Data Analysis with Rough Set Theory

Zdzisław Pawlak
Institute of Theoretical and Applied Informatics
Polish Academy of Sciences
ul. Baltycka 5, 44 000 Gliwice, Poland
e-mail: zpw@ii.pw.edu.pl

1 Introduction

We present here a new approach to data analysis called rough set theory (Pawlak, 1982) The rough set philosophy is founded on the assumption that with every object of the universe of discourse we associate some information (data, knowledge). Objects characterized by the same information are indiscernible (similar) in view of the available information about them. The indiscernibility relation generated in this way is the mathematical basis of rough set theory.

Any set of all indiscernible (similar) objects is called an elementary set, and form a basic granule (atom) of knowledge about the universe. Any union of some elementary sets is referred to as crisp (precise) set – otherwise the set is rough (imprecise, vague).

Consequently each rough set has boundary-line cases, i.e., objects which cannot be with certainty classified as members of the set or of its complement. Obviously crisp sets have no boundary-line elements at all. That means that boundary-line cases cannot be properly classified by employing the available knowledge.

Thus, the assumption that objects can be "seen" only through the information available about them leads to the view that knowledge has granular structure. Due to the granularity of knowledge some objects of interest cannot be discerned and appear as the same (or similar). As, a consequence vague concepts, in contrast to precise concepts, cannot be characterized in terms of information about their elements. Therefore in the proposed approach we assume that any vague concept is replaced by a pair of precise concepts – called the lower and the upper approximation of the vague concept. The lower approximation consists of all objects which surely belong to the concept and the upper approximation contains all objects which possible belong to the concept. Obviously, the difference between the upper and the lower approximation constitute the boundary region of the vague concept. Approximations are two basic operations in the rough set theory.

Rough set theory overlaps to a certain degree many other mathematical theories. Particularly interesting is the relationship with fuzzy set theory and Dempster-Shafer theory of evidence. The concepts of rough set and fuzzy set are different since they refer to various aspects of imprecision (Pawlak and Skowron, 1994) whereas the connection with theory of evidence is more substantial (Skowron and Grzymał-Busse, 1994). Besides, rough set theory is related to discriminant analysis (Krusińska *et al.,* 1992), Boolean reasoning methods (Skowron and Rauszer, 1992) and others. The relationship between rough set theory and decision analysis is presented in (Pawlak and Słowiński, 1994, Słowiński, 1993). More details concerning these relationships can be found in the references.

Despite of the relationships rough set theory can be viewed in its own rights, as an the independent discipline.

Rough set theory has found many interesting applications. The rough set approach seems to be of fundamental importance to AI and cognitive sciences, especially in the areas of machine learning, knowledge acquisition, decision analysis, knowledge discovery from databases, expert systems, inductive reasoning and pattern recognition. It seems of particular importance to decision support systems.

The main advantage of rough set theory is that it does not need any preliminary or additional information about data – like probability in statistics, or basic probability assignment in Dempster-Shafer theory and grade of membership or the value of possibility in fuzzy set theory.

The rough set theory has been successfully applied in many real-life problems in medicine, pharmacology, engineering, banking, financial and market analysis and others. Some exemplary applications are listed below.

There are many applications in medicine (Grzymała-Busse and Woolerly 1994, Słowiński K., *et al.,* 1988, Słowiński K., 1992, Słowiński K., and Sharif 1993, Słowiński K., *et al.,* 1995, Tanaka *et al.*, 1992). In pharmacology the analysis of relationships between the chemical structure and the antimicrobial activity of drugs (Krysiński 1990, 1992, 1992, 1995) has been successfully investigated. Banking applications include evaluation of a bankruptcy risk (Słowiński R., and Zopounidis 1993, 1994) and market research (Golan and Edwards 1993, Ziarko and Katzberg 1989). Very interesting results have been also obtained in speaker independent speech recognition (Brindle 1994, Czyżewski 1995, Czyżewski and Kaczmarek 1993, 1995, 1995) and acoustics (Kostek 1995, 1995, 1995, 1995). The rough set approach seems also important for various engineering applications, like diagnosis of machines using vibroacoustics, symptoms (noise, vibrations) (Nowicki *et al.,* 1992, 1992, 1992), material sciences (Jackson *et al.,* 1994) and process control (Lin 1995, Mr ozek 1992, Munakata 1995, Płonka and Mrózek 1995, Szladow and Ziarko 1992, Ziarko 1992, Ziarko and Katzberg 1989). Application in linguistics (Grzymała-Busse *et al.,* 1995, Grzymała-Busse and Than 1993, Kobayashi and Yokomori 1995, Moradi *et al.,* 1995) and environment (Gunn *et al.*, 1994), databses (Beaubouef and Petry 1995, Beaubouef *et al.,* 1995, Cercone and Han 1993, Shenoi 1995, Ziarko 1991) are other important domains.

More about applications of the rough set theory can be found in (Grzymała-Busse 1995, Lin 1994, Słowiński R., 1992, Wang 1995, Ziarko 1993). Besides, many other fields of application, e.g., time series analysis, image processing and character recognition, are being extensively explored.

Application of rough sets requires a suitable software. Many software systems for workstations and personal computers based on rough set theory have been developed. The most known include LERS (Grzymała-Busse 1992), Rough DAS and Rough Class and DATALOGIC (Szladow 1993). Some of them are available commercially.

One of the most important and difficult problem in software implementation of the presented approach is optimal decision rule generation from data. Many various approaches to solve this task can be found in (Bazan *et al.,* 1995, 1994, Grzymała-Busse *et al.,* 1995, Skowron 1995, Skowron and Stepaniuk 1994, Tsumoto and Tanaka 1995, Wróblewski 1995). The relation to other methods of rule generation is dwelt in (Grzymała-Busse *et* *al.*, 1995).

