easy to see that for all $X, Y \subseteq U$,

$$(5.5) \qquad \mu_X(x) = \begin{cases} 1, & \text{iff } x \in \underline{X}, \\ 0, & \text{iff } x \notin \overline{X}, \end{cases}$$

$$(5.6) \qquad \mu_{U \setminus X}(x) = 1 - \mu_X(x),$$

$$(5.7) \qquad \mu_{X \cup Y}(x) \geq \max(\mu_X(x), \mu_Y(x)),$$

$$(5.8) \qquad \mu_{X \cap Y}(x) \leq \min(\mu_X(x), \mu_Y(x)).$$

The cases, where equality holds in (5.7) and (5.8) - and when, consequently, μ_X is a fuzzy membership function, as well as an efficient algorithm to compute rough membership functions are given in [49]. Unlike fuzzy membership, the μ_X values are obtained from the internal structure of the data, and no outside information is needed.

Rough sets and fuzzy sets are geared towards different situations: While at the heart of RSDA is the concept of granularity, mathematically expressed by classes with crisp boundaries, within which no information is available, fuzzy sets describe vagueness of a concept where boundaries among classes are ill–defined. Frameworks to combine the two points of view have been proposed, among others, in [7, 70, 82]. Hybrid systems, where RSDA is used as a pre-processing device for fuzzy methods have been given, for example, in [50, 72]

6 Rule significance and validation

In the early development of RSDA, the static point of view was predominant: Reducts which were obtained from one sample of data were studied, along with the contribution of single attributes, measured by the drop of the γ approximation quality, e.g. [68, 73]. Over time, it became clear, however, that for prediction, i.e. the classification of new elements, these tools were not well suited and other measures had to be found.

> "The rough set methods developed so far are not always sufficient for extracting laws from decision tables. The set of all decision rules generated from all conditional attributes can be too large and/or can contain many chaotic rules not appropriate for unseen object classification" [1].

Indeed, a re-analysis of the results of [68] and [73] showed that not all the claimed rules were useful, and that some potentially valuable rules were overlooked [12].

Any rule based system faces the problems of *rule significance* and *rule validity*. Whereas significance tests measure the probability that a rule is due to chance, i.e. random behaviour, validation of a rule system investigates, how well the particular hypothesis, expressed by the rule, is replicable. Clearly, both problems need to be addressed before RSDA can claim to be a fully fledged instrument for data analysis and prediction. Of particular interest are those significance testing and validation procedures which stay within the non–invasive RSDA paradigm of (1.1), and do not rely on extra model assumptions.

One might suppose that the γ approximation quality is a suitable tool, since it measures classification success; however, it has been shown in [12], that a high approximation quality is not a guarantee that the result of a rough set analysis is either significant or valid. If, for example, rough set analysis discovers a rule of the form $Q \Rightarrow d$, and if the rule is based on only a few observations, classification may be due to chance. A similar observation can be made with regard to the influence of one attribute in a set of independent attributes.

A solution of the significance problem was given in [12] on the basis of randomisation procedures. These are particularly appropriate to RSDA as a non–invasive method, since they do not require any representability assumption and use only the information of the given sample [17, 33]. A special case of the procedures developed in [12] is the investigation of the contribution of single attributes on a

rule $Q \rightarrow d$; an attribute $q \in Q$ is called *conditionally casual*, if, loosely speaking, there are only a few observations in which q is needed to predict d. Removing conditionally casual attributes reduces the overfitting of data which is common in the traditional RSDA approach based on γ.

Significance of a rule is a minimal but necessary requirement for evaluating the predictive power of a rule $Q \rightarrow d$. To test the prediction quality of a rule, additional validation is needed. Rule validation in RSDA is usually performed by customary cross–validation methods. It can be seen, however, that the jackknife validation ("leave-one-out") may result in an overly optimistic estimation of prediction success [3]. Another problem with this method, particular to RSDA, is the case when the element which is left out can only be predicted by its own rule. In this case, invasive techniques such as the adoption of metric information (Hamming distance, Euclidean distance etc.) need to be applied, which assume more than the data itself tell us, and are outside the RSDA paradigm (1.1).

