A Mathematic Method for Calculating Possible Rules

Jia-Yuarn Guo[#1] ,Ching-Chou Huang[*2]

[#]*Dept. of Industrial Engineering and Management,Nan Kai University of Technology Taiwan ROC*

[1]jxg87@nkut.edu.tw

[*]*Dept. of Information Science and Management,Nan Kai University of Technology Taiwan ROC*

[2]hccit@nkut.edu.tw

***Abstract*—Data mining is a process of extracting potentially and previously unknown information technique which have been applied to study the real-life research increasingly. Many techniques have been developed to extract decision rules from an incomplete information system. A key factor among them is to use different methods to manage the missing data (unknown values).In this paper, we propose a mathematic method based on the Binomial Distribution to calculate the plausibility and probability of possible rules from original incomplete information systems, especially for diagnosis database.**

***Keywords*：Data Mining, Rough sets, Decision rules, Incomplete information systems.**

1. INTRODUCTION

It is noted that due to the typically huge size of today's information systems, real-world data tend to be incomplete due to missing some values. Hence, discover knowledge from incomplete information systems has received more and more attention in recent years.

The rough sets theory, proposed by Pawlak [1], provides a natural method to cope with incomplete or inconsistent information which has been the mainly impediment to the classification and rule induction of objects [2, 3]. By using the concept of lower and upper approximation of rough set theory, knowledge invisible in information systems may be disclosed and expressed in the form of decision rules via an objective knowledge induction process for decision making. Various approaches have been proposed to induce decision rules from data sets taking the form of complete decision systems [4-11] Whereas, due to the abundant existence of incomplete information systems in real life, many applications developed extensions of Pawlak's rough set model to extract decision rules from an incomplete information system [12-18].A key factor among them is using different methods to manage the missing data (unknown values) . The simplest is removing the objects with unknown values [19]. More complex approaches which provide tactics to deal with null values in terms of statistics are proposed in [20-23]. For example, the techniques proposed in [23] predict the unknown values of an attribute on the basis of values of other attributes of an object and relevant class information. Another method deals with null values in the source system by replacing a set of assumed objects in the intention system. In truth, the converted complete system is the simple combination of all completions of the source system. All these methods all try to transform an incomplete system into a complete system by smoothing or extending the data.

Other groups of techniques deal with the incomplete systems without changing the size of the data sets or making assumption of the missing values [24, 25]. For example, Deng et al. [26], Hong et al. [27, 28], Jensen and Shen [29] and Wang et al. [30], used rough set models to handle fuzzy and quantitative data. These methods intend to induce every certain rule directly from the original data sets.

Like the second group, our approach uses a rule generation algorithm to induce all certain rules and possible rules from the original incomplete data. In this paper, we propose a mathematic method based on the Binomial Distribution to calculate the plausibility and probability of possible rules from original incomplete information systems.

2. BACKGROUND AND DEFINITIONS

2.1. Rough sets theory preliminary (R.S.T)

The rough sets theory provides a natural method to deal with incomplete or inconsistent information which has been the mainly obstacle to the classification and rule induction of objects. In this section we recall some basic notation of R.S.T that related to our research, and we assume that reader is familiar with basic principle of R.S.T. For more detail introduction of R.S.T, see references [1, 31].

.Incomplete information systems

An incomplete information system *IS* with two-tuple can be seen as a system: $IS = (U, \overline{A})$, where $\overline{A}$ is the set of attributes (features, variables) containing unknown values. Each attribute $a \in \overline{A}$ defines an information function $\overline{f_a} : U \rightarrow \overline{V_a}$, where $\overline{V_a}$ is the set of values of a including unknown values (*M*), called the domain of attribute a. A decision table is any information system with a decision attribute d: $\overline{T} =(U, \overline{A} \cup d)$.

Example.1 Transfer the incomplete medical diagnosis data (table1) into rough sets decision table.