The theory has many important advantages. Some of them are listed below.

- Provides efficient algorithms for finding hidden patterns in data.
- Finds minimal sets of data (data reduction).
- Evaluates significance of data.
- Generates sets of decision rules from data.
- It is easy to understand.
- Offers straightforward interpretation of obtained results.
- Most algorithms based on the rough set theory are particularly suited for parallel processing, but in order to exploit this feature fully, a new computer organization based on rough set theory is necessary.

Although rough set theory has many achievements to its credit, nevertheless several theoretical and practical problems require further attention.

Especially important is widly accessible efficient software development for rough set based data analysis, particularly for large collections of data analysis.

Despite of many valuable methods of efficient, optimal decision rule generation methods from data, developed in recent years based on rough set theory – more research here is needed, particularly, when quantitative attributes are involved. In this context also further discretization methods for quantitative attribute values are badly needed. Also an extensive study of a new approach to missing data is very important. Comparison to other similar methods still requires due attention, although important results have been obtained in this area. Particularly interesting seems to be a study of the relationship between neural network and rough set approach to feature extraction from data.

Last but not least, rough set computer is badly needed for more serious computations in decision support. Some research in this area is already in progress.

For basic ideas of rough set theory the reader is referred to (Grzymała-Busse 1995, Nakamura *et al* 1996, Pawlak 1991, Pawlak *et al* 1995, Słowiński 1995, Szladow and Ziarko 1993).

2 Decision Tables and Decision Rules

Data are often presented as a table, columns of which are labeled by *attributes*, rows by *objects* of interest and entries of the table are *attribute values*. An example of such table is shown below.

Store	E	Q	L	P
1	high	good	no	profit
2	med.	good	no	loss
3	med.	good	no	profit
4	no	avg.	no	loss
5	med.	avg.	yes	loss
6	high	avg.	yes	profit

Table 1

In the table six stores are characterized by four attributes:

E – empowerment of sales personnel,
Q – perceived quality of merchandise,
L – high traffic location,
P – store profit or loss.

Sometimes we distinguish in such a table two classes of attributes, called *condition* and *decision* (*action*) attributes. For example in Table 1 attributes E, Q, L are condition attributes, whereas the attribute P, is a decision one. Such tables will be referred to as *decision tables.*

Each row of a decision table deteremines a *decision rule*, which specifies *decisions* (*actions*) that should be taken when conditions pointed out by *condition* attributes are satisfied. Decision rules 2) and 3) in Table 1 have the same conditions by different decisions. Such rules are called *inconsistent* (*nondeterministic, conflicting*); otherwise the rules are referred to as *consistent* (*certain, deterministic, nonconflicting*). Decision tables containing inconsistent decision rules are called *inconsistent* (*nondeterministic, conflicting*); otherwise the table is *consistent* (*deterministic, non conflicting*).

The number of consistent rules to all rules in a decision table can be used as *consistency measure* of the decision table, and will be denoted by $\gamma(C, D)$, where C and D are condition and decision attributes respectively. Thus if $\gamma(C, D) = 1$ the decision table is consistent and if $\gamma(C, D) \neq 1$ the decision table is inconsistent. For example for Table 1 $\gamma(C, D) = 4/6$.

Decision rules are often presented as implications and are called "*if... then...*" rules. For example rule 1) in Table 1 can be presented as implication

if (H, high) and (Q, good) and (L, no) then (P, profit).

3 Rough Sets and Approximations

As mentioned in the introduction, the starting point of the rough set theory is the indiscernibility relation, generated by information about objects of interest. The indiscernibility relation is intended to express the fact that due to the lack of knowledge we are unable to discern some objects employing the available information. That means that, in general, we are unable to deal with single objects but we have to consider clusters of indiscernible objects, as fundamental concepts of our theory.

Now we present above considerations more formally.

Suppose we are given two finite, non-empty sets U and A, where U is the *universe*, and A – a set *attributes*. With every attribute $a \in A$ we associate a set V_a, of its *values*, called the *domain* of a. Any subset B of A determines a binary relation $I(B)$ on U, which will be called an *indiscernibility relation*, and is defined as follows:

$xI(B)y$ if and only if $a(x) = a(y)$ for every $a \in A$,
where $a(x)$ denotes the value of attribute a for element x.

Obviously $I(B)$ is an equivalence relation. The family of all equivalence classes of $I(B)$, i.e., partition determined by B, will be denoted by $U/I(B)$, or simple U/B; an equivalence class of $I(B)$, i.e., block of the partition U/B, containing x will be denoted by $B(x)$.

If (x, y) belongs to $I(B)$ we will say that x and y are *B-indiscernible*. Equivalence classes of the relation $I(B)$ (or blocks of the partition U/B) are refereed to as *B-elementary sets*. In the rough set approach the elementary sets are the basic building blocks (concepts) of our knowledge about reality.

The indiscernibility relation will be used next to define basic concepts of rough set theory. Let us define now the following two operations on sets

$$B_*(X) = \{x \in U : B(x) \subseteq X\},$$

$$B^*(X) = \{x \in U : B(x) \cap X \neq \emptyset\},$$

assigning to every subset X of the universe U two sets $B_*(X)$ and $B^*(X)$ called the *B-lower* and the *B-upper approximation* of X, respectively. The set

$$BN_B(X) = B^*(X) - B_*(X)$$

will be referred to as the *B-boundary region* of X.

If the boundary region of X is the empty set, i.e., $BN_B(X) = \emptyset$, then the set X is *crisp* (*exact*) with respect to B; in the opposite case, i.e., if $BN_B(X) \neq \emptyset$, the set X is to as *rough* (*inexact*) with respect to B. Let us depict the above ideas by an example refering to Table 1.