A non-invasive validation technique is based on a significance test [12]. The reasoning behind the method is the assumption that prediction success should not depend on the way the data set is split up into a training and a testing set: A new attribute SPLIT is added and randomly distributed over half of the data set. If SPLIT is α significant in more than $\alpha \cdot 100\%$ of the simulated data sets, then the split is not random, and the prediction success is dependent on the choice of training and testing set. It follows that the rule cannot be validated.

7 Model selection

Attribute reduction is an inherent feature of RSDA, and usually has no unique solution. If several attribute sets exist from which prediction is possible, then the problem arises which of these are most suitable by some predefined criterion. We will present two very different recent methods of model selection. The first one is selection by *dynamic reducts* [1]:

> "The underlying idea of dynamic reducts stems from the observation that reducts generated from information systems are unstable in the sense that they are sensitive to changes in the information system introduced by removing a randomly chosen set of objects. The notion of dynamic reduct encompasses the stable reducts, i.e. reducts that are the most frequent reducts in random samples created by subtables of the given decision table" [65].

For $\mathcal{I} = \langle U, \Omega, V_q, f_q, d \rangle_{q \in \Omega}$ and $U' \subseteq U'$, we call $\mathcal{I}' = \langle U', \Omega, V_q, f_q, d \rangle_{q \in \Omega}$ a *subsystem of $\mathcal{I}$*. If $\mathbf{F}$ is a family of subsystems of $\mathcal{I}$, then

(7.1) $$DRed(I, \mathbf{F}) = Red(\mathcal{I}) \cap \bigcap \{\mathcal{J} : \mathcal{J} \in \mathbf{F}\}$$

is called the *family of $\mathbf{F}$-dynamic reducts of $\mathcal{I}$*. It is easily seen that in most cases, this is too restrictive for practical use, and thus, a threshold is introduced: Let $0 \leq \epsilon \leq 1$. The *family of $(\epsilon, \mathbf{F})$-dynamic reducts of $\mathcal{I}$* is defined by

(7.2) $$D_\epsilon Red(I, \mathbf{F}) = \{Q \in Red(\mathcal{I}) : 1 - \epsilon \leq s_{\mathbf{F}}(Q)\}$$

13

where

(7.3)
$$s_{\mathbf{F}}(Q) = \frac{|\{\mathcal{J} \in \mathbf{F} : Q \in Red(\mathcal{J})\}|}{|\mathbf{F}|}$$

is the $\mathbf{F}$- *stability coefficient of* Q. Model selection proceeds in four steps:

1. Choose numbers n, k_j, $j \le n$, $1 \le k_j \le \frac{|U|}{2}$, and a threshold ϵ.

2. For each $1 \le j \le n$ generate a subsystem $\mathcal{I}_j$ of $\mathcal{I}$ by randomly deleting a k_j objects of U, and set $\mathbf{F} = \{\mathcal{I}_j : 1 \le j \le n\}$.

3. Find the reducts for $\mathcal{I}$ and each $\mathcal{I}_j$.

4. Choose all reducts Q with $1 - \epsilon \le s_{\mathbf{F}}(Q)$ for further processing.

From these "true dynamic reducts", decision rules for classification are computed, and the final decision is taken by "majority voting".

The method of dynamic reducts employs a kind of internal cross-validation in order to improve the external prediction quality. We observe that the researcher has to make some subjective choices in step 1. of the procedure, which are not contained in the data. The huge complexity of step 3. forces applications of heuristic techniques, such as combinatorial or genetic algorithms. Extensive experiments reported in [2] show that the dynamic reduct approach fares considerably better than the traditional RSDA method and compares well with customary procedures. For extensions of the method and similar approaches we refer the reader to [2, 40, 65].

A different approach, reported in [14], does not rely on reducts at all – thus leaving the traditional RSDA way of reducing feature complexity – and uses information theoretic techniques and the minimum description length principle [58, 59]. The model selection criterion is an entropy value $H(Q \to d)$ which aggregates

- The complexity of coding the hypothesis Q, measured by the entropy $H(Q)$ of the partition of its associated equivalence relation θ_Q and

