# A Probabilistic Rough Set Approach to Rule Discovery

Hossam A. Nabwey
*Department of Engineering Basic science, Faculty of Engineering,*
*Menofia University, Menofia, Egypt*
*eng_hossam21@yahoo.com*

***Abstract***

*Rough set theory is a relative new tool that deals with vagueness and uncertainty inherent in decision making. This paper introduce a new probabilistic approach for reducing dimensions and extracting rules of information systems using expert systems. The core of the approach is a soft hybrid induction system called the Generalized Distribution Table and Rough Set System (GDT-RS) for discovering classification rules, Which is based on a combination of Generalized Distribution Table (GDT) and the Rough Set methodologies. The probabilistic properties of the Decision rules are discussed and the proposed probabilistic rough set approach was applied to discover grade rules of transformer evaluation when there is a missing failure symptom of transformer. The results show that the proposed approach represents explicitly the uncertainty of a rule, it can flexibly select biases for search control and it can effectively handle noisy and missing data.*

***Keywords:*** *Knowledge discovery, Rough set, Rule extraction, Attributes Reduction Missing attribute values, Generalized Distribution Table (GDT), Rule discovery*

## 1. Introduction

Classical rough set theory developed by Professor Z. Pawlak in 1982 has made a great success in knowledge acquisition in recent years [1]. In Rough set theory, knowledge is represented in information systems. An information system is a data set represented in a table, this table is called decision table [2]. Each row in the table represents an object, for example a case or an event. Each column in the table represents an attribute, for instance a variable, an observation or a property. To each object (row) there are assigned some attribute values.

One of the disadvantages of rough set theory is its dependence on complete information systems i.e., A decision table to be processed must be complete and its all objects values must be known [3]. But in real-life applications, Due to measurement errors, miscomprehension, access limitation and misoperation in register, etc, information systems with missing values often occur in knowledge acquisition. Information systems with missing data, or, in different words, the corresponding decision tables are incompletely specified, is called incomplete information systems [4]. For simplicity, incompletely specified decision tables will be called incomplete decision tables.

The core of the proposed approach is a soft hybrid induction system called the Generalized Distribution Table and Rough Set System (GDT-RS) for discovering classification rules. The system is based on a combination of Generalized Distribution Table (GDT) and the Rough Set methodologies.

## 2. Rough Set and Missing Attribute Values

Missing attribute values commonly exist in real world data set. They may come from the data collecting process or redundant diagnose tests, unknown data and so on. Since the main concern is learning from examples, and an example with a missing decision value, (i.e., not classified) is useless [5], we will assume that only attribute values may be missing. Discarding all data containing the missing attribute values cannot fully preserve the characteristics of the original data. So In data analysis two main strategies are used to deal with missing attribute values in data tables.

***The former strategy*** is based on conversion of incomplete data sets (i.e., data sets with missing attribute values) into complete data sets and then acquiring knowledge. The process to change the incomplete data set into complete data set, say to transform the missing data into specified data via some technique, is called completeness of data set. Multiple approaches on filling in the missing attribute values were introduced [6],[7], such as selecting the “most common attribute value”, the “concept most common attribute value”, “assigning all possible values of the attribute restricted to the given concept”, “ignoring examples with unknown attribute values”, “treating missing attribute values as special values”, “event covering method” and so on. In this strategy conversion of incomplete data sets to complete data sets is a preprocessing to the main process of data mining.

***In the later strategy***, knowledge is acquired from incomplete data sets taking into account that some attribute values are missing. The original data sets are not converted into complete data sets. The later strategy is exemplified by the C4.5 approach to missing attribute values [8] or by a modified LEM2 algorithm [9, 10]. In both algorithms original data sets with missing attribute values are not preprocessed.

This paper will concentrate on the later strategy used for rule induction, i.e., it will be assumed that the rule sets are induced from the original data sets, with missing attribute values, not preprocessed as in the former strategy.

The next basic assumption is that there are three approaches to missing attribute values [11]:

***The first approach*** is that an attribute value, for a specific case, is lost. For example, originally the attribute value was known; however, due to a variety of reasons, currently the value is not available. Maybe it was recorded but later it was erased.

***The second approach*** is that an attribute value was not relevant , the case was decided to be a member of some concept, i.e., was classified, or diagnosed, in spite of the fact that some attribute values were not known. For example, it was feasible to diagnose a patient in spite of the fact that some test results were not taken (here attributes correspond to tests, so attribute values are test results). Since such missing attribute values do not matter for the final outcome, we will call them "do not care" conditions.