Let us observe that each store has different description in terms of attributes E, Q, L and P, thus all stores may be distinguished (discerned) employing information provided by all attributes. However, stores 2 and 3 are indiscernible in terms of attributes E, Q and L, since they have the same values of these attributes. Similarly, stores 1, 2 and 3 are indiscernible with respect to attributes Q and L, etc.

Each subset of attributes determines a partition (classification) of all objects into classes having the same description in terms of these attributes. For example, attributes Q and L aggregate all stores into the following classes $\{1, 2, 3\}, \{4\}, \{5, 6\}$. Thus, each information table determines a family of classification patterns which are used as a basis of further considerations.

Suppose we are interested in the following problem: what are the characteristic features of stores having profit (or loss) in view of information available in Table 1. In other words, the question is whether we are able to describe set (concept) $\{1, 3, 6\}$ (or $\{2, 4, 5\}$) in terms of attributes E, Q and L. It can be easily seen that this is impossible, since stores 2 and 3 display the same features in terms of attributes E, Q and L, but store 2 makes a profit, whereas store 3 has a loss. Thus information given in Table 1 is not sufficient to answer this question. However, we can give a partial answer to this question. Let us observe that if the attribute E has the value *high* for a certain store, then the store makes a profit, whereas if the value of the attribute E is *low*, then the store has a loss. Thus, in view of information contained in Table 1, we can say for sure that stores 1 and 6 make a profit, stores 4 and 5 have a loss, whereas stores 2 and 3 cannot be classified as making a profit or having a loss. Therefore we can give approximate answers only. Employing attributes E, Q and L, we can say that stores 1 and 6 *surely* make a profit, i.e., *surely* belong to the set $\{1, 3, 6\}$, whereas stores 1,2,3 and 6 *possibly* make a profit, i.e., possibly belong to the set $\{1, 3, 6\}$. Thus the set $\{1, 6\}$ is the *lower approximation* of the set (concept) $\{1, 3, 6\}$, and the set $\{1, 2, 3, 6\}$ – is the *upper approximation* of the set $\{1, 3, 6\}$. The set $\{2, 3\}$, being the difference between the upper approximation and the lower approximation is refered to as the *boundary region* of the set $\{1, 3, 6\}$.

Rough set can be also characterized numerically by the following coefficient

$$\alpha_B(X) = \frac{|B_*(X)|}{|B^*(X)|}$$

called *accuracy of approximation*, where $|X|$ denotes the cardinality of X. Obviously $0 \leq \alpha_B(X) \leq 1$. If $\alpha_B(X) = 1$, X is *crisp* with respect to B (X is *precise* with respect to B), and otherwise, if $\alpha_B(X) < 1$, X is *rough* with respect to B (X is *vague* with respect to B).

Rough sets can be also defined using a *rough membership function*, defined as

$$\mu_X^B(x) = \frac{|X \cap B(x)|}{|B(x)|}.$$

Obviously

$$\mu_X^B(x) \in [0, 1].$$

Value of the membership function $\mu_X(x)$ is kind of conditional probability, and can be interpreted as a degree of *certainty* to which x belongs to X (or $1 - \mu_X(x)$, as a degree of *uncertainty*).

4 Dependency of Attributes

Another important issue in data analysis is discovering dependencies between attributes. Intuitively, a set of attributes D depends totally on a set of attributes C, denoted $C \Rightarrow D$, if all values of attributes from D are uniquely determined by values of attributes from C. In other words, D depends totally on C, if there exists a functional dependency between values of D and C.

Formally dependency can be defined in the following way. Let D and C be subsets of A. We say that B *depends totally* on C, if and only if $I(C) \subseteq I(D)$. That means that the partition generated by C is finer than the partition generated by D. Notice, that the concept of dependency discussed above corresponds to that considered in relational databases.

We would need also a more general concept of dependency of attributes, called a *partial dependency* of attributes.

Formally, the above idea can be formulated as follows. Let D and C be subsets of A. We say that D *depends in degree* $k, 0 \leq k \leq 1$, on C, denoted $C \Rightarrow_k D$, if

$$k = \frac{|POS_C(D)|}{|U|},$$

where

$$POS_C(D) = \bigcup_{X \in U/I(D)} C_*(X).$$

The expression $POS_C(D)$, called a *positive region* of the partition U/D with respect to C, is the set of all elements of U that can be uniquely classified to blocks of the partition U/D, by means of C.

Thus the coefficient k expresses the ratio of all elements of the universe, which can be properly classified to blocks of the partition U/D, employing attributes C. Notice that for $k = 1$ we get the previous definition of total dependency.

Obviously, a decision table is consistent if and only if $k = 1$, otherwise, i.e., if $k \neq 1$, the decision table is inconsistent; if $k = 0$ we will say that the decision table is totally inconsistent.

Obviously dependency between attributes can be defined using the consituency factor i.e.,

$$C \Rightarrow_k D,$$

where $k = \gamma(C, D)$.

Summing up: D is *totally* (*partially*) dependent on C, if *all* (*some*) elements of the universe U can be uniquely classified to blocks of the partition U/D, employing C.

5 Data Reduction (Compression)

We often face a question whether we can remove some data from a data-table preserving its basic properties, that is – whether a table contains some superfluous data.

Very often we are interested in reducing the number of condition attributes preserving the degree of dependency between decision and condition attributes. That means that we want to preserve the ability to classify objects to decision classes using smaler number of conditions attributes – or, in other words, we want to make decisions employing less conditions.

To this end we define the concept of a *recuct* of *attributes*. $B \subseteq C$ is a *D-reducts* of C, if B is a minimal subset of C such that

$$\gamma(B, D) = \gamma(C, D).$$

For example, in Table 1 we have two reducts of $\{E, Q, L\}$ – mamely $\{E, Q\}$ and $\{E, L\}$.

