ISSN: 2321-7782 (Online)
Volume 4, Issue 1, January 2016
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Rough Set Reasoning Based Classification Model Generating Decision Rule on Early Stage of Chronic Kidney Disease

Vidyullatha Pellakuri[1]
Research scholar, Department of CSE
KL University
Guntur, Andhra Pradesh, India

D. Rajeswara Rao[2]
Department of Computer Science & Eng
KL University
Guntur, Andhra Pradesh, India

Abstract: ***For handling imperfect data, the rough set founded classification model is one of the methods for decision rule making in intelligent systems. This paper presents classification process utilizing rough set conception as part of data mining. In view f this point, present paper proposes a rough set procedure for generating decision rules from the 400 sufferers on early stage of chronic kidney sickness. This gain knowledge of tested that the speculation of rough sets by all accounts a useful decision making for knowledge learning and large support for constructing expert systems.***

Key words: classification; Data mining; Decision rule; Rough set theory; Reduction; Rule generation.

I. INTRODUCTION

Z. Pawlak was offered Rough set theory in 1982. Grzymala-Busse 1988 showed rough set theory [1] represents an objective approach to imperfections in data. Its methodology is concerned with the classification and analysis of imprecise, unsure or incomplete information and knowledge and of it is considered as one of the vital non-statistical approaches in data analysis. Rough set theory constitutes a sound basis for decision support system and data mining. It offers method to the problem of discretization, attribute selection, data reduction and decision rule generation. Rough set on classification strategy is effectively utilized to real world utility. Rough set concept for knowledge mining offers valuable software tools for discovering patterns hidden in information in many elements and excellent potential to handle qualitative data. Rough Set [3] can be used in exceptional phases of the knowledge discovery process, as attribute selection, attribute extraction, information reduction, decision rule generation and pattern extraction. Moreover, latest extensions of rough set theory have brought new methods of decomposition in tremendous data sets, information mining in dispensed and multi-agent based environments & granular computing. It entails mechanisms for outlining partial memberships [2] of sets, however does not introduce extra measures of chances or degrees of membership. The basic proposal is that there's some expertise or data associated with each object in the universe of discourse. Based on this understanding, it's possible to tell one of the objects apart, at the same time others are inconceivable to differentiate. The latter objects are indiscernible [4] from every other, and kind a set. Each set of indiscernible objects is a potential granule, and so they kind the building blocks for expertise in regards to the universe. The rough set neighborhood has been an extraordinarily lively study neighborhood considering the fact that its inception within the eighties, and a massive quantity of rough set [5] ways for knowledge discovery and knowledge mining had been developed. The entire competencies discovery process has been area to study and a huge variety of contributions has been made. Knowledge mining technological know-how data supplies a new thought for organizing and managing big knowledge. Rough set [6] conception is one of the principal ways for competencies discovery. This approach can analyze intact information, receive unsure advantage and offer potent software by using reasoning. Information mining with rough set is a multi-phase method consisted of probably: discretization; Reduct and rules iteration on training set; classification on test set. Rough Set concept, considering it used to be put ahead, has been greatly utilized in information Mining, and it has essential functions in the expression, be trained and

conclusion of uncertain competencies [7]. The essential focal point is to show how rough set approaches may also be employed as process to the challenge of information mining and knowledge extraction.

The rough set methodology has vital significance to AI and cognitive sciences, notably in computer studying, know-how obtaining, and choice rule iteration, knowledge discovery [8] from databases, knowledgeable techniques, inductive reasoning and sample awareness. It appears of particular importance to determination aid methods and data mining [9]. The main ability of rough set theory is that it does not need any preliminary or further know-how about information. The rough set concept has been successfully applied in many real-life problems in remedy, pharmacology, engineering, banking, fiscal and market evaluation & others. The rough set approach [10] seems also important for various engineering applications, like diagnosis of machines using vibro-acoustics symptoms (noise, vibrations) and process control. This paper organized the rough set theory analyzed using Rough set exploration system RSES2.2 software tool. The subject of this paper is to present the Rough Set Theory, important concepts, and software tools for data mining, special applications in analysis of data in chronic kidney diagnosis.