***The third approach*** is a partial "do not care" condition; we assume that the missing attribute value belongs to the set of typical attribute values for all cases from the same concept. Such a missing attribute value will be called an attribute-concept value. Calling it *concept "do not care" condition* would be perhaps better, but this name is too long

In the sequel it is assumed that all decision values are specified. Also, all missing attribute values are denoted either by "?" or by "*", or by "-", lost values will be denoted by "?", "do not care" conditions will be denoted by "*", and attribute-concept value will be denoted by "-". Additionally, it is assume that for each case at least one

attribute value is specified. An example of an incompletely specified table is presented in Table 1.

**Table 1: An incompletely Specified Decision Table**

| | Attributes | | | Decision |
|---|---|---|---|---|
| **transformer** | **Valid utilization degree** | **Maintenance cost** | **Reliability** | **Grade** |
| 1 | $a_1$ | $b_1$ | $c_1$ | II |
| 2 | - | $b_0$ | $c_0$ | I |
| 3 | $a_1$ | $b_0$ | $c_1$ | II |
| 4 | $a_1$ | $b_0$ | * | I |
| 5 | $a_0$ | $b_0$ | $c_0$ | II |
| 6 | $a_0$ | ? | - | II |
| 7 | $a_1$ | ? | $c_1$ | II |
| 8 | $a_1$ | $b_1$ | $c_0$ | I |

Obviously, in rough set theory any decision table defines a function ρ that maps the set of ordered pairs (case, attribute) into the set of all values [12]. For example, in Table 1, ρ (1, Valid utilization degree) = $a_1$

Rough set theory is based on the idea of an indiscernibility relation [13]. The indiscernibility relation IND(B) is an equivalence relation. Equivalence classes of IND(B) are called elementary sets of B and are denoted by [x]B.

The indiscernibility relation IND(B) may be computed using the idea of blocks of attribute-value pairs. Let a be an attribute and let v be a value of a for some case. For complete decision tables if t = (a, v) is an attribute-value pair then a block of t, denoted [t], is a set of all cases from U that for attribute a have value v.

For incomplete decision tables the definition of a block of an attribute-value pair must be modified as follow:

- If for an attribute $\boldsymbol{a}$ there exists a case x such that ρ(x, $\boldsymbol{a}$) = ?, i.e., the corresponding value is lost, then the case x should not be included in any block [( $\boldsymbol{a}$, v )] for all values v of attribute $\boldsymbol{a}$.
- If for an attribute $\boldsymbol{a}$ there exists a case x such that the corresponding value is a "do not care" condition, i.e., ρ (x, a) = *, then the corresponding case x should be included in blocks [(a, v)] for all specified values v of attribute a.
- If for an attribute a there exists a case x such that the corresponding value is a attribute-concept value, i.e., ρ (x, a) = –, then the corresponding case x should be included in blocks [(a, *v*)] for all specified values *v* of attribute $\boldsymbol{a}$ that are members of the set $V(x, \boldsymbol{a})$, where

$$V(x, \boldsymbol{a}) = \{ \rho (y, \boldsymbol{a}) \mid y \in U , \rho (y, d) = \rho (x, d)\}, \quad \text{and } d \text{ is the decision.}$$

Thus, For Table 1,

$$\begin{aligned} &[(\text{Valid utilization degree}, a_1)] = \{1, 2, 3, 4, 7, 8\}, \\ &[(\text{Valid utilization degree}, a_0)] = \{5, 6\}, \\ &[(\text{Maintenance cost}, b_1)] = \{1, 8\}, \\ &[(\text{Maintenance cost}, b_0)] = \{2, 3, 4, 5\}, \end{aligned} \tag{1}$$

[(Reliability, $^{c_1}$)] = {1, 3, 4, 6, 7},

[(Reliability, $^{c_0}$)] = {2, 4, 5, 6}.

These modifications of the definition of the block of attribute-value pair are consistent with the interpretation of missing attribute values [11] lost, "do not care" conditions, and attribute-concept values. Also, note that the attribute-concept value is the most universal, since if $V(x, a) = \emptyset$, the definition of the attribute-concept value is reduced to the lost value, and if $V(x, a)$ is the set of all values of an attribute $a$, the attribute-concept value becomes a "do not care" condition.

## 3. Generalized Distribution Table

Generalized Distribution Table (GDT) is a table in which the probabilistic relationships between concepts and instances over discrete domains are represented [14], [15]. Any GDT consists of three components: possible instances, possible generalizations of instances, and probabilistic relationships between possible instances and possible generalizations.