- The conditional coding complexity $H(d|Q)$ of d, given by the values of attributes in Q,

so that

(7.4)
$$H(Q \to d) = H(Q) + H(d|Q).$$

The value $H(Q \to d)$ is obtained in the following way: Suppose that θ_Q generates the partition of U with classes M_i, $i \le k$, each having cardinality m_i. In compliance with the principle of indifference – the statistical basis of RSDA – we suppose that the elements of U are randomly distributed within the classes, so that the probability of an element x being in class M_i is just $\frac{m_i}{|U|}$. This leads to a probability distribution $\{\hat{\pi}_i : i \le k\}$ defined by

$$\hat{\pi}_i = \frac{m_i}{|U|}.$$

Suppose that the union V of deterministic classes of $Q \to d$ is not empty, and that the deterministic classes are $M_1, \dots, M_c$. The underlying idea is the consideration that each observation $y \in U \setminus V$ is the result of a random process whose characteristics are unknown; therefore, each such y should be viewed as a realization of an unknown probability distribution with uncertainty $\frac{1}{n} \log_2(|U|)$. This observed probability distribution is given by $\{\hat{\psi}_i : i \leq c + |U \setminus V|\}$ with

(7.5)
$$\hat{\psi}_i \stackrel{\text{def}}{=} \begin{cases} \hat{\pi}_i, & \text{if } i \leq c, \\ \frac{1}{|U|}, & \text{otherwise.} \end{cases}$$

We now define the *entropy of rough prediction* (with respect to $Q \to d$) as

$$H(Q \to d) \sum_i \hat{\psi}_i \cdot \log_2(\frac{1}{\hat{\psi}_i})$$

and have

$$H(Q \to d) = \sum_{i \leq c} \hat{\pi}_i \cdot \log_2(\frac{1}{\hat{\pi}_i}) + |U \setminus V| \cdot \frac{1}{|U|} \cdot \log_2(|U|)$$

$$= \underbrace{\sum_{i \leq c} \hat{\pi}_i \cdot \log_2(\frac{1}{\hat{\pi}_i})}_{\text{Knowledge}} + \underbrace{(1 - \gamma) \cdot \log_2(|U|)}_{\text{Guessing}} .$$

where $\gamma = \gamma(Q \to d)$. This gives us

$$H(d|Q) = H(Q \to d) - H(Q)$$

$$= (1 - \gamma) \cdot \log_2(|U|) - \sum_{i > c} \hat{\pi}_i \cdot \log_2(\frac{1}{\hat{\pi}_i}).$$

The importance of $H(Q \to d)$ is due to the fact that it aggregates the uncertainty $H(d|Q)$ and the effort $H(Q)$ of coding the hypothesis, i.e. the predicting elements. Thus, one can compare different attribute sets Q_i in terms of a common unit of measurement, which cannot be done by a conditional measure of prediction success such as γ approximation or dynamic reducts. Furthermore, there are no additional parameters required, so that the method is well in accord with the RSDA basic principle (1.1). The model selection (and attribute reduction) process SORES (**S**earching **O**ptimal **R**ough **E**ntropy **S**ets) compares well with standard numerical methods, without using invasive model assumptions. Detailed results can be found in [14] and on the website `http://www.psycho.uni-osnabrueck.de/sores/`.

It is worthy of mention that these methods are universal, and they take place before ad hoc tuning for specific situation can be applied.

8 Data discretisation

Attributes with continuous values pose a severe problem for RSDA, indeed, for any rule based data analysis method: If there are many values, the equivalence classes belonging to the attribute set in

15

question may have many small classes, and a rule will be based on a few instances only. Therefore, its significance will be low, and its ability to classify new objects is rather limited. There are many methods to discretise a continuous domain, most of which require additional parameters and model assumptions, and thus, are outside the rough set paradigm. We invite the reader to consult [37, 38, 64] for such methods in the context of RSDA.

The general claim that RSDA is not applicable to most real life problems, because it cannot handle continuous variables is not supported by recent research results [3], and [36, 37, 39] show that RSDA can be supplemented by numerical discretisation procedures. The success of applications of fuzzy controlling, which also requires discretisation of continuous data, shows that the distinction of "continuous data" vs. "discrete data" does not necessarily imply that there is a need for different "continuous methods" , respectively, "discrete methods", to handle these different types of data.

A filtering procedure which uses only internal information has been put forward in [13]. The main idea is to collect attribute values within one attribute domain, which do not split a class of the decision attribute; the procedure makes makes no use of the ordering of attribute values. For example, if there is a rule

$$\text{If } q = 1.2 \text{ or } q = 4.6 \text{ or } q = 4.7, \text{ then } d = \text{green,}$$

then one can collect 1.2, 4.6, and 4.7 into a single attribute value of q. Formally, the descriptor function $f_q : U \to V_q$ is replaced by a function $f'_q : U \to 2^{V_q}$.