II. ROUGH SET THEORY CONCEPT

Over the years rough Set theory has emerged as a valuable tool [11] in the resolution of more than a few issues, comparable to: representation of unsure or imprecise knowledge; expertise evaluation; evaluation of quality and availability of knowledge with respect to consistency and presence of data patterns [12]; identification and evaluation of data dependency; reasoning headquartered an uncertain and Reduct of knowledge. The rough set technique may also be viewed as a formal framework for locating details from imperfect data [13]. The results of the rough set process are presented within the type of classification or decision rules. In a data table, the columns are named as attributes the place rows as objects and entries are values of an attribute. Almost always, there are two types of attributes namely condition and decision attributes [14] where the rows of a decision table are known as "if...Then..." *decision rules*, which offer conditions are primary to make selections indicated through the decision attributes. In rough set the essential idea is the lower and higher approximations [15] of a set which is being the formal classification of advantage regarding the interest domain. In rough set, the every subset is outlined through higher and lower approximations. The minimize approximations are generated by way of subset which is characterized by means of objects a good way to type a part of an interest subset, whereas the upper approximations [16] are characterized via objects so one can very likely form a part of an interest subset. Formally, an information system IS can also be seen as a procedure so that [17] IS = (U, A) Where U is the universe where U=(x1, x2……xn) and A is the set of attributes. Each attribute aϵ A defines an information functionfa:$U \rightarrow Va$, where Va is the set of values of a, referred to as the domain of attribute a. In indiscrenibility Relation For each set of attributes B $\subseteq$ A, an indiscernibility relation IND(B) is outlined within the following means: two objects, x_i and x_j , are indiscernible by way of the set of attributes B in A, if b(x_i) = b(x_j) for every b $\subset$ B. The equivalence class of IND(B) is known as elementary set in B considering the fact that it represents the smallest discernible groups of objects. For any element xi of U, the equivalence class of x_i in relation IND(B) is represented as [x_i]IND(B) . The construction of elementary sets is the first in classification with rough set. The elements that doubtlessly belong to the set, and the elements the possibly belong to the set. Let X denotes the subset of elements of the universe U (X $\subseteq$ U). The lower approximation of X in B(B $\subseteq$A) , denoted as $\underline{BX}$, is defined as the union of all these elementary sets which are contained in X [18]. More formally: $\underline{BX}$={$x_i \in U$ |[x_i]$Ind(B) \subset X$} The above statement is to be read as: the lower approximation of the set X is a set of object x_i , which belong to the elementary sets contained in X (in the space B).The upper approximation of the set X, denoted as BX, is the union of these elementary sets, which have a non-empty intersection with X: BX={$x_i \in U$ |[x_i]IND(B)$\bigcap X \neq 0$}, For any object x_i of the lower approximation of X (i.e,$xi \in \underline{BX}$) , it is certain that it belong to X. for any object x_i of the upper approximation of X(i.e,$xi \in$BX), we can only say that x_i may belong to X. The difference: BNX = BX – $\underline{BX}$ is called a boundary of X in U. The Accuracy of Approximation [19] measure of the set X in B$\subseteq$ A is defined as µB(X) = card ($\underline{BX}$)/card(BX) The cardinality of a set is the number of objects contained in the lower (upper) approximation of the set X as one can notice $0 \leq \mu B$ (X) ≤ 1. If X is definable in U theμB (X) = 1, if X is indefinable in U then μB (X) < 1. Core and Reduct

[20] of Attributes in rough set theory, information table is used for describe of object in the universe, it consist of two dimensions, each row is an object, and each column is an attribute. Rough set theory classifies attribute in two types according to their roles of information table: core attribute and redundant attribute [21]. Here the minimum condition attributes set can be received, which is called reduction. One information table might have a several different reduction simultaneously. The intersection of the reduction is the core [22] of information table and the core attribute are the important attribute that influences attribute classification. A subset B of a set of attribute C is the reduction of C with respect to R if and only if $(R) = POSC\ (R)$, and $POSB-\{a\}\ (R) \neq POSC\ (R)$, for any a B. And the core defined by the equation given below $COREC\ (R) = \{c \in C\ | \forall c \in C, POSC\ (R)\}$.