**The possible instances**, which are represented at the top row of GDT, are defined by all possible combinations of attribute values from a database, and the number of the possible instances is

$$\prod_{i=1}^{m} n_i \qquad (2)$$

Where m is the number of attributes, n is the number of different data values in each attribute.

**The possible generalizations** for instances, which are represented by the left column of a GDT, are all possible cases of generalization for all possible instances, and the number of the possible generalizations is

$$\left(\prod_{i=1}^{m}\left(n_i+1\right)\right) - \left(\prod_{i=1}^{m} n_i\right) - 1 \qquad (3)$$

A wild card ` * ' denotes the generalization for instances, For simplicity, the wild card will sometimes be omitted in the paper. For example, the generalization $a_0 * c_0$ means that the attribute b is superfluous (irrelevant) for the concept description. In other words, if an attribute b takes values from $\{b_0, b_1\}$ and both $a_0b_0c_0$ and $a_0b_1c_0$ describe the same concept, the attribute b is superfluous, i.e. the concept can be described by $a_0c_0$. Therefore, the generalization $a_0*c_0$ used to describe the set $\{a_0\ b_0\ c_0, a_0\ b_1\ c_0\}$

**The probabilistic relationships** between possible instances and possible generalizations, represented by entries $G_{ij}$ of a given GDT, are defined by means of a probabilistic distribution describing the strength of the relationship between every possible instance and every possible generalization. The prior distribution is assumed to be uniform if background knowledge is not available. Thus, it is defined by

$$G_{ij} = p(PI_j \setminus PG_i)$$
$$= \left\{ \begin{array}{ll} \frac{1}{N_{PG_i}} & \text{if } PG_i \text{ is a generalization of } PI_j \\ 0 & \text{otherwise} \end{array} \right\} \quad (4)$$

where

$PI_j$ is the j th possible instance,

$PG_i$ is the ith possible generalization,

and $N_{PG_i}$ is the number of the possible instances satisfying the ith possible generalization , that is ,

$$N_{PG_i} = \prod_j^m n_j \quad (5)$$

where j = 1,. . . , m, and j # the attribute that is contained by the ith possible generalization (i.e., j just contains the attributes expressed by the wild card ) .

**Rule Strength**

In this approach, the rules are expressed in the following form: $X \rightarrow$ Y with S That is, "if X then Y with strength S". Where

X: denotes the conjunction of the conditions that a concept must satisfy,

Y: denotes a concept that the rule describes, and

S: is a "measure of strength" of which the rule holds.

The strength of a given rule reflects the incompleteness and uncertainty in the process of rule inducing influenced by both unseen instances and noise. It is defined by

$$S(X \rightarrow Y) = s(X) \,.\, \left[1 - r(X \rightarrow Y)\right] \quad (6)$$

where **s**( X ) : The strength of the generalization X and r : noise rate function .

**s( X ) :** ***The strength of the generalization X*** (i.e., the condition of the rule) it represents explicitly the prediction for unseen instances. It is given by Eq. (7).

$$s(PG_i) = \sum_j p(PI_j \setminus PG_i) = \frac{N_{ins\text{-}rel,i}}{N_{PG_i}} \quad (7)$$

Where $N_{ins\text{-}rel,i}$ is the number of the observed instances satisfying the $i^{th}$ generalization.

***r : noise rate function***

It shows the quality of classification measured by the number of the instances satisfying the generalization X which cannot be classified into class Y. The user can specify an allowed noise level as a threshold value. Thus, the rule candidates with a noise level larger than the given threshold value will be deleted. It is defined by ,

$$r(X \to Y) = \frac{N_{ins-rel}(X) - N_{ins-class}(X,Y)}{N_{ins-rel}(X)} \tag{8}$$

where

$N_{ins-rel}(x)$ is the number of the observed instances satisfying the generalization X ,

$N_{ins-class}(X,Y)$ is the number of the instances belonging to the class Y within the instances satisfying the generalization X .

From the GDT, we can see that a generalization is 100% true if and only if all of instances belonging to this generalization appear. Let us use the example shown in Table 1. Considering the generalization $\{b_0 , c_1\}$, if instances both $\{a_0\ b_0\ c_1\}$ and $\{ a_1\ b_0\ c_1\}$ appear, the strength $s(\{b_0 , c_1\})$ is 1; if only one of $\{ a_0\ b_0\ c_1\}$ and $\{ a_1\ b_0\ c_1\}$ appears, the strength $s( \{b_0\ c_1\})$ is 0.5, as shown in Figure 1.