Using the terminology of the rough sets theory this data set can be considered as follows:

$U = \{x_0 , x_1 , x_2 , x_3 , x_4 , x_5 , x_6 ,. ....x_9\}$

$\overline{A} = \{F_1, F_2, F_3\}$

d =(Temperature, Dry-cough , Headache)

TABLE 1
MEDICAL DIAGNOSIS DATA

Row no	Attributes			Decision
	Temperature	Dry-cough	Headache	Influenza
0	Missing	Absent	Absent	Absent
1	Normal	Absent	Present	Absent
2	Subfebrile	Absent	Present	Present
3	Subfebrile	Missing	Absent	Absent
4	Subfebrile	Present	Absent	Present
5	High	Absent	Missing	Absent
6	High	Present	Absent	Absent
7	High	Present	Absent	Present
8	High	Present	Present	Present
9	High	Present	Present	Present

TABLE 2
DECISION TABLE

Obj	F1	F2	F3	F4
X_0	M	0	0	0
X_1	0	0	1	0
X_2	1	0	1	1
X_3	1	M	0	0
X_4	1	1	0	0
X_5	2	0	M	0
X_6	2	1	0	0
X_7	2	1	0	1
X_8	2	1	1	1
X_9	2	1	1	1

The domains of the particular attributes are:

$\overline{V_1}$ *={0,1,2,M}, 0=normal, 1=* Subfebrile, *2=high, M=missing.*

$\overline{V_2}$ *={0,1,M}, 0=absent, 1=present, M=missing.*

$\overline{V_3}$ *={0,1,M}, 0=absent, 1=present, M=missing.*

$\overline{V_d}$ *={0,1}, 0=absent, 1=present.*

i.e., the domain of each attribute is the set of values of this attribute. The decision table for this system is presented in Table 2.

2.2. Reduct generation

In R.S.T based applications for classification and rule induction, the knowledge induction from reducts is the most important concept. By definition, a reduct is defined as minimal sufficient sets of features necessary for the description of all features A, (Pawlak, 1991)[32]. Nevertheless, a rule reduct, r-reduct, is a subset of features that can define all basic concepts for each object. In other words, a r-reduct is utilizing part of input features to uniquely identify output feature for each object. Furthermore, each r-reduct represents a decision rule. Consider a single-feature r-reduct with four input features and one output feature:2 x x x 1 , the entry 'x' mean that corresponding feature does not affect the determination of the feature output, only entry'2' does, and the decision rule can be expressed as:

$$If\ F_1=2\ Then\ d=1$$

For two-feature r-reduct: 1 x 2 x 2 , only features F_1 and F_3 affect the determination of the output feature , and the decision rule is:

$$If\ F_1=1 \land F_3=2\ Then\ d=2$$

Usually, there exists more than one r-reduct for each object. However, the knowledge rule of data not only can be inducted from lower and upper approximation of R.S.T but also can be directly analysed from all r-reducts.

3. PROBABILITY ANALYSIS FOR POSSIBLE RULES

If we determine all the certain rules of table1 using data mining tools, the result will come out as following:

one feature r-reduct:

F1	F2	F3	d	Obj
0	Nan	Nan	0	2

two feature r-reduct:

F1	F2	F3	d	Obj
Nan	0	0	0	1
1	1	Nan	0	3
1	Nan	0	0	3
2	0	Nan	0	5
1	Nan	1	1	7

(NaN denote "x")

And finally the certain rules are :

Certain rules

Symbols for Influenza = absent
IF (Temperature = normal)
IF (Dry-cough = absent) and (Headache = absent)
IF (Temperature = subfebrile) and (Dry-cough = present)
IF (Temperature = high) and (Dry-cough = absent)
IF (Temperature = subfebrile) and (Headache = absent)

Symbols for Influenza = present
IF (Temperature = subfebrile) and (Dry-cough = present)

However, in real world applications, certain rules induced directly from incomplete information systems may not provide enough knowledge for enterprises or decision makers to predict uncertain situations or provide strategies. Hence, generating possible rules by probability analysis can improve the expediency of rule extraction approach for incomplete information systems.