III. Experimental Results

Illustration: For implementation of rough set theory [23], chronic kidney disease database is used from the UCI database to generate classification decision rule model. Database consists 400 instances and 24 condition attributes reminiscent of age in years, blood pressure(bp in mm/Hg), specific gravity(sg - 1.005,1.010,1.015,1.020,1.025), albumin(al-0,1,2,3,4,5), sugar(su - 0,1,2,3,4,5), red blood cells(rbc in normal, abnormal), puscell(pc-normal, abnormal), pus cell clumps(pcc - present, not present), bacteria(ba-present, not present), blood glucose random(bgr in mgs/dl), blood urea(bu in mgs/dl), serum creatinine(sc in mgs/dl),sodium(sod in mEq/L), potassium(pot in mEq/L), hemoglobin (hemo in gms), packed cell volume(pcv numerical), white blood cell count(wc in cells/cumm), hypertension(htn- yes,no), diabetes mellitus(dm - yes, no), coronary artery disease(cad - yes, no) appetite(appet - good, poor), pedal edema(pe - yes, no), anemia(ane - yes, no) and one decision attribute class - (ckd, notckd). The dataset consists of missing attribute values which are denoted by "?". We have built classification model using RSES2.2 to categorize a diagnosing early stages of chronic kidney disease in the following section. This section presents the results of experiments in which the rough set classifier [24] is used diagnosing the early stages of chronic kidney disease. Each of 400 records of database describes representative of chronic kidney disease who is checked or not checked (two classes) and is characterized by 24 features. The ensemble shows the best performance in case of missing data. RSES 2.2 - Rough Set Exploration System [25] is a software tool which has been used here that provides means for analysis of tabular data sets with use of various methods, in particular those based on Rough Set Theory.

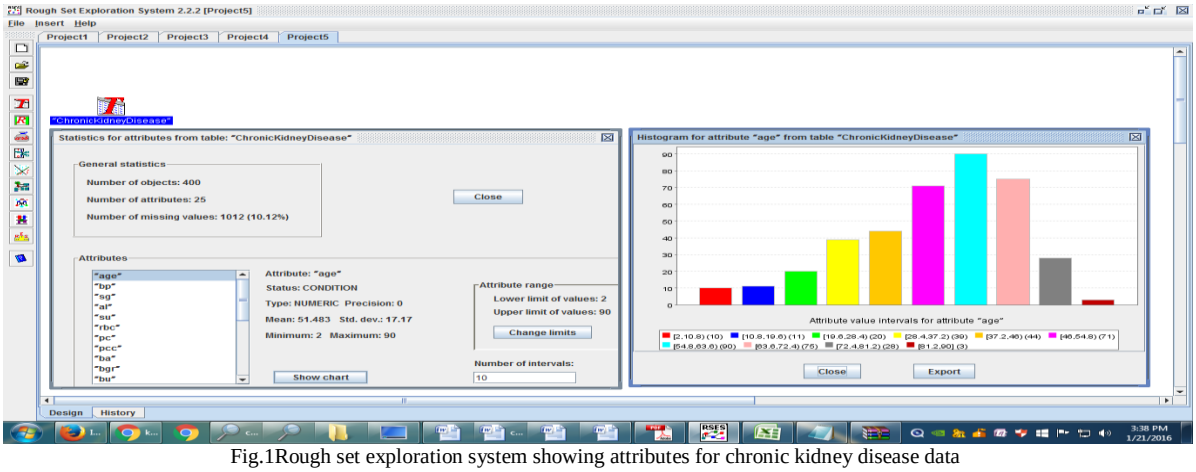


Fig.1Rough set exploration system showing attributes for chronic kidney disease data