That means that either Q or L can be removed from the table without changing the degree of consistency of the table.

The intersection of all reducts is called the *D-core* of condition attributes, i.e.,

$$CORE_D(C) = \bigcap RED_D(C).$$

In Table 1 the core is the attribute E. The core can be interpreded as a set of the "most important" attributes, which cannot be removed from data, without effecting the consitancy factor of the table.

References

[1] Bazan, J., Skowron, A., and Synak, P., (1995), "Discovery of decision rules from experimental data", in: T.Y. Lin and A.M. Wildberger (eds.), *The Third International Workshop on Rough Sets and Soft Computing Proceedings* (*RSSC'94*), San Jose State University, San Jose, California, USA, November 10–12, 276–279.

[2] Bazan, J., Skowron, A., and Synak, P., (1994), "Dynamic reducts as a tool for extracting laws from decision tables", *Proc. of the Symp. on Methodologies for Intelligent Systems*, Charlotte, NC, October 16–19, 1994, Lecture Notes in Arificial Intelligence Vol. 869, Springer Verlag, 346–355.

[3] Bazan, J., Skowron, A., and Synak P., (1994), "Discovery of decision rules from experimental data", in: T.Y. Lin (ed.), *The Third International Workshop on Rough Sets and Soft Computing Proceedings* (*RSSC'94*), San Jose State University, San Jose, California, USA, November 10–12, 526–535.

[4] Beaubouef, T., and Petry, F.E., (1993), "A rough set model for relational databases", in: W. Ziarko (ed.), *Rough Sets, Fuzzy Sets and Knowledge Discovery. Proceedings of the International Workshop on Rough Sets and Knowledge Discovery* (*RSKD'93*), Banff, Alberta, Canada, October 12–15, Springer-Verlag, Berlin, 100–107.

[5] Beaubouef, T., and Petry, F.E., (1995), "Rough querying of crisp data in relational databases", in: T.Y. Lin and A.M. Wildberger (eds.), (1995), *The Third International Workshop on Rough Sets and Soft Computing Proceedings* (*RSSC'94*), San Jose State University, San Jose, California, USA, November 10–12, 85–88.

[6] Beaubouef, T., Petry, F.E., and Buckles, B.P., (1995), "Extension of the relational Database and its algebra with rough set techniques", *Computational Intelligence: An International Journal Vol. 11*, 233–245.

[7] Brindle, D., (1994), "Speaker-independent speech recognition by rough sets analysis", in: T.Y. Lin (ed.), *The Third International Workshop on Rough Sets and Soft Computing Proceedings* (*RSSC'94*), San Jose State University, San Jose, California, USA, November 10–12, 376–383.

[8] Czyżewski , A., (1995) , "Speaker-Independent Recognition of Digits - Experiments with Neural Networks, fuzzy logic and rough sets", *Journal of the Intelligent Automation and Soft Computing* (to appear).

[9] Czyżewski, A., and Kaczmarek, A., (1993), "Multilayer knowledge based system for speaker-independent recognition of isolated words", in: W. Ziarko (ed.), *Rough Sets, Fuzzy Sets and Knowledge Discovery. Proceedings of the International Workshop on Rough Sets and Knowledge Discovery* (*RSKD'93*), Banff, Alberta, Canada, October 12–15, Springer-Verlag, Berlin, 387–394.

[10] Czyżewski, A., and Kaczmarek, A., (1995), "Speaker-independent recognition of isolated words using rough sets", in: P.P. Wang (ed.), *Second Annual Joint Conference on Information Sciences PROCEEDINGS*, September 28 – October 1, Wrightsville Beach, North Carolina, USA, 397–400.

[11] Czyżewski A., and Kaczmarek, A., (1995), "Speech recognition systems based on rough sets and neural networks", in: T.Y. Lin and A.M. Wildberger (eds.), *The Third International Workshop on Rough Sets and Soft Computing Proceedings* (*RSSC'94*), San Jose State University, San Jose, California, USA, November 10-12, 97–100.

[12] Czogała, E., Mr ozek, A. and Pawlak, Z. (1995). "The Idea of Rough-Fuzzy Controller". *International Journal of Fuzzy Sets and Systems*, 72, 61–63.

[13] Dubois, D. and Prade, H. (1992),"Putting Rough Sets and Fuzzy Sets Together", in: R. Słowi nski (ed.), *Intelligent Decision Support – Handbook of Advances and Applications of the Rough Set Theory*, , Kluwer Academic Publishers, Dordrecht, Boston, London, 203–232.

[14] Frege, G. (1903). Grundgesetze der Arithmetik, 2, in: *Selections from the Philosophical Writings of Gotlob Frege*, Blackweil, Oxford, 1970.

[15] Golan, R., and Edwards, D., (1993), "Temporal rules discovery using datalogic/R+ with stock market data", in: W. Ziarko *Rough Sets, Fuzzy Sets and Knowledge Discovery. Proceedings of the International Workshop on Rough Sets and Knowledge Discovery* (*RSKD'93*), Banff, Alberta, Canada, October 12–15, Springer-Verlag, Berlin, 74–81.

[16] Grzymała-Busse, J.W., (1988), "Knowledge acquisition under uncertainty - A rough set approach", *Journal of Intelligent* & *Robotic Systems*, 1/1, 3–16.

[17] Grzymała-Busse J.W., (1991), *Managing Uncertainty in Expert Systems*. Kluwer Academic Publishers, Dordrecht, Boston, London.

[18] Grzymała-Busse, J.W., (1992), "LERS-a system for learning from examples based on rough sets", in: R. Słowiński (ed.), *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory*, Kluwer Academic Publishers, Dordrecht, 3–18.