The important feature of this procedure is that the internal dependency structure of the system is kept intact, and that no additional parameters are needed. Nevertheless, this simple "nominal" filtering method is surprisingly powerful, as its application to the Iris data set (Table 2) shows: Table 5 gives the filtering for the data. There, for example, the sepal length values 43 – 48 and 53 are collected into one common value 46; in this attribute, filtering reduces the number of classes from 35 to 22. Observe the large reduction in the number of classes of the petal attributes. Extensions of this method to multi-attribute filtering can be found in [79, 80].

9   Extensions and variations of RSDA

In this section we will briefly describe other directions into which RSDA has branched, some of which are only beginning to be investigated.

The *variable precision rough set model* (VPRS), introduced in [83], is a generalisation of the original RSDA in the direction of relaxing the strict boundaries of equivalence classes; it assumes that rules are only valid within a certain part of the population, and it is able to cope with measurement errors:

> "Hypotheses derived from sample data should not . . . be based only on error – free classification rules observed in the sample data. Also, partially incorrect classification should be taken into account. Any partially incorrect classification rule provides valuable trend

Table 5: Filtering of Iris data [3]

Attribute	Filter	No of classes Before	No of classes After
Sepal length:	43–48, 53 → 46		
	66,70 → 70	35	22
	71–79 → 77		
Sepal width:	35, 37, 39–44 → 35	23	16
	20, 24 → 24		
Petal length:	10–19 → 14		
	30–44,46,47 → 46	43	8
	50, 52, 54–69 → 50		
Petal width:	1–6 → 2		
	10–13 → 11	22	8
	17, 20–25 → 17		

information if the majority of available data to which such a rule applies can be correctly classified" [83].

The first step in the VPRS analysis is a numerical form of set inclusion in analogy to the rough membership function of (5.4). If $X, Y \subseteq U$, then the *relative classification error of X with respect to Y* is the value

(9.1)
$$c(X,Y) = \begin{cases} 1 - \frac{|X \cap Y|}{|X|}, & \text{if } X \neq \emptyset, \\ 0, & \text{otherwise.} \end{cases}$$

We now choose a tolerance threshold $0 \le \beta \lneq 0.5$ and set

(9.2)
$$X \stackrel{\beta}{\subseteq} Y \iff c(X,Y) \le \beta.$$

Given an approximation space $\langle U, \theta \rangle$ we define lower and upper β – approximation in analogy to (2.3) and (2.4) by

(9.3)
$$\underline{X}_\beta \stackrel{\text{def}}{=} \bigcup \{ \theta x : \theta x \stackrel{\beta}{\subseteq} X \}$$

(9.4)
$$\overline{X}^\beta \stackrel{\text{def}}{=} \bigcup \{ \theta x : c(\theta x, X) \le \beta \}.$$

In this spirit, one can define a β – dependent approximation quality $\gamma(Q \stackrel{\beta}{\to} P)$, and β – approximate rules. Details can be found in [83], and some application examples in [84].

The advantage of the VPRS is that it uses only two parameters, the external precision parameter β and the internal approximation quality γ, to describe the quality of a rule system; its disadvantages are that precision and γ are partially exchangeable, and that there is as yet no theoretical background to judge which combination is best suited to the data.

17

A much more general form of "rough inclusion" was proposed in [51] which is modelled after Leś niewski's mereology [28, and also the reprint [29]]. Mereology is an alternative to set theory which avoids the pitfalls of Russel's paradox. Unlike set theory, which distinguishes between $\in$ and $\subseteq$, its only connective is a "part of" relation P, which is assumed to be asymmetric, and transitive; thus, P is a strict partial order. It has been recognised that mereology is often more suitable as a language for reasoning about complex objects than first order logics based on Cantorian set theory; applications of mereological systems arise, for example, in spatial reasoning [4, 78] and natural language processing [32].