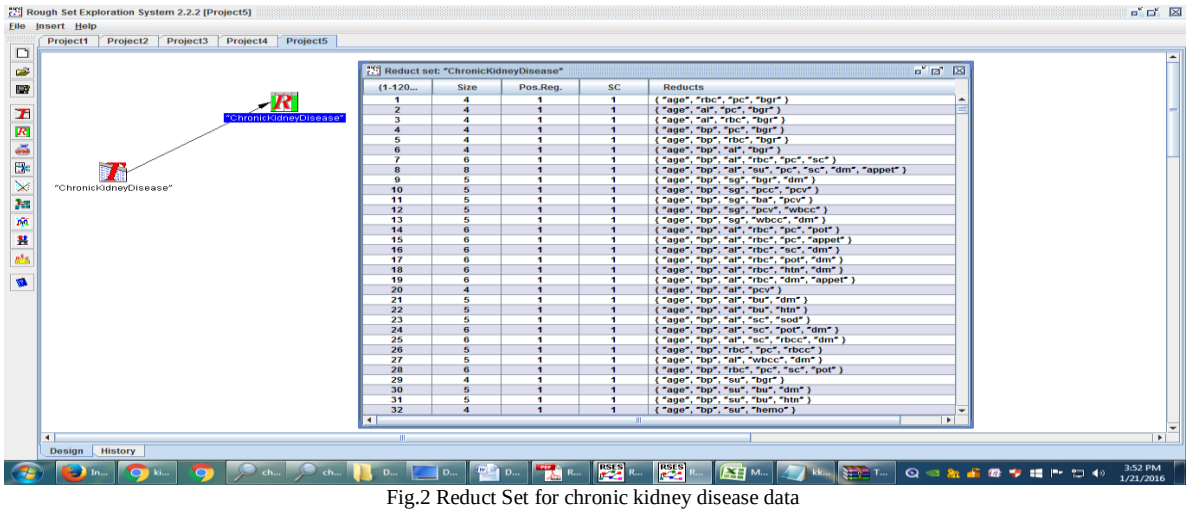


Fig.2 Reduct Set for chronic kidney disease data

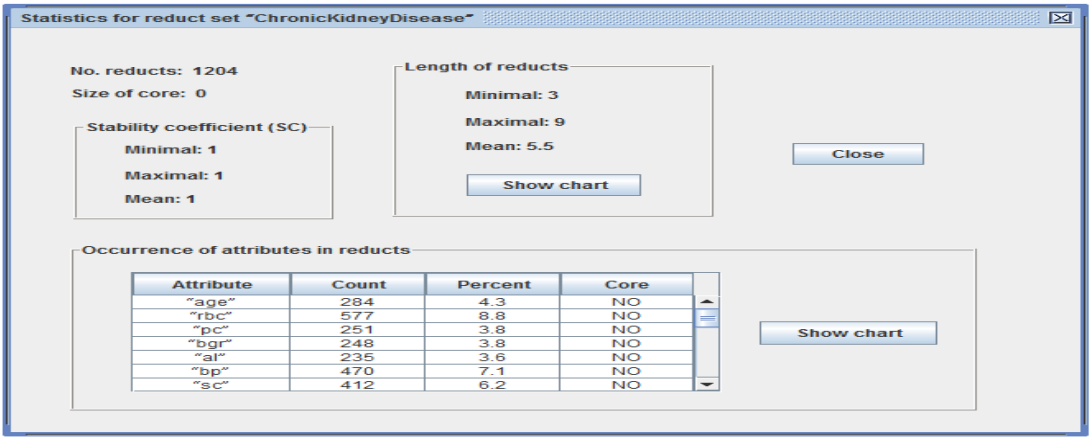


Fig.3 Statistics for Reduct Set on chronic kidney disease data