[19] Grzymała-Busse, J. (1995). "Rough Sets", *Advances in Imaging and Electrons Physics*, 94, to appear.

[20] Grzymała-Busse, J.W., Sedelow, S.Y., and Sedelow, W.A. Jr., (1995), "Machine learning & knowledge acquisition, rough sets, and the English semantic code". *Proc. of the Workshop on Rough Sets and Database Mining, 23rd Annual ACM Computer Science Conference CSC'95*, Nashville, TN, March 2, 91–109.

[21] Grzymała-Busse, Stefanowski, J., and Ziarko W., (1995), "Rough sets: facts versus misconceptions", *ICS Research Report 61/95*.

[22] Grzymała-Busse, J.W., and Than, S. (1993), "Data compression in machine learning applied to natural language". *Behavior Research Methods, Instruments, & Computers* 25, 318–321.

[23] Grzymała-Busse, and Woolerly, L., (1994), "Improving prediction of preterm birth using a new classification scheme and rule induction", *Proc. of the 18-th Annual Symposium on Computer Applications in Medical Care (SCAMC)*, Washington D.C. November 5–9, 730–734.

[24] Gunn, J.D., and Grzymała-Busse, J.W., (1994), "Global temperature stability by rule induction: an interdisciplinary bridge", *Human Ecology*, 22, 59–81.

[25] Hadjimichael, M., and Wasilewska, A. (1992), "Rough sets-based study of voter preference in 1988 USA presidential election", in: R. Słowiński (ed.), *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory*, Kluwer Academic Publishers, Dordrecht, 137–152.

[26] Haines, S., (1995), "The practical application of rough sets to semantics and simulation", in: T.Y. Lin and A.M. Wildberger (eds.), *The Third International Workshop on Rough Sets and Soft Computing Proceedings* (*RSSC'94*), San Jose State University, San Jose, California, USA, November 10–12, 77–80.

[27] Haines, S., Longshaw, T., and Magee, G., (1994), "The practical application of rough sets to sementics and simulation", in: T.Y. Lin (ed.), *The Third International Workshop on Rough Sets and Soft Computing Proceedings* (*RSSC'94*), San Jose State University, San Jose, California, USA, November 10–12, 392–398.

[28] Hu, X., Cercone, N., and Han, J., (1993), "An attribute-oriented rough set approach for knowledge discovery in databases", in: W. Ziarko (ed.), *Rough Sets, Fuzzy Sets and Knowledge Discovery. Proceedings of the International Workshop on Rough Sets and Knowledge Discovery* (*RSKD'93*), Banff, Alberta, Canada, October 12–15, Springer-Verlag, Berlin. 90–99.

[29] Jackson, A.G., Ohmer, M., and Al-Kamhawi, H., (1994), "Rough sets analysis of chalcopyrite semiconductor band gap data", in: T.Y. Lin (ed.), *The Third International Workshop on Rough Sets and Soft Computing Proceedings* (*RSSC'94*), San Jose State University, San Jose, California, USA, November 10–12, 408–417.

[30] Jackson, A.G., LeClair, S.R., Ohmer, M.C., Ziarko. W., and Al-Kamhwi, H., "Rough sets applied to material data", *Acta Metallurgica et Materialia* (to appear).

[31] Jelonek, J., Krawiec, K., Słowiński, R., Stefanowski, J., and Szymas, J., (1993), "Neural networks and rough sets - comparison and combination for classification of histological pictures", in: W. Ziarko (ed.), *Rough Sets, Fuzzy Sets and Knowledge Discovery. Proceedings of the International Workshop on Rough Sets and Knowledge Discovery* (*RSKD'93*), Banff, Alberta, Canada, October 12–15, Springer-Verlag, Berlin, 426–433.

[32] Jelonek, J., Krawiec, K., Słowiński, R., and Szymas, J., (1994), "Rough set reduction of features for picture-based reasoning", in: T.Y. Lin (ed.), *The Third International Workshop on Rough Sets and Soft Computing Proceedings* (*RSSC'94*), San Jose State University, San Jose, California, USA, November 10–12, 418–425.

[33] Kobayashi, S., Yokomori, T., (1995), "Approximately learning regular languages with respect to reversible languages: A rough set based analysis", in: P.P. Wang (ed.), *Second Annual Joint Conference on Information Sciences PROCEEDINGS*, September 28 – October 1, Wrightsville Beach, North Carolina, USA, 91–94.

[34] Kostek B., (1995), "Rough classification as a tool for acoustical analysis", in: T.Y. Lin and A.M. Wildberger (eds.), *The Third International Workshop on Rough Sets and Soft Computing Proceedings* (*RSSC'94*), San Jose State University, San Jose, California, USA, November 10–12, 81–84.

[35] Kostek B., (1995), "Computer based recognition of musical phrases using the rough set apprach", in: P.P. Wang (ed.), *Second Annual Joint Conference on Information Sciences PROCEEDINGS*, September 28 – October 1, Wrightsville Beach, North Carolina, USA, 401–404.

[36] Kostek, B., (1995), "Statistical versus artificial intelligence based processing of subjective test results", *98th Convention of the Audio Engineering Society*, Paris, February 25–28, Preprint No. 4018.

[37] Kostek B., (1995), "Rough set and fuzzy set methods applied to acoustical analyses", *Journal of the Intelligent Automation and Soft Computing* (to appear).

[38] Krusińska E., Słowiński R., and Stefanowski J., (1992), "Discriminant versus rough set approach to vague data analysis", *Applied Stochastic Models and Data Analysis*, 8, 43–56.

[39] Krysiński, J., (1990), "Rough set approach to the analysis of structure - activity relationship of quaternary imidazolium compounds", *Arzneimittel Forschung / Drug Research*, 40/II, 795–799.