A *model of rough mereology* is a structure $\mathfrak{M} = \langle X, \mu, n, Fu \rangle$, such that X is a collection of objects, $\mu : X \times X \to [0, 1]$ a function, and $n \in X$; we denote by X^+ the set $X \setminus \{n\}$. Furthermore, $Fu : 2^{X^+} \to X$ is a function, called *fusion*, such that

(9.5)
$$x = Fu(U) \iff (\forall z \in X^+ y)[\mu(z, x) = 1 \Rightarrow$$
$$(\exists w \in U, t \in X^+)(\mu(w, z) = 1 \wedge \mu(w, t) = 1 \wedge \mu(t, x) = 1)].$$

There are the following axioms:

(9.6)
$$(\forall w \in X)[\mu(n, w) = 1 \wedge (\mu(w, n) = 1 \Rightarrow \mu(n, w) = 0)].$$

For all $x, y, z \in X$,

(9.7)
$$\mu(x, x) = 1,$$

(9.8)
$$\mu(x, y) = 1 \Rightarrow \mu(z, y) \geq \mu(z, x).$$

The distinguished n will be referred to as the *null element*[1]. The function μ is called a *rough inclusion*, and it is interpreted as "x is part of y to degree $\mu(x, y)$". An example is the function $\mu_U : 2^U \times 2^U \to [0, 1]$ defined by

(9.9)
$$\mu(X, Y) = \begin{cases} \frac{|X \cap Y|}{|X|}, & \text{if } X \neq \emptyset, \\ \emptyset, & \text{otherwise.} \end{cases}$$

It is shown in [51] that rough mereology extends classical mereology, and that, furthermore,

> " ... we can interpret rough inclusions as global fuzzy membership functions on the universe of discourse which satisfy certain general requirements responsible for their regular mathematical properties.".

Rough inclusions are the basis for reasoning about the synthesis of complex systems from their simpler parts and have widespread applications, for example, computer-aided design, re-engineering, and distributed problem solving. We invite the reader to consult [65] for an overview of recent results in this area.

---

[1]It may be interesting to remark, that Leśniewski was vehemently opposed to the introduction of a null element, [see e.g. 29, p.18ff].

Table 6: State complexity

# of attr. values	# of attributes		
	10	20	30
	$\log_{10}$ (states)		
2	3.01	6.02	9.03
3	4.77	9.54	14.31
4	6.02	12.04	18.06
5	6.99	13.98	20.97

It has been argued that the condition of transitivity for an indiscernibility relation is too strong, viz. the "Sorites Paradox" [34]. Thus, similarity relations, which are reflexive and symmetric, but not necessarily transitive, have been studied inside RSDA in some detail (for example, in rough mereology), and many of the notions of indiscernibility based RSDA have been translated and adjusted to the new situation [23, 27, 55, 67]. The logical structure of the resulting systems has been investigated in [26, 76, 77].

Ordinal prediction takes into account numerical information of the domain of the decision attribute; its rules predict intervals rather than unique values and are of the form

$$\text{If } f_q(x) = a \ldots, \text{then } a \leq f_d(x) \leq b,$$

where $a, b \in V_d$ are real numbers. A rough set approach to ordinal prediction has been given in [11], and a similar problem, applied to the rough approximation of a preference relation, has been studied in [20].

We also should like to mention that the rough approximation of relations is of considerable importance, for example, in social choice, preference relations, or spatial reasoning [5, 6, 8, 16, 81].

10 Concluding remarks

RSDA is the analysis of data based purely on the description of granularity expressed by equivalence relations on the object set U, where the only explicit model assumption is the operationalisation of data in form of an information system designed by a researcher. In that, RSDA differs from most other data analysis methods, and this is where its strengths, but also its weaknesses lie. The strength of requiring no outside parameters is that the method is applicable (and, indeed, applied) to all situations in which data is given in the RSDA operationalisation of information systems. In most cases, real life problems have few data points with respect to state complexity, and show many attribute dependencies (Table 6). Therefore, traditional statistical models are not always optimal tools for knowledge discovery because of their model assumption of representativeness of the sample, and their sensitivity to irrelevant features. In these cases, RSDA can serve as a useful pre-processing tool, removing

superfluous attributes and, with the ROUGHIAN extensions, suggesting validated rules and models on an objective and global basis, i.e without any tuning for a particular application.

If useful and validated background knowledge is available, then the use of hybrid methods, with RSDA as a basis, has proved to be rather successful. In particular, the combination of rough and fuzzy methods turns out to be a powerful tool for soft data analysis, and we invite the reader to consult the collection [46] for a more detailed description and many exemplary case-studies.