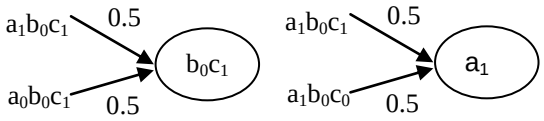


**Fig. 1. Probability of a Generalization Rule**

It is obvious that one instance can be expressed by several possible generalizations, and several instances can be also expressed by one possible generalization. For the example shown in Table 1, the instance $\{a\ b_0\ c_1\}$ can be expressed by $\{a_1\ b_0\}$, $\{b_0\ c_1\}$……., or $\{ c_1\}$.

Every generalization in upper levels contains all generalizations related to it in lower levels. That is,

$$\{a_1\} \supset \{a_1\ b_0\} , \{ a_1\ c_1\} ,$$

$$\{a_1\ b_0\} \supset \{ a_1\ b_0\ c_1\}$$

In other words, if the rule $\{a_1\} \to y$ is true, the rule $\{a_1\ b_0\} \to y$ and $\{ a_1\ c_1\} \to y$ are also true. Otherwise, if $\{a_1\ b_0\} \to y$ or $\{ a_1\ c_1\} \to y$ is false, the rule $\{a_1\} \to y$ is also false. Figure 2 gives the relationship among generalizations.

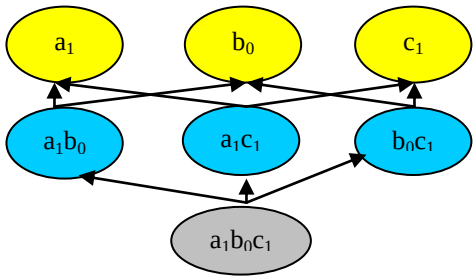


**Fig. 2 The Relationship Among Generalizations**

A generalization that contains the instances with different classes is contradictory, and it cannot be used as a rule. In contrast, a generalization that contains the instances with the same class is consistent, so From Table 1, we can see that the generalizations

can be divided into three groups: contradictory, belonging to class y, and belonging to class n.

## 4. Searching Algorithm for an Optimal Set of Rules

We now outline the idea of a searching algorithm for a set of rules based on the GDT-RS methodology. a sample decision table shown in Table 1 is used to illustrate the idea.

**Algorithm**

***Step 1.*** ***Create the GDT***.

Since :

over-current protection action $\in \{ a_0 , a_1 \} \Rightarrow n_1 = 2$

Exceeding of winding insulation resistance $\in \{ b_0 , b_1 \} \Rightarrow n_2 = 2$

Unbalance of three-phase winding direct current resistance $\in \{ c_0 , c_1 \} \Rightarrow n_3 = 2$

Hence :

the number of attributes ( m ) = 3 ,

from Eq.(2) number of the possible instances is 8 ,

from Eq.(3) number of the possible generalizations is 18 ,

***Step 2 . simplify the GDT.***

By deleting all of the instances and generalizations un-appeared in the example database shown in Table 1:

From table 1 The instances appeared with respect to cases 1, 3, 5, 8 are $\{a_1 b_1 c_1\}$, $\{a_1 b_0 c_1\}, \{a_0 b_0 c_0\}, \{a_1 b_1 c_0\}$, respectively.

From Eq. ( 1 ) and table 1 the instance appeared with respect to case 2 is $\{a_1 b_0 c_0\}$ ;

From Eq. ( 1 ) and table 1 the instance appeared with respect to case 4 may be one of $\{\{a_1 b_0 c_0\}, \{a_1 b_0 c_1\}\}$ ;

From table 1, the instance appeared with respect to case 6 may be one of $\{\{a_0 b_0 c_0\}, \{a_0 b_0 c_1\}, \{a_0 b_1 c_0\}, \{a_0 b_1 c_1\}\}$

Similarly, the instance appeared with respect to case 7 may be one of $\{\{a_1 b_0 c_1\}, \{a_1 b_1 c_1\}\}$ but $\{a_1 b_0 c_1\}$ is not consistent with table. 2 . so the appeared instance is $\{a_1 b_1 c_1\}$ .