[40] Krysiński, J., (1992), "Grob Mengen Theorie in der Analyse der Struktur Wirkungs Beziehungen von quartaren Pyridiniumverbindungen", *Pharmazie*, 46/12, 878–881.

[41] Krysiński, J., (1992), "Analysis of structure - activity relationships of quaternary ammonium compounds", in: R. Słowiński (ed.), *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory*, Kluwer Academic Publishers, Dordrecht, 119–136.

[42] Krysiński, J., (1995), "Application of the rough sets theory to the analysis of structure-activity-relationships of antimicrobial pyridinium compounds", *Die Pharmazie*, 50, 593–597.

[43] Kryszkiewicz, M., and Rybiński, H., (1993), "Finding Reducts in Composed Information Systems", in: W. Ziarko (ed.), *Rough Sets, Fuzzy Sets and Knowledge Discovery. Proceedings of the International Workshop on Rough Sets and Knowledge*

Discovery (*RSKD'93*), Banff, Alberta, Canada, October 12–15, Springer-Verlag, Berlin, 261–273.

[44] Lenarcik, A., and Piasta, Z., (1993), "Rough classfiers", in: W. Ziarko (ed.), *Rough Sets, Fuzzy Sets and Knowledge Discovery. Proceedings of the International Workshop on Rough Sets and Knowledge Discovery* (*RSKD'93*), Banff, Alberta, Canada, October 12–15, Springer-Verlag, Berlin, 298–316.

[45] Lin, T.Y., (ed.), (1994), *The Third International Workshop on Rough Sets and Soft Computing Proceedings* (*RSSC'94*), San Jose State University, San Jose, California, USA, November 10–12.

[46] Lin, T.Y., (1995), "Topological view of stability in rough-fuzzy controllers", in: P.P. Wang (ed.), *Second Annual Joint Conference on Information Sciences PROCEEDINGS*, September 28 - October 1, Wrightsville Beach, North Carolina, USA.

[47] Lin, T.Y., and Wildberger A.M., (eds.), (1995), *The Third International Workshop on Rough Sets and Soft Computing Proceedings* (*RSSC'94*), San Jose State University, San Jose, California, USA, November 10–12.

[48] Lau, S.S.Y., (1993), "Image segmentation based on the indiscernibility relation", in: W. Ziarko (ed.), *Rough Sets, Fuzzy Sets and Knowledge Discovery. Proceedings of the International Workshop on Rough Sets and Knowledge Discovery* (*RSKD'93*), Banff, Alberta, Canada, October 12–15, Springer-Verlag, Berlin, 395–403.

[49] Mrózek, A., (1989), "Rough sets and dependency analysis among attributes in computer implementations of expert's inference models", *International Journal of Man-Machine Studies*, 30, 457–471.

[50] Mrózek, A., (1992), "Rough sets in computer implementation of rule-based control of industrial processes", in: R. Słowiński (ed.), *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory*, Kluwer Academic Publishers, Dordrecht, 19–31.

[51] Moradi, H., Grzymała-Busse, J., and Roberts, J., (1995), "Entropy of English text: Experiments with humans nad machine learning system based on rough sets", in: P.P. Wang (ed.), *Second Annual Joint Conference on Information Sciences PROCEEDINGS*, September 28 – October 1, Wrightsville Beach, North Carolina, USA, 87–88.

[52] Munakata, T., (1995), "Rough control: Basic ideas and applications", in: Wang, P.P., (ed.), *Second Annual Joint Conference on Information Sciences PROCEEDINGS*, September 28 – October 1, Wrightsville Beach, North Carolina, USA, 340–343.

[53] Nakamura, A., Tsumoto, S., Tanaka, H., and Kobayashi, S., (1996), "Rough set theory and its application", *J. of Japanese Society for Artificial Intelligence*, 11, 35–41, (in Japanese).

[54] Nakamura, A., (1996), "Rough sets – Its theory and applications", *J. of Japan Society for Fuzzy Thory and Systems*, 8, 594–603, (in Japanese).

[55] Nowicki, R., Słowiński, R., and Stefanowski, J.,(1992), "Rough sets analysis of diagnostic capacity of vibroacoustic symptoms", *Journal of Computers and Mathematics with Applications*, 24/2, 109–123.

[56] Nowicki, R., Słowiński, R., and Stefanowski, J., (1992), "Evaluation of vibroacoustic diagnostic symptoms by means of the rough sets theory", *Journal of Computers in Industry*, 20, 141–152.

[57] Nowicki, R., Słowiński, R., and Stefanowski, J., (1992), "Analysis of diagnostic symptoms in vibroacoustic diagnostics by means of the rough set theory", in: R. Słowiński (ed.), *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory*, Kluwer Academic Publishers, Dordrecht, 33–48.

[58] Pawlak Z., (1982), "Rough sets". *International Journal of Computer and Information Sciences*, 11, 341–356.

[59] Pawlak Z., (1991), *Rough Sets - Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Dordrecht, Boston, London.

[60] Pawlak Z., Grzymała-Busse J. W., Słowiński R., and Ziarko, W., (1995),"Rough sets", *Communication of the ACM*, 38, 88–95.

[61] Pawlak Z., and Skowron A., (1994), "Rough membership functions", in: R.R Yaeger, M. Fedrizzi and J. Kacprzyk (eds.), *Advances in the Dempster Shafer Theory of Evidence*, John Wiley & Sons, Inc., New York, Chichester, Brisbane, Toronto, Singapore, 251–271.

[62] Pawlak Z., and Słowiński R., (1994), "Rough set approach to multi-attribute decision analysis, Invited Review", *European Journal of Operational Research*, 72, 443–459.