References

[1] Bazan, J., Skowron, A. & Synak, P. (1994). Dynamic reducts as a tool for extracting laws from decision tables. In *Proc. of the Symp. on Methodologies for Intelligent Systems, Charlotte, NC*, Lecture Notes in Artificial Intelligence, 346–355, Berlin. Springer–Verlag.

[2] Bazan, J. G. (1998). A comparison of dynamic and non–dynamic rough set methods for extracting laws from decision tables. In [53], 321–365.

[3] Browne, C., Düntsch, I. & Gediga, G. (1998). IRIS revisited: A comparison of discriminant and enhanced rough set data analysis. In [54], 345–368.

[4] Clarke, B. L. (1981). A calculus of individuals based on 'connection'. *Notre Dame Journal of Formal Logic*, **22**, 204–218.

[5] Cohn, A. G. & Gotts, N. M. (1996). The 'egg-yolk' representation of regions with indeterminate boundaries. In P. Burrough & A. M. Frank (Eds.), *Proc. of the GISDATA Specialist Meeting on Geographical Objects with Undetermined Boundaries*, 171–187. Francis Taylor.

[6] Comer, S. (1993). On connections between information systems, rough sets, and algebraic logic. In C. Rauszer (Ed.), *Algebraic Methods in Logic and Computer Science*, vol. 28 of *Banach Center Publications*, 117–124. Warszawa: Polish Academy of Science.

[7] Dubois, D. & Prade, H. (1992). Putting rough sets and fuzzy sets together. In [69], 203–232.

[8] Düntsch, I. (1994). Rough relation algebras. *Fundamenta Informaticae*, **21**, 321–331.

[9] Düntsch, I. (1997a). A logic for rough sets. *Theoretical Computer Science*, **179**, 427–436.

[10] Düntsch, I. (1997b). Rough sets and algebras of relations. In [42], 95–108.

[11] Düntsch, I. & Gediga, G. (1997a). Relation restricted prediction analysis. In A. Sydow (Ed.), *Proc. 15th IMACS World Congress, Berlin*, vol. 4, 619–624, Berlin. Wissenschaft und Technik Verlag.

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

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

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

[15] Düntsch, I. & Gediga, G. (1999). ROUGHIAN – Rough Information Analysis. *International Journal of Intelligent Systems*. To appear.

[16] Düntsch, I. & Orłowska, E. (1999). Approximating regions. Preprint.

[17] Edgington, E. S. (1987). Randomization Tests, vol. 31 of *Statistics: Textbooks and Monographs*. New York and Basel: Marcel Dekker.

[18] Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Ann. Eugen.*, **7**, 179–188.

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

[20] Greco, S., Matarazzo, B. & Słowinski, R. (1998). Rough approximation of a preference relation in a pairwise comparison table. In [54], 13–36.

[21] Grzymała-Busse, J. & Zou, X. (1998). Classification strategy using certain and possible rules. In [52], 37–44.

[22] Grzymała-Busse, J. W. (1992). LERS – A system for learning from examples based on rough sets. In [69], 3–18.

[23] Hu, X. & Cercone, N. (1995). Rough set similarity based learning from databases. In *Proc. of the First International Conference on Knowledge Discovery and Data Mining*, 162–167.

[24] Iwinski, T. B. (1987). Algebraic approach to rough sets. *Bull. Polish Acad. Sci. Math.*, **35**, 673–683.

[25] Komorowski, J., Pawlak, Z., Polkowski, L. & Skowron, A. (1999). Rough sets: A tutorial. In [46], 3–98.

[26] Konikowska, B. (1994). A logic for reasoning about similarity. In H. Rasiowa & E. Orłowska (Eds.), *Reasoning with incomplete information*, vol. 58 of *Studia Logica*, 185–226.

[27] Krętowski, M. & Stepaniuk, J. (1996). Selection of objects and attributes in a tolerance rough set approach. In *Proc. of the Poster Session of the Ninth International Symposium on Methodologies for Intelligent Systems*, 169–180.

[28] Leśniewski, S. (1927 – 1931). O podstawach matematyki. *Przeglad Filozoficzny*, **30–34**.

[29] Leśniewski, S. (1983). On the foundation of mathematics. *Topoi*, **2**, 7–52.