Let $U/R = \{E_1\ E_2\ E_3 \ldots\ldots E_m\}$ be a partition space of universe of objects U, where E_i is set of objects in the same equivalence class defined by R. For example, for a one-feature F_i, R may be defined as the set of all possible outcomes of F_i and E_j is the set of objects having the same j^{th} outcome of F_j. Let $D_i=\{x \in U \mid d(x)= d_i\}$ be the set of all objects whose output feature are classified as d_i. A "possible" rule: "If input feature $F_{ij} = a_{ij}$ then output feature $d = d_i$" can be associated with a plausibility as :

$$\frac{\#\text{of objects k with } a_{kj} = a_{ij} \text{ and } d(k) = d_i}{\#\text{of objects k with } a_{kj} = a_{ij}}$$

or expressed as $: \dfrac{|E_j \cap D_i|}{|E_j|}$ (1)

As we know, "plausibility" indices only calculates the frequency of the rule, which are not "probability analysis" that allow more precise statements of reliability and quality of those rules. Hence, in this section, we will build a probability function combining the plausibility and probability of missing values to compute the possible rules for incomplete information systems.

3.1. The Binomial Distribution

The binomial distribution is a discrete distribution model to calculate the probability of an experiment that has two possible outcomes for each trial, called "success" and "failure". Let the probability of success and failure be B and 1-B respectively, if the experiment consists of t repeated trials, the binomial model computes the probability of θ successes in t trials as a random variable X as follows:

$$P(X=\theta) = \binom{t}{\theta} B^{\theta} (1-B)^{t-\theta} \quad for\ \theta = 0, 1, \ldots, t \qquad (2)$$

Where

x = a random variable representing the number of successes

$\binom{t}{\theta}$ = the number of ways θ objects can be chosen from a set of t objects.

i.e. $\binom{t}{\theta} = \dfrac{t!}{\theta!(t-\theta)!}$

The binomial distribution model is appropriate when satisfying these assumptions:

1) Binary outcome for each trial.
2) The probability B is identical for each trial.
3) Fixed experiment size, t.

4) The outcomes of t trials are all independent.

Next we will show how the binomial model can be applied to build the probability function of possible rules after inducing certain rules from incomplete information systems.

When considering the plausibility of an uncertain rule-reduct *p*, which include features containing missing value M_{ij}, the outcome of M_{ij} will affect the probability of that possible rule. For each missing M_{ij}, there exist two possible outcomes, $M_{ij} = a_{ij}$ (success) or $M_{ij} \neq a_{ij}$ (failure). Hence, we can consider the outcomes of M_{ij} as a binomial distribution:

$$P(X=\theta)=\binom{t_p}{\theta} B^{\theta} (1-B)^{tp-\theta} \quad for\ \theta= 0, 1...,tp \quad (3)$$

where

θ = the number of success outcomes($M_{ij}=a_{ij}$) for possible reduct p

t_p = the total number of M_{ij}

B = the probability that $M_{ij}=a_{ij}$.

Recall that for a given set of features $\Im$, let $\Im_k$ be the k^{th} possible combination of values of the features in $\Im$, and there are m possible combinations. In case $\Im$ consists of a single feature F_j, $\Im_k$ is the k^{th} possible combination of F_j.

Then we define $E_k = \{ x \in U | \Im(x) = \Im_k \}$, k=1, …, *m*

Also define $D_i = \{x \in U | d(x) = d_i\}$ where d_i is the i^{th} possible value of decision feature *d*. Assume there are *n* possible values of *d*, i.e. *i*=1, …, *n*.

Finally we define A_{ik} as the plausibility that $\Im_k$ the *kth* possible combination of outcomes of features in $\Im$, signifies the decision variable d_i. Thus:

$$A_{ik} = \frac{| E_k \cap D_i |}{| E_k |} \qquad k = 1, \ldots, m;\ i = 1, \ldots, n$$

where $| S |$ = number of members in the set *S*, cardinality of *S*.

Note that

$$\sum_{i=1}^{n} A_{ik} = \sum \frac{| E_k \cap D_i |}{| E_k |} = \frac{| E_j |}{| E_j |} = 1$$

Now suppose each possible combination of outcomes $\Im_k$ has some missing values, nevertheless, we still would like to construct a reduct p_{ij} using the available outcomes in $\Im_k$ to infer decision d_i. There is obviously some uncertainties associated with the accuracy of the prediction of reduct P_{ij} due to the fact the missing values may or may not result in conflicting outcomes. Reducts generated in this case are therefore not “certain” reducts, but “possible” reducts with some probability of being right. We combine the concepts of plausibility A_{ij} above with the binomial model of missing values described earlier to construct the probability associated with reducts p_{ij}’s as follows:

Before we give a general formula for computing the probability of a general reduct, we consider a special simple case of reduct built on one-feature reduct first to illustrate the ideas involved.