[63] Pawlak, Z., Wong, S.K.M. and Ziarko, W. (1988). "Rough Sets: Probabilistic versus Deterministic Approach", *International Journal of Man Machine Studies*, 29, 81–85.

[64] Piasta, Z., (1995), "A comparative analysis of classifiers' performance by using a simulation study", in: T.Y. Lin and A.M. Wildberger (eds.), *The Third International Workshop on Rough Sets and Soft Computing Proceedings* (*RSSC'94*), San Jose State University, San Jose, California, USA, November 10–12, 65–68.

[65] Peterson, G.I., (1993), "Rough classification of pneumonia patients using a clinical database", in: W. Ziarko (ed.), *Rough Sets, Fuzzy Sets and Knowledge Discovery. Proceedings of the International Workshop on Rough Sets and Knowledge Discovery* (*RSKD'93*), Banff, Alberta, Canada, October 12–15, Springer-Verlag, Berlin, 412–419.

[66] Płonka, L., and Mrózek, A., (1995), "Rule-based stabilization of the inverted pendulum", *Computational Intelligence: An International Journal*, 11, 348–356.

[67] Raś, Z., Kudelska, A., and Chilumula, N., (1995), "Can we simplify international physical performance test profile using rough sets approach?", in: P.P. Wang (ed.), *Second Annual Joint Conference on Information Sciences PROCEEDINGS*, September 28 – October 1, Wrightsville Beach, North Carolina, USA, 393–396.

[68] Shenoi, S., (1995), "Rough sets in fuzzy databases", in: P.P. Wang (ed.), *Second Annual Joint Conference on Information Sciences PROCEEDINGS*, September 28 – October 1, Wrightsville Beach, North Carolina, USA, 263–264.

[69] Skowron, A., (1995), "Extracting laws from decision tables", *Computational Intelligence*, 11/2, 371–388.

[70] Skowron, A., (1995), "Synthesis of adaptive decision systems from experimantal data", in: A. Aamadt and J. Komorowski (eds.), *Proc. of the Fifth Scandinavian Conference on Artificial Intelligence SCAI-95*, Amsterdam: IOS Press , 220–238.

[71] Skowron A., and Grzymała-Busse, J.W., (1994), "From rough set theory to evidence theory", in: R.R Yaeger, M. Fedrizzi and J. Kacprzyk (eds.), *Advances in the Dempster Shafer Theory of Evidence*, John Wiley & Sons, Inc., New York, Chichester, Brisbane, Toronto, Singapore, 193–236.

[72] Skowron A.,and Rauszer, C., (1992), "The discernibility matrices and functions in information systems", in: R. Słowiński (ed.), *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory*, Kluwer Academic Publishers, Dordrecht, 311–362.

[73] Skowron, A., and Stepaniuk, J., (1994), "Decision rules based on discernibility matrices and decision", in: T.Y. Lin (ed.), *The Third International Workshop on Rough Sets and Soft Computing Proceedings* (*RSSC'94*), San Jose State University, San Jose, California, USA, November 10–12, 602–609.

[74] Słowiński, R., (ed.), (1992), *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory*, Kluwer Academic Publishers, Dordrecht.

[75] Słowiński, R., (1993), "Rough set learning of preferential attitude in multi-criteria decision making", in: J. Komorowski and Z.W. Raś (eds.), *Methodologies for Intelligent Systems. Lecture Notes in Artificial Intelligence Vol. 689*, Springer-Verlag, Berlin, 642–651.

[76] Słowiński, R. (1995). "Rough Set Approach to Decision Analysis", *AI Expert*, 10, 18–25.

[77] Słowiński, R., and Stefanowski, J., (1992), "RoughDAS and RoughClass' software implementations of the rough sets approach", in: R. Słowiński (ed.), *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory*, Kluwer Academic Publishers, Dordrecht, 445–456.

[78] Słowiński R., and Stefanowski J., (1994), "Rough classification with valued closeness relation", in: E.Diday, Y.Lechevallier, M.Schrader, P.Bertrand and B.Burtschy (eds.), *New Approaches in Classification and Data Analysis*, Springer-Verlag, Berlin, 482–489.

[79] Słowiński, R., and Zopounidis, C., (1993), "Applications of the rough set approach to evaluation of bankruptcy risk", Working Paper 93-08, Decision Support System Laboratory, Technical University of Crete, Chania, June.

[80] Słowiński, R., and Zopounidis, C., (1994), "Rough set sorting of firms according to bankruptcy risk", in: M. Paruccini (ed.), *Applying Multiple Criteria Aid for Decision to Environmental Management*, Kluwer, Dordrecht, Netherlands, 339–357.

[81] Słowiński, K., (1992), "Rough classification of HSV patients", in: R. Słowiński (ed.), *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory*, Kluwer Academic Publishers, Dordrecht, 77–93.

[82] Słowiński, K., Słowiński, R., and Stefanowski, J., (1988), "Rough sets approach to analysis of data from peritoneal lavage in acute pancreatitis", *Medical Informatics*, 13/3, 143–159.

[83] Słowiński, K., and Sharif, E.S., (1993), "Rough sets approach to analysis of data of diatnostic peritoneal lavage applied for multiple injuries patients", in: W. Ziarko (ed.), *Rough Sets, Fuzzy Sets and Knowledge Discovery. Proceedings of the International Workshop on Rough Sets and Knowledge Discovery* (*RSKD'93*), Banff, Alberta, Canada, October 12–15, Springer-Verlag, Berlin, 420–425.

[84] Słowiński, K., Stefanowski, J., Antczak, A. and Kwias, Z., (1995), "Rough sets approach to the verification of indications for treatment of urinary stones by extracorporeal shock wave lithotripsy (ESWL)", in: T.Y. Lin and A.M. Wildberger (eds.), *The Third International Workshop on Rough Sets and Soft Computing Proceedings* (*RSSC'94*), San Jose State University, San Jose, California, USA, November 10–12, 93–96.