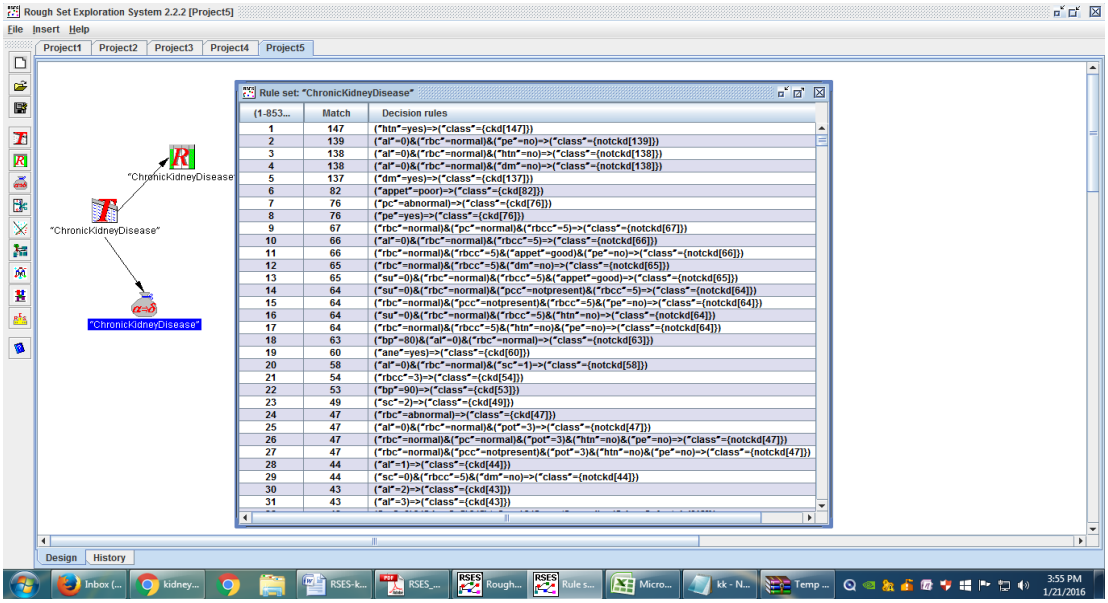


Fig.4 Rule Set for chronic kidney disease data

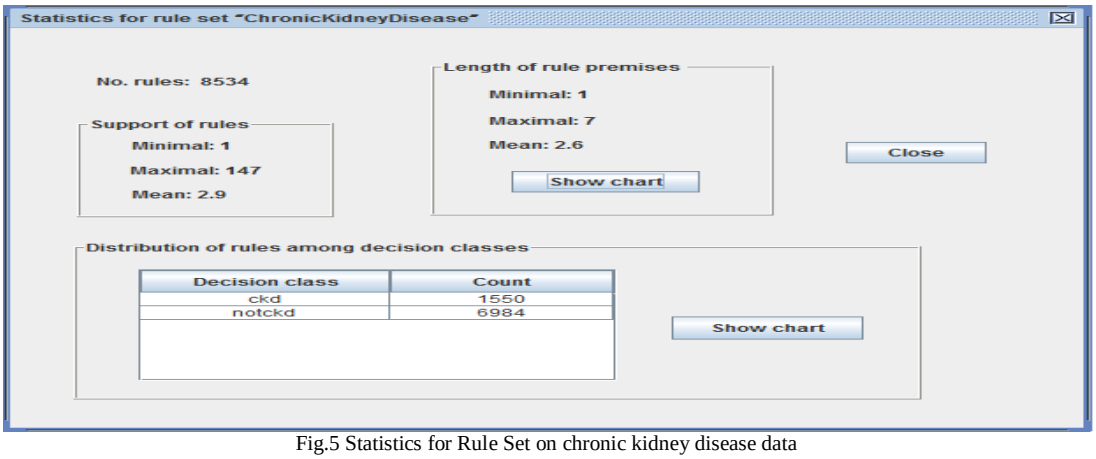
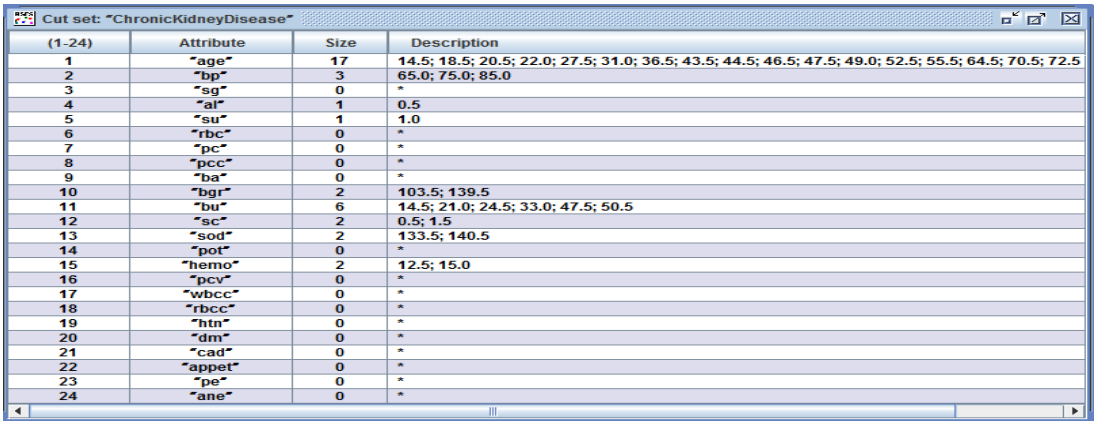