So the simplified GDT is shown in table 2

**Table 2: The Simplified GDT for the Decision Table Shown in Table 1**
(Note the elements that are not displayed are all zero)

| | $a_0 b_0 c_0$ | $a_0 b_0 c_1$ | a0 b1 c0 | a0 b1 c1 | a1 b0 c0 | a1 b0 c1 | a1 b1 c0 | a1 b1 c1 |
|---|---|---|---|---|---|---|---|---|
| $* b_0 c_0$ | 1/2 | | | | 1/2 | | | |
| $* b_0 c_1$ | | 1/2 | | | | 1/2 | | |
| $* b_1 c_0$ | | | 1/2 | | | | 1/2 | |
| $* b_1 c_1$ | | | | 1/2 | | | | 1/2 |
| $a_0 * c_0$ | 1/2 | | 1/2 | | | | | |
| $a_0 * c_1$ | | 1/2 | | 1/2 | | | | |
| $a_1 * c_0$ | | | | | 1/2 | | 1/2 | |
| $a_1 * c_1$ | | | | | | 1/2 | | 1/2 |
| $a_0 b_0 *$ | 1/2 | 1/2 | | | | | | |
| $a_0 b_1 *$ | | | 1/2 | 1/2 | | | | |
| $a_1 b_0 *$ | | | | | 1/2 | 1/2 | | |
| $a_1 b_1 *$ | | | | | | | 1/2 | 1/2 |
| $* * c_0$ | 1/4 | | 1/4 | | 1/4 | | 1/4 | |
| $* * c_1$ | | 1/4 | | 1/4 | | 1/4 | | 1/4 |
| $a_0 * *$ | 1/4 | 1/4 | 1/4 | 1/4 | | | | |
| $a_1 * *$ | | | | | 1/4 | 1/4 | 1/4 | 1/4 |
| $* b_0 *$ | 1/4 | 1/4 | | | 1/4 | 1/4 | | |
| $* b_1 *$ | | | 1/4 | 1/4 | | | 1/4 | 1/4 |

***Step 3 . Group the Generalizations***

Generalizations can be divided into three groups contradictory, belonging to class yes, and belonging to class no. The contradictory generalizations, containing the instances belonging to different decision classes, cannot be used as the rules. Hence they are ignored. In other words, we are just interested in the generalizations belonging to class yes or no, which will be selected as the rules.

**Table 3. The Generalizations Belonging to Class Yes**

| | $a_1 b_0 c_0$ | $a_0 b_1 c_0$ | $a_1 b_1 c_1$ |
|---|---|---|---|
| $a_1 * c_0$ | 1/2 | 1/2 | |
| $a_1 b_1 *$ | | 1/2 | 1/2 |

**Table 4. The Generalizations Belonging to Class No**

| | $a_0 b_0 c_0$ | $a_0 b_0 c_1$ | $a_0 b_1 c_0$ | $a_0 b_1 c_1$ |
|---|---|---|---|---|
| $a_0 * c_0$ | 1/2 | | 1/2 | |
| $a_0 * c_1$ | | 1/2 | | 1/2 |
| $a_0 b_0 *$ | 1/2 | 1/2 | | |
| $a_0 b_1 *$ | | | 1/2 | 1/2 |
| $a * *$ | 1/4 | 1/4 | 1/4 | 1/4 |

***Step 4. Rule Selection***

There are several possible ways for rule selection. For example :

- Selecting the rules that contain as many instances as possible.
- Selecting the rules in the levels of generalization as high as possible according to the number of " * " in a generalization .
- Selecting the rules with larger strengths.

Since the purpose is to simplify the decision table and simpler results of generalization (i.e., more general rules) are preferred, the first priority will be to the rules that contains more instances, then to the rules corresponding to an upper level of generalization. and the third priority to The rules with larger strengths .

Thus, from table 3 and table 4 the final rule set is

$$\{a_1 c_o\} \rightarrow yes \ , \ \text{with S} = 1$$
$$\{a_1 b_1\} \rightarrow yes \ , \ \text{with S} = 1$$
$$\{a_0\} \rightarrow no \ , \ \text{with S} = 1$$

***Results***

The induced Rules can be written as:

- **If** (Valid utilization degree, $a_1$ ) and (Maintenance cost, , not appearing) **then** (Grade, II)
- **If** (Valid utilization degree, $a_1$ ) and (Maintenance cost, $b_1$) **then** (Grade, II)
- **If** (Valid utilization degree, not appearing) **then** (Grade, I)

## 5. Conclusions

- Rough set theory and statistics are related to analyze the data from the rough set perspective.
- Three approaches to missing attribute values are presented in a unified way. It is shown that all three approaches to missing attribute values may be described using the same idea of attribute-value blocks.
- An approach of rule discovery based on Rough Sets and Generalization Distribution Table was presented. The basic concepts and an implementation of the methodology was described. Main features of that methodology can be summarized as follows:
  - ✓ It can discover **If-Then** rules from very large, complex databases .