[85] Stefanowski, J., and Vanderpooten, D., (1993), "A general two-stage approach to inducing rules from examples", in: W. Ziarko (ed.), *Rough Sets, Fuzzy Sets and Knowledge Discovery. Proceedings of the International Workshop on Rough Sets and Knowledge Discovery* (*RSKD'93*), Banff, Alberta, Canada, October 12–15, Springer-Verlag, Berlin, 317–325.

[86] Szladow, A. J., and Ziarko, W., (1992), "Knowledge-based process control using rough sets", in: R. Słowiński (ed.), *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory*, Kluwer Academic Publishers, Dordrecht, 49–60.

[87] Szladow, A., (1993), "Datalogic/R: Mining the knowledge in databases", *PC AI* , 7/1, 40–41.

[88] Szladow, A., and Ziarko W., (1993), "Rough sets: Working with imperfect data", *AI Expert*, 7, 36–41.

[89] Tanaka, H., Ishibuchi, H., and Shigenaga, T., (1992) "Fuzzy inference system based on rough sets and its application to medical diagnostic", in: R. Słowiński (ed.), *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory*, Kluwer Academic Publishers, Dordrecht, 111–117.

[90] Teghem, J., and Charlet J.-M., (1992), "Use of 'rough sets' method to draw premonitory factors for eathquakes by emphasing gas geochemistry: The case of a low seismic activiti context, in Belgium", in: R. Słowiński (ed.), *Intelligent Decision*

Support. Handbook of Applications and Advances of the Rough Set Theory, Kluwer Academic Publishers, Dordrecht, 165–180.

[91] Tsumoto, S., and Tanaka, H., (1994), "Characterization of structure of decision trees based on rough sets and greedoid theory", in: T.Y. Lin (ed.), *The Third International Workshop on Rough Sets and Soft Computing Proceedings* (*RSSC'94*), San Jose State University, San Jose, California, USA, November 10–12, 450–461.

[92] Tsumoto, S., and Tanaka, H., (1993), "PRIMEROSE: "Probabilistic rule induction method based on rough set theory", in: W. Ziarko (ed.), *Rough Sets, Fuzzy Sets and Knowledge Discovery. Proceedings of the International Workshop on Rough Sets and Knowledge Discovery* (*RSKD'93*), Banff, Alberta, Canada, October 12–15, Springer-Verlag, Berlin, 274–281.

[93] Tsumoto, S., and Tanaka, H., (1995), "PRIMEROSE: "Probabilistic rule induction method based on rough set and resampling methods", *Computational Intelligence: An International Journal*, 11/2, 389–405.

[94] Tsumoto, S., and Tanaka, H., (1995), "Induction of expert system rules from databases based pm rough set theory and resampling methods", in: P.P Wang (ed.), *Second Annual Joint Conference on Information Sciences PROCEEDINGS*, September 28 – October 1, Wrightsville Beach, North Carolina, USA.

[95] Wang, P.P., (ed.), (1995), *Second Annual Joint Conference on Information Sciences PROCEEDINGS*, September 28 – October 1, Wrightsville Beach, North Carolina, USA.

[96] Woolery, L.K. and 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.

[97] Wróblewski, J., (1995), "Finding minimal reducts using genetic algorithm (extended version)", in: P.P. Wang (ed.), *Second Annual Joint Conference on Information Sciences PROCEEDINGS*, September 28 – October 1, Wrightsville Beach, North Carolina, USA, 186–189.

[98] Wu, B.H., (1995), "An intelligent tutoring system using a rough set approach", in: P.P. Wang (ed.), *Second Annual Joint Conference on Information Sciences PROCEEDINGS*, September 28 – October 1, Wrightsville Beach, North Carolina, USA, 409–412.

[99] Ziarko, W., (1989), "Data analysis and case-based expert system development tool 'Rough'", in: *Proc. Case-Based Reasoning Workshop*, Morgan Kaufmann, Los Altos, CA, 356–361.

[100] Ziarko, W., (1991), "The discovery, analysis and representation of data dependencies in databases", in: G. Piatetsky-Shapiro, and W. J. Frawley (eds.),*Knowledge Discovery in Databases*, AAAI Press/MIT Press, 177–195.

[101] Ziarko, W., (1992), "Acquisition of control algorithms from operation data", in: R. Słowiński (ed.), *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory*, Kluwer Academic Publishers, Dordrecht, 61–75.

[102] Ziarko, W., (1993), "Analysis of uncertain information in the framework of variable precision rough sets", *Foundations of Computing and Decision Sciences* 18, 3/4 381–396.

[103] Ziarko, W., (ed.), (1993), *Rough Sets, Fuzzy Sets and Knowledge Discovery. Proceedings of the International Workshop on Rough Sets and Knowledge Discovery* (*RSKD'93*), Banff, Alberta, Canada, October 12–15, Springer-Verlag, Berlin.

[104] Ziarko, W. (1993). "Variable Precision Rough Set Model", *Journal of Computer and System Sciences*, 40, 39–59.

[105] Ziarko, W., Golan, R., and Edwards, D., (1993), "An application of DATALOGIC/R knowledge discovery tool to identify strong predictive rules in stock market data" in: *Proc. AAAI Workshop on Knowledge Discovery in Databases*, Washington, DC, 89–101

[106] Ziarko, W., and Katzberg, J., (1989), "Control algorithms acquisition, analysis and reduction: machine learning approach", in: *Knowledge-Based Systems Diagnosis, Supervision and Control*, Plenum Press, Oxford, 167–178.