Fig.5 Statistics for Rule Set on chronic kidney disease data



(1-24)	Attribute	Size	Description
1	"age"	17	14.5; 18.5; 20.5; 22.0; 27.5; 31.0; 36.5; 43.5; 44.5; 46.5; 47.5; 49.0; 52.5; 55.5; 64.5; 70.5; 72.5
2	"bp"	3	65.0; 75.0; 85.0
3	"sg"	0	*
4	"al"	1	0.5
5	"su"	1	1.0
6	"rbc"	0	*
7	"pc"	0	*
8	"pcc"	0	*
9	"ba"	0	*
10	"bgr"	2	103.5; 139.5
11	"bu"	6	14.5; 21.0; 24.5; 33.0; 47.5; 50.5
12	"sc"	2	0.5; 1.5
13	"sod"	2	133.5; 140.5
14	"pot"	0	*
15	"hemo"	2	12.5; 15.0
16	"pcv"	0	*
17	"wbcc"	0	*
18	"rbcc"	0	*
19	"htn"	0	*
20	"dm"	0	*
21	"cad"	0	*
22	"appet"	0	*
23	"pe"	0	*
24	"ane"	0	*

Fig.6 Cut set for chronic kidney disease data

IV. CONCLUSION

In data analysis the rough set methodology has numerous imperative favorable circumstances which permits both qualitative and quantitative data, identifies relationships that would not be found using statistical methods, provides efficient algorithms for discovering hidden patterns in data, finds minimal sets of data i.e., data reduction, evaluates significance of data and generates sets of decision rules from data. The efficient software development for rough set based data analysis is rough set exploration system RSES2.2 particularly for large collections of data. In this context 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 future interesting seems to be study of the relationship between neural network and rough set approach to feature extraction from data.

References

1. Atiq Islam, Syed M. S. Reza, and Khan M. Iftekharuddin, Senior Member, IEEE, "Multifractal Texture Estimation for Detection and Segmentation of Brain Tumors" IEEE Transactions On Biomedical Engineering, vol-60, NO.11, NOVEMBER 2013.

2. Vidyullatha p, D Rajeswara Rao, "Designing and Simulating of soft computing method oriented on fuzzy logic systems" international journal of current research, vol-7, issue 09, pp.20539-20543, September 2015.