- ✓ It represents explicitly the uncertainty of a rule including the prediction of possible instances in the strength of the rule.
- ✓ Lost values are considered during the process of rule induction .
- ✓ It can flexibly select biases for search control.
- ✓ It can effectively handle noisy data, missing data

## References

[1] Guoyin Wang , Extension of Rough Set under Incomplete Information Systems ,National Science Foundation of china ( No. 69803014 ), PD program of P.R. China

[2] Grzymala-Busse J. W., Three Approaches to Missing Attribute Values - A Rough Set Perspective , Accepted for the Workshop on Foundations of Data Mining, associated with the fourth IEEE International Conference on Data Mining, Brighton, UK, November 1–4, 2004

[3] Grzymala-Busse J. W. and Wang A. Y., Modified algorithms LEM1 and LEM2 for rule induction from data with missing attribute values. Proc. of the Fifth International Workshop on Rough Sets and Soft Computing (RSSC'97) at the Third Joint Conference on Information Sciences (JCIS'97), Research Triangle Park, NC, March 2–5, 1997, 69–72.

[4] Grzymala-Busse J. W. and M. Hu, A comparison of several approaches to missing attribute values in data mining. Proceedings of the Second International Conference on Rough Sets and Current Trends in Computing RSCTC'2000, October 16–19, 2000, Banff, Canada, 340–347.

[5] Grzymala-Busse J. W., Data with missing attribute values: Generalization of indiscernibility relation and rule induction. Transactions on Rough Sets, Lecture Notes in Computer Science Journal Sub line, Springer-Verlag, vol. 1, 2004, 78–95.

[6] Grzymala-Busse J. W., On the unknown attribute values in learning from examples. Proc. of the ISMIS-91, 6th International Symposium on Methodologies for Intelligent Systems, Charlotte, North Carolina, October 16–19, 1991, 368–377, Lecture Notes in Artificial Intelligence, vol. 542, Springer-Verlag, Berlin, Heidelberg, New York, 1991.

[7] Grzymala-Busse J. W., Rough Set Strategies to Data with Missing Attribute Values , Proceedings of the Workshop on Foundations and New Directions in Data Mining, associated with the third IEEE International Conference on Data Mining, November 19–22, 2003, Melbourne, FL, USA, 56–63

[8] Marte Skarstein Bjanger , vibration Analysis in Rotating Machinery using Rough Set theory and ROSETTA . Tech. report, Univ. of Norwegian, 1999

[9] Zhong N. and Obsuga S., Using Generalization Distribution Tables as a Hypotheses Search Space for Generalization, Proc. 4th International Workshop on Rough Sets, Fuzzy Sets, and Machine Discovery (RSFD-96), 1996, 396-403.

[10] Zhong N., Dong J. Z., and Ohsuga S., "Discovering Rules in the Environment with Noise and Incompleteness", Proc. 10th International Florida AI Research Symposium (FLAIRS-97) edited in the Special Track on Uncertainty in AI, 1997, 186-191.

[11] Zhong N., Dong J. Z., and Ohsuga S., "Soft Techniques to Rule Discovery in Data", Proc. the Fifth Congress on Intelligent Techniques and Soft Computing (EUFUT-97) edited in the Invited Session on Soft Techniques in Knowledge Discovery, 1997, 212-217.

[12] Zhong N., Fujitsu S. and Ohsuga S, Generalization Based on the Connectionist Networks Representation of a Generalization Distribution Table, Proc. First Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD-97), World Scientific, 1997, 183-197.

[13] Ning Zhong, Andrzej Skowron, A rough set-based knowledge discovery process , Int. J. Appl. Math. Comput. Sci., 2001, Vol.11, No.3, 603-619

[14] Weihua Zhu , Wei Zhang, Yunqing Fu , An Incomplete Data Analysis Approach using Rough Set Theory , Proceedings of the 2004 international Conference on Intelligent Mechatronics and Automation, Chengdu , China August 2004

[15] Pawlak Z., Busse J. G., R. Slowinski and Ziarko W., Rough Sets, Communication of the ACM, vol. 38:11 , pp. 89-95 , 1995

[16] Quinlan J. R., C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.