[30] Lin, T. Y. & Cercone, N. (Eds.) (1997). Rough sets and data mining, Dordrecht. Kluwer.

[31] Lin, T. Y., Liu, Q. & Yao, Y. Y. (1994). Logic systems for approximate reasoning: via rough sets and topology. In Z. W. Raś & M. Zemankova (Eds.), *Methodologies for intelligent systems*, 64–74. Berlin: Springer.

[32] Link, G. (1998). Algebraic Semantics in Language and Philosophy, vol. 74 of *CSLI Lecture Notes*. Stanford: Center for the Study of Language and Information.

[33] Manly, B. F. J. (1991). Randomization and Monte Carlo Methods in Biology. London: Chapman and Hall.

[34] Marcus, S. (1998). The paradox of the heap of grains in respect to roughness, fuzziness, and negligibility. In [52], 19–23.

[35] Munakata, T. (1998). Fundamentals of the new artificial intelligence: Beyond traditional paradigms. Springer–Verlag.

[36] Nguyen, H. (1998). From optimal hyperplanes to optimal decision trees. *Fundamenta Informaticae*, **34**, 145–174.

[37] Nguyen, H. S. & Nguyen, S. H. (1998a). Discretization methods in data mining. In [53], 452–482.

[38] Nguyen, H. S., Nguyen, S. H. & Skowron, A. (1996). Searching for features defined by hyperplanes. In Z. W. Ras & M. Michalewicz (Eds.), *ISMIS-96, Ninth International Symposium on Methodologies for Intelligent Systems*, vol. 1079 of *Lecture Notes in Artificial Intelligence*, 366–375, Berlin. Springer–Verlag.

[39] Nguyen, S. & Nguyen, H. (1998b). Pattern extraction from data. *Fundamenta Informaticae*, **34**, 129–144.

[40] Nguyen, S., Skowron, A. & Synak, P. (1998). Discovery of data patterns with applications to decomposition and and classification problems. In [54], 55–97.

[41] Orłowska, E. (1984). Modal logics in the theory of information systems. *Zeitschr. f. Math. Logik und Grundlagen der Math.*, **30**, 213–222.

[42] Orłowska, E. (Ed.) (1997a). Incomplete Information – Rough Set Analysis. Heidelberg: Physica – Verlag.

[43] Orłowska, E. (1997b). Introduction: What you always wanted to know about rough sets. In [42], 1–20.

[44] Orłowska, E. & Pawlak, Z. (1987). Representation of nondeterministic information. *Theoretical Computer Science*, **29**, 27–39.

[45] Pagliani, P. (1997). Rough sets theory and logic-algebraic structures. In [42], 109–190.

[46] Pal, S. & Skowron, A. (Eds.) (1999). Rough Fuzzy Hybridization. Springer–Verlag.

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

[48] Pawlak, Z. (1991). Rough sets: Theoretical aspects of reasoning about data, vol. 9 of *System Theory, Knowledge Engineering and Problem Solving*. Dordrecht: Kluwer.

[49] Pawlak, Z. & Skowron, A. (1994). Rough membership functions. In L. Zadeh & J. Kacprzyk (Eds.), *Fuzzy logic for the management of uncertainty*, 251–271, New York. Wiley.

[50] Peters, J. F. (1998). Time and clock information systems: Concepts and roughly fuzzy Petri net models. In [54], 385–417.

[51] Polkowski, L. & Skowron, A. (1996). Rough mereology: A new paradigm for approximate reasoning. *International Journal of Approximate Reasoning*, **15**, 333–365.

[52] Polkowski, L. & Skowron, A. (Eds.) (1998a). Rough sets and current trends in computing, vol. 1424 of *Lecture Notes in Artificial Intelligence*. Springer.

[53] Polkowski, L. & Skowron, A. (Eds.) (1998b). Rough sets in knowledge discovery, Vol. 1. Heidelberg: Physica–Verlag.

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

[55] Polkowski, L., Skowron, A. & Żytkow, J. (1995). Tolerance based rough sets. In T. Lin & A. Wildenberger (Eds.), *Soft Computing*, 55–58.

[56] Pomykala, J. & Pomykala, J. A. (1988). The Stone algebra of rough sets. *Bull. Polish Acad. Sci. Math.*, **36**, 495–508.

[57] Rauszer, C. (1991). Reducts in information systems. *Fundamenta Informaticae*, **15**, 1–12.