3. T.Rajesh, R. Suja Mani Malar "Rough Set Theory and Feed Forward Neural Network Based Brain Tumor Detection in Magnetic Resonance Images" Proceedings of the International Conference on Advanced Nanomaterials & Emerging Engineering Technologies" (ICANMEET-20 13).

4. L.-Xiang-wei,Q.Yian-fang " A Data Preprocessing algorithm for classification Model Based on rough sets", International conference on Solid state Device and Material Science (2012).

5. Y. Yao, "Notes on rough set approximations and associated measures", Journal of zhejiang ocean university(natural science), vol. 29, no. 5,(2010), pp. 399-410.

6. Marcin Korytkowski, Robert Nowicki, Rafał Scherer, Leszek Rutkowski, "Ensemble of Rough-Neuro-Fuzzy Systems for Classification with Missing Features" 2008 IEEE, 2008 IEEE International Conference on Fuzzy Systems (FUZZ 2008)

7. Y. Yao and Y. Zhao, "Attribute reduction in decision-theoretic rough set models", Information sciences, vol.178, no. 17, (2008), pp. 3356-3373.

8. S Kotsiantis, D. Kanellopoulos and P.Pintelas, "Data preprocessing for supervised leaning," Internal Journal of Computer Science, (2006), vol.1, no.2.

9. S. Xiao-xue, "Rough sets theory and its application," journal of xianyang normal university, vol. 20, no.2, (2005), pp. 30–35.

10. T. Jian-guo, T. Ming-shu, "On finding core and reduction in rough set theory", Control and Decision, vol. 18, no.4, (2003), pp. 449–457.

11. B. Walczak, D. L. Massart, "tutorial: rough set theory", Chemometrics and intelligent laboratory system, vol. 47, (1999),pp. 1-16.

12. T.Y. Lin and N. Cercone, Rough Sets and Data Mining - Analysis of Imperfect Data, Kluwer Academic Publishers, Boston, London, Dordrecht, 1997, P.430.

13. Wang, Joint Conference of Information Sciences, Vol. 3. Rough Sets and Computer Sciences, Duke University, 1997, P.449.

14. T.Y. Lin and N. Cercone, Rough Sets and Data Mining - Analysis of Imperfect Data, Kluwer Academic Publishers, Boston, London, Dordrecht, 1997, P.430.

15. T.Y. Lin and A.M.Wildberger, The Third International Workshop on Rough Sets and Soft Com-puting Proceedings RSSC'94), San Jose State University, San Jose, California, USA, 1995.

16. A.G. Jackson, M. Ohmer and H. Al-Kamhawi, in The Third International Workshop on Rough Sets and Soft Computing Proceedings (RSSC'94), (T.Y. Lin ed.), San Jose State University, San Jose, California, USA, 1994.

17. W. P Ziarko (Ed.), Rough Sets, Fuzzy sets and Knowledge Discovery, Springer, New York, (1994).

18. 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, 1993, P.476.

19. A. Skowron and C. Rauszer, in Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory, (R. S³owiński ed.), Kluwer Academic Publishers, Boston, London, Dordrecht, 1992, P.471.

20. J.W. Grzyma³a-Busse, in Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory, (R. S³owiński ed.), Kluwer Academic Publishers, Boston, London, Dordrecht, 1992, P.471. 2.

21. R. Slowinski (Ed.), "Intelligent Decision Support", Handbook of Application and Advances of the rough set theory, Kluwer Academic Publishers, Dorderecht, (1992).

22. Z. Pawlak, Rough Sets - Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Boston, London, Dordrecht, 1991, P.229.

23. Di Kai-chang Li De-ren and Li De-yi, "Rough set theory and its application in attribute analysis and kenowledge discovery in gis", Journal of wuhan technical university of surveying and mapping, vol. 24, no.1, (1990), pp. 1–10.

24. Z. Pawlak, "Rough set", Int.j.Inf. Comput. Sci, vol. 11, (1982), pp. 341-356.

25. http://logic.mimuw.edu.pl/rses