One-feature reduct case:

Let $\Im$ consists of a single feature *F* who j^{th} possible outcome is f_j.

Let p_{ij} be the one-feature reduct “ For an object $x \in U$, *if* $F(x) = f_j$, *then* $d(x) = d_i$”.

Let $E_j = \{ x \in U | F_j(x) = f_j\}$ and recall $D_i = \{x \in U | d(x) = d_i\}$

Now for any $x \in U$, if $F_j(x)$ does not exist, we say x has a missing value under F_j and we denote the missing value by M_{xj}. Let the number of M_{xj}(under feature F_j) be t_j.

If the missing value $M_{xj} = f_j$ and $x \in D_i$, then the object *x* supports or strengthens reduct p_{ij}. We call this case a “success”. We can assume that the probability that $M_{xj} = f_j$ is constant and equals *B*. Let θ be the number of missing values M_{xj} that are equal to f_j, which is a random variable with the binomial distribution as discussed earlier.

Now out of t_j missing values that can possibly affect the accuracy of reduct p_{ij}, there are $L_\theta = \binom{t_j}{\theta}$ combinations that θ of those missing values will be equal to f_j.

Let $\theta_\ell \subseteq U$ be the l^{th} combination of such combinations.

Define $E_{\theta\ell} = \{ x \in U | F(x) = f_j \} \cup \{ x \in \theta_\ell | M_{xj} = f_j \}$

Clearly the first and second sets in the union on the left hand side are exclusive. Hence:

$$|E_{\theta\ell} \cap D_i|$$
$$= |\{x \in U \mid F(x) = f_j\} \cap D_i|$$
$$+ |\{x \in \theta_\ell \mid M_{xj} = f_j\} \cap D_i|$$
$$= \eta_{ij} + \theta_{i\ell} \qquad (4)$$

We note that given the number of missing values assigned to f_{jk} being equal to θ, the probability that it will have combination θ_ℓ is $1/L_\theta$ since each of the L_θ combination is equally likely. Thus if we define $A_{i\theta}$ as the probability that if θ of the t_j missing values $M_{xj} = f_j$, then the reeduct p_{ij} will predict d_i correctly:

$$A_{i\theta} = \sum_{l=1}^{L_\theta} \left(\frac{|E_{\theta\ell} \cap D_i|}{|E_{\theta\ell}|} \right) \times p(\theta_\ell \mid \theta)$$

$$= \sum_{\ell=1}^{L_\theta} \left(\frac{\eta_{ij} + \theta_{i\ell}}{\eta_j + \theta} \right) \frac{1}{L_\theta} = \frac{\eta_{ij}}{\eta_j + \theta} + \frac{\sum_{\ell=1}^{L_\theta} \theta_{i\ell}}{(\eta_j + \theta) L_\theta} \text{-------}(5)$$

where

$$\eta_j = |\{x \in U \mid F(x) = f_j\}|.$$

Note that $\sum_i \eta_{ij} = \eta_j$ and $\theta = \sum_i \theta_{i\ell}$ which is independent of ℓ. Now we also note that:

$$\sum_i A_{i\theta} = \sum_i \left(\frac{\eta_{ij}}{\eta_j + \theta}\right) + \sum_i \left(\frac{\sum_{\ell=1}^{L_\theta} \theta_{i\ell}}{(\eta_j + \theta) L_\theta}\right)$$

$$= \frac{1}{\eta_j + \theta} \left[\left(\sum_i \eta_{ij} \right) + \frac{\sum_{\ell=1}^{L_\theta} \left(\sum_i \theta_{i\ell} \right)}{L_\theta} \right] = \frac{\eta_j + \theta}{\eta_j + \theta} = 1$$

Moreover, we can now estimate the probability that reduct p_{ij} will predict correctly (regardless of how the missing values M_{xj} turn out) as:

$$p(d(x) = d_i \mid F_j(x) = f_j)$$
$$= \sum_{\theta=0}^{t_j} A_{i\theta}\, p(\theta)$$
$$= \sum_{\theta=0}^{t_j} A_{i\theta} \binom{t_j}{\theta} B_j^{\theta} (1 - B_j)^{t_j - \theta}$$
$$= \sum_{\theta=0}^{t_j} \binom{t_j}{\theta} A_{i\theta} B_j^{\theta} (1 - B_j)^{t_j - \theta} \text{------------}(6)$$

Note that (6) is a legitimate probability quantity since

$$\sum_i p(d(x) = d_i \mid F_j(x) = f_j)$$
$$= \sum_{\theta=0}^{t_j} \sum_{\theta=0}^{t_j} A_{i\theta}\, p(\theta) = \sum_{\theta=0}^{t_j} \left(\sum_i A_{i\theta}\right) p(\theta)$$
$$= \sum_{\theta=0}^{t_j} p(\theta) = 1 \text{ as desired}$$

Multiple-feature reduct case

Now we consider a more general case where the set of features that is used to form a reduct p consists of more than one feature.

Let there be J features in $\Im$. Recall that we define $\Im_k$ as the k^{th} combination of possible values of the J features in $\Im$. A reduct p_{ik} may, hence be defined as: "If the features in $\Im$ have a combination of values $\Im_k$ for object x, then $d(x)=d_i$". If there are missing values M_{xj} for feature j in $\Im$ for object x in U, then reduct p_{ik} is not a "certain" rule, since the actual value of M_{xj} can support or conflict with reduct p_{ik}. The probability that reduct p_{ik} will give an accurate prediction can be computed as follows:

Let

B_j : be the probability that missing value $M_{xj} = f_{jk}$ where f_{jk} is the value of feature in $\Im$ that is part of the combination $\Im_k$ in reduct p_{ik}. Normally we estimate B_j as $1 / V_j$, where V_j is the possible values of feature j.

t_j : be the total number of missing values of feature j in $\Im$ that will affect reduct p_{ik} among all object $x \in U$.

Hence there will be $t = \sum_{j \in \Im} t_j$ missing values for all features in $\Im$ that will affect reduct p_{ik}. Since each missing value can either support (i.e $M_{xj} = f_{jk}$ and $d(x) = d_i$), have no effect (i,e $M_{xj} \neq f_{jk}$ and $d(x) = d_i$) or contradict (i,e $M_{xj} = f_{jk}$ and $d(x) \neq d_i$) reduct p_{ik}. there are $T=2^t$ cases to be considered. Let q be the index of these cases, where $q = 1, \ldots, T$.

Now for each case q, let

M_{xjq} = the assignment of the missing value M_{xj} under case q.

θ_{jq} = the number of objects in U who missing values under feature j, M_{xjq}, is assigned f_{jk}.

ζ_{iq} = the number of objects in D_i whose missing values are assigned values in support of reduct p_{jk} under q, i,e

$$\zeta_{iq} = \left| \bigcap_{j\in\Im} \{x \in D_i \mid M_{xjq} = f_{jk}\} \right|$$

ζ_q = the number of objects in U whose missing values are assigned under case q the exact same pattern as in reduct p_{jk}, i,e

$$\zeta_q = \left| \bigcap_{j\in\Im} \{x \in U \mid M_{xjq} = f_{jk}\} \right|$$

Again it is easy to see that $\zeta_q = \sum_i \zeta_q$

Also let $E_k = \{x\in U \mid F_j(x) = f_{jk}, j\in \Im\}$, which is the set of objects that have exactly the right combination of values of features in $\Im$ used to form reduct p_{jk} (but not necessarily have their decision values equal to d_i).

$\eta_k = \mid E_k \mid$, which is the number of objects in E_k.

$E_{ik} = E_k \cap D_i$, which is the number of objects that completely support reduct p_{jk}.

$\eta_{ik} = \mid E_{ik} \mid$, which is the number of objects in E_{ik}.

Finally, let E_{kq} be the set of objects whose feature values $F_j(x)$ and /or missing values M_{xjq} are assigned f_{jk}, $j\in \Im$ under case q. That is:

$$E_{kq} = E_k \cup \{x \in U \mid M_{xjq} = f_{jk},\ j\in \Im\} = E_k \cup (\bigcap_{j\in\Im}\{x\in U \mid M_{xjq} = f_{jk}\})$$

Due to the disjoint of the two sets in the union on the right hand side, gives this:

$$|E_{kq}| = |E_k| + |(\bigcap_{j\in\Im}\{x\in U \mid M_{xjq}= f_{jk}\})| = \eta_k + \zeta_q$$

Thus the probability that reduct p_{ik} will predict correctly if the associated missing values have their values distributed according to case q is:

$$p\Big([d(x) = d_i \mid F_j(x) = f_j, j\in\Im] \mid case\ \ q\Big)$$

$$= A_{iq} = \frac{|E_{kq} \cap D_i|}{|E_{kq}|}$$

$$= \frac{|E_k \cap D_i| + |\{x\in D_i \mid M_{xjq} = f_{jk}, j\in\Im\}|}{|E_{kq}|}$$

(Due to the disjoint of the two sets in the union in the numerator)

Hence $A_{iq} = \dfrac{\eta_{ik} + \zeta_{iq}}{\eta_k + \zeta_q}$ ----------------------(7)

Note that

$$\sum_i A_{iq} = \frac{\sum_i \eta_{ik} + \sum_i \zeta_{iq}}{\eta_k + \zeta_q} = \frac{\eta_k + \zeta_q}{\eta_k + \zeta_q} = 1$$

as required by the property of probability functions.

Now we can compute the probability P_{ik} that reduct p_{ik} will predict correctly under any scenario of the missing values as:

$$P_{ik} = p\big(d(x) = d_i \mid F_j(x) = f_j, j\in\Im\big) = \sum_q p\Big([d(x) = d_i \mid F_j(x) = f_j, j\in\Im] \mid case\ \ q\Big) \quad (8)$$

$P(case\ q) = \sum_q A_{iq} P(case\ q)$

where $P(case\ q) = P(\theta_{jq}, j\in\Im)$

$$= \prod_{j\in\Im} (B_j)^{\theta_{jq}} (1-B_j)^{t_j-\theta_{jq}} \quad \text{---(9)}$$

$B_j = prob(M_{xjq} = f_{jk}) = \dfrac{1}{V_j}$, V_j = number of possible values of feature j Thus

$$P_{ik} = \sum_{q=1}^{T} A_{iq} \left(\prod_{j\in\Im} (B_j)^{\theta_{jq}} (1-B_j)^{t_j-\theta_{jq}} \right) \text{-------(10)}$$

Equation (10) is a legitimate probability distribution since; for reduct p_{ik}

$$\sum_i P_{ik} = \sum_i \left(\sum_{q=1}^{T} \left(\prod_{j\in\Im} (B_j)^{\theta_{jq}} (1-B_j)^{t_j-\theta_{jq}} \right) A_{iq} \right) =$$

$$\sum_{q=1}^{T} \left(\prod_{j\in\Im} (B_j)^{\theta_{jq}} (1-B_j)^{t_j-\theta_{jq}} \right) \left(\sum_i A_{iq} \right) =$$

$$\sum_{q=1}^{T} \left(\prod_{j\in\Im} (B_j)^{\theta_{jq}} (1-B_j)^{t_j-\theta_{jq}} \right) \text{----------------(11)}$$

Since $\sum_i A_{iq} = 1$

Without lost of generality, let $\Im = \{1, 2, \ldots, J\}$. Now for each j say $j = 1$, if we hold the values of M_{xj} for all $j \neq 1$ and $x \in U$ at a particular pattern (out of $2^{\sum_{j\neq 1} t_j}$ patterns), there will be 2^{t_1} possible ways that we can assign M_{x1}, $x \in U$. The summation of $B_j^{\theta_{1q}} (1-B_j)^{t_1-\theta_{1q}}$ over q over all those 2^{t_1} patterns is equal to

$$\sum_{\substack{q\in 2^{t_1}\ patterns\\ of M_{x1},x\in U}} \left(B_j^{\theta_{1q}}(1-B_j)^{t_1-\theta_{1q}}\right) =$$

$$\sum_{\theta_{1q}=0}^{t_1}\binom{t_1}{\theta_{1q}} B_j^{\theta_{1q}}(1-B_j)^{t_1-\theta_{1q}} = 1$$

Thus

$$\sum_i P_{ik} = \sum_{\substack{q\in 2^{\sum_{j\neq 1} t_j}\ patterns of\\ M_{x1},j\in\Im,j\neq 1,x\in U}} \prod \left(B_j^{\theta_{jq}}(1-B_j)^{t_{j1}-\theta_{jq}}\right)$$

Applying the same argument to $j = 2, 3, \ldots$ until $j=J$, we thus have:

$$\sum_i P_{ik} = \sum_{\theta_{Jq}=0}^{t_J}\binom{t_J}{\theta_{Jq}} B_j^{\theta_{Jq}}(1-B_j)^{t_J-\theta_{Jq}} = 1$$

Hence we have shown the require property for P_{ik} be a proper probability function.

The following examples illustrate the use of the above formula.

Example 2

F1	*F2*	*F3*	*d*
M_{11}	0	0	0
0	0	1	0
1	1	0	0
1	M_{42}	0	0
2	0	0	0
2	1	M_{63}	0
1	0	1	1
2	1	0	1
2	1	1	1
2	1	1	1

If we like to compute the probability of a two-feature reduct *p* " *1 0 NaN 1*" (*If F1=1 &F2=0 then d=1)* from the decision table above. The element sets for feature 1 and 2 before considering the possibility of missing values M_{11} and M_{42} are:

$E_1=\{2\}$ *for* $F_1=0$, $F_2=0$
$E_2=\{3\}$ *for* $F_1=1$, $F_2=1$
$E_3=\{5\}$ *for* $F_1=2$, $F_2=0$
$E_4=\{6, 8, 9, 10\}$ *for* $F_1=2$, $F_2=1$
$E_5=\{7\}$ *for* $F_1=1$, $F_2=0$
$D_1=\{1, 2, 3, 4, 5, 6\}$ *for* $d_1=0$.
$D_2=\{7, 8, 9, 10\}$ *for* $d_2=1$.

We know that reduct *p* is induced from object $7\in E_5$ and $d(7) = 1\in d_2$. As shown in decision table, there are two missing values, M_{11} and M_{42}, will affect the probability of reduct *p*. Hence $t=t_1+t_2=2$ and $T=2^2=4$ cases. We say $M_{11}=1$ and $M_{42}=0$ are successes and $B_1=1/3$, $B_2=1/2$.

First we compute the probability A_{iq} for each case *q*.

Let $i=2$, $k=5$.

Case 1: $M_{11} \neq 1$, $M_{42} \neq 0$

Then there is no object whose missing values are assigned f_{jk}. We get:

$\theta_{11}=0$, $\theta_{21}=0$, $\zeta_{21}=0$, $\zeta_1=0$ and

$\eta_5 = |E5| = |\{7\}| = 1$

$\eta_{25} = |E_{51}\cap D_2| = |E_5\cap D_2| = |\{7\}| = 1$

From (4)

$$A_{21} = \frac{\eta_{25}+\zeta_{21}}{\eta_5+\zeta_1} = \frac{1+0}{1+0} = 1$$

Case 2: $M_{11} \neq 1$, $M_{42} = 0$

Then there is one object (x_4) whose missing values are assigned f_{jk}. We get:

$\theta_{12}=0$, $\theta_{22}=1$, $\zeta_{22}=0$, $\zeta_2=1$ and

$$A_{22} = \frac{\eta_{25}+\zeta_{22}}{\eta_5+\zeta_2} = \frac{1+0}{1+1} = 1/2$$

Case 3: $M_{11}=1$, $M_{42} \neq 0$

Then there is one object (x_1) whose missing values are assigned f_{jk}. We get:

$\theta_{13}=1$, $\theta_{23}=0$, $\zeta_{23}=0$, $\zeta_3=1$ and

$$A_{23} = \frac{1+0}{1+1} = 1/2$$

Case 4: $M_{11}=1$, $M_{42}=0$

Then there are two objects (x_1 and x_4) whose missing values are assigned f_{jk}.

We get:

$\theta_{14}=1$, $\theta_{24}=1$, $\zeta_{24}=0$, $\zeta_4=2$ and

$$A_{24} = \frac{\eta_{25}+\zeta_{22}}{\eta_5+\zeta_2} = \frac{1+0}{1+2} = 1/3$$

Therefore, from (5) we compute the probability P_{ik} of reduct *p*:

$$P_{25} = \sum_{q=1}^{4} A_{iq}\left(\prod_{j=1}^{2}(B_j)^{\theta_{jq}}(1-B_j)^{t_j-\theta_{jq}}\right)$$

$$= 1\times\left(\frac{1}{3}\right)^0\left(1-\frac{1}{3}\right)^1\left(\frac{1}{2}\right)^0\left(1-\frac{1}{2}\right)^1$$

$$+ \frac{1}{2}\times\left(\frac{1}{3}\right)^0\left(1-\frac{1}{3}\right)^1\left(\frac{1}{2}\right)^1\left(1-\frac{1}{2}\right)^0$$

$$+ \frac{1}{2}\times\left(\frac{1}{3}\right)^1\left(1-\frac{1}{3}\right)^0\left(\frac{1}{2}\right)^0\left(1-\frac{1}{2}\right)^1$$

$$+ \frac{1}{3}\times\left(\frac{1}{3}\right)^1\left(1-\frac{1}{3}\right)^0\left(\frac{1}{2}\right)^1\left(1-\frac{1}{2}\right)^0$$

$=23/36=63.89\%$

4. CONCLUSION

In this paper, we propose a mathematic method to calculate the plausibility and probability of possible rules Based on the binomial distribution model, the probabilistic function (P_{ik}) depends on the probability of missing value B_j for each feature j in $\Im$. Here, we assume that each missing entry will have an equally likely chance to assume each of the V_j possible values of the corresponding feature *j*.

An alternative approach to estimate B_j is based on the degree of belief of the assessor, who may in turn have relevant knowledge and experiences to provide inputs for such estimate. If this is the case, then Dempster-Shafer's theory of evidence is an appropriate mathematical tool to help in the estimation.

REFERENCES

[1] Z. Pawlak,. Rough Sets, International Journal of Computer and Information Sciences, 11(1982).341-356.

[2] J. Grzymala-Busse., Managing Uncertainty in Expert Systems, Kluwe Boston, (1991).

[3] L.P. Khoo, S.B.Tor, .L.Y. Zhai, Rough Set-based Approach to Classification and Rule Induction, International Journal of Advanced Manufacturing Technology, Sprigег-Verlag, 15(1999)438-444.

[4] J. W. Grzymala-Busse., On the unknown attribute values in learning from examples, Methodologies for Intelligence Systems, Sixth International Symposium (ISMIS' 91),

[5] Lecture Notes in Artificial Intelligence, Charlotte, NC, 542(1991)16-19.

[6] W. Ziarko., Analysis of uncertain information in the framework of variable precision rough sets, Fund. Computing Decision Science, 18(1993)381-396.

[7] S. Greco, B. Matarazzo, R. Slowinski, Rough approximation by dominance relation, International Journal of Intelligent Systems 17(2002) 153–171.

[8] Z. Pawlak, A. Skowron, Rudiments of rough sets, Information Sciences 177 (2007) 3–27.

[9] Z. Pawlak, A. Skowron, Rough sets: some extensions, Information Sciences 177 (2007) 28–40.

[10] S.K. Pal, Soft data mining computational theory of perceptions and rough-fuzzy approach, Information Sciences 163 (1–3) (2004) 5–12.

[11] S. Tsumoto, Mining diagnostic rules from clinical databases using rough sets and medical diagnostic model, Information Science 162(2) (2004) 65–80.

[12] Y.-Y. Guan, H.-K. Wang, Set-valued information systems, Information Sciences 176 (2006) 2507–2525

[13] T.P. Hong, L.-H. Tseng