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

[59] Rissanen, J. (1985). Minimum – description – length principle. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of Statistical Sciences*, 523–527, New York. Wiley.

[60] Shafer, G. (1976). A mathematical theory of evidence. Princeton University Press.

[61] Skowron, A. (1990). The rough sets theory and evidence theory. *Fundamenta Informaticae*, **13**, 245–262.

[62] Skowron, A. (1993). A Boolean reasoning for decision rules generation. In J. Komorowski & Z. W. Raś (Eds.), *Methodologies for intelligent systems*, vol. 689 of *Lecture Notes in Artificial Intelligence*, 295–305. Berlin: Springer.

[63] Skowron, A. & Grzymała-Busse, J. W. (1993). From rough set theory to evidence theory. In R. Yager, M. Fedrizzi & J. Kasprzyk (Eds.), *Advances in the Dempster–Shafer Theory of Evidence*, 193–236. New York: Wiley.

[64] Skowron, A. & Nguyen, H. S. (1996). Quantization of real value attributes: Rough set and Boolean reasoning approach. *Bulletin of International Rough Set Society*, **1**, 5–16.

[65] Skowron, A. & Polkowski, L. (1997). Synthesis of decision systems from data tables. In [30], 259–299.

[66] Skowron, A. & Rauszer, C. (1992). The discernibility matrices and functions in information systems. In [69], 331–362.

[67] Skowron, A. & Stepaniuk, J. (1996). Tolerance approximation spaces. *Fundamenta Informaticae*, **27**, 245–253.

[68] Słowiński, K. (1992a). Rough classification of HSV patients. In [69], 77–94.

[69] Słowiński, R. (Ed.) (1992b). Intelligent decision support: Handbook of applications and advances of rough set theory, vol. 11 of *System Theory, Knowledge Engineering and Problem Solving*. Dordrecht: Kluwer.

[70] Słowinski, R. & Stefanowski, J. (1996). Rough set reasoning about uncertain data. *Fundamenta Informaticae*, **27**, 229–243.

[71] Stefanowski, J. (1998). On rough based approaches to induction of decision rules. In [53], 500–529.

[72] Tanaka, M., Ishibuchi, H. & Shigenaga, T. (1992). Fuzzy inference system based on rough sets and its application to medical diagnosis. In [69], 111–118.

[73] Teghem, J. & Charlet, J.-M. (1992). Use of "rough sets" method to draw premonitory factors for earthquakes by emphasing gas geochemistry: the case of a low seismic activity context in Belgium. In [69], 165–179.

[74] Torgerson, W. S. (1958). Theory and methods of scaling. New York: Wiley.

[75] Tsumoto, S. & Tanaka, H. (1996). A common algebraic framework for empirical learning methods based on rough sets and matroid theory. *Fundamenta Informaticae*, **27**, 273–288.

[76] Vakarelov, D. (1991). A modal logic for similarity relations in Pawlak knowledge representation systems. *Fundamenta Informaticae*, **15**, 61–79.

[77] Vakarelov, D. (1997). Information systems, similarity relations, and modal logics. In [42], 492–550.

[78] Varzi, A. C. (1996). Parts, wholes, and part–whole relations: The prospect of mereotopology. *Data & Knowledge Engineering*, **20**, 259–286.

[79] Wang, H., Düntsch, I. & Bell, D. (1998a). Data reduction based on hyper relations. In R. Agrawal, P. Stolorz & G. Piatetsky-Shapiro (Eds.), *Proceedings of KDD'98*, 349–353, New York.

[80] Wang, H., Düntsch, I. & Gediga, G. (1998b). Classificatory filtering in decision systems. Submitted for publication.

[81] Worboys, M. (1998). Imprecision in finite resolution spatial data. *Geoinformatica*, **2**, 257–280.

[82] Yao, Y. (1997). Combination of rough and fuzzy sets based on α-level sets. In [30], 301–321.

[83] Ziarko, W. (1993). Variable precision rough set model. *Journal of Computer and System Sciences*, **46**.

[84] Ziarko, W. (1998). Rough sets as a methodology in data mining. In [53], 554–573.

[85] Ziarko, W. & Shan, N. (1993). A method for computing all maximally general rules in attribute–value systems. *Computational Intelligence*, **2**, 2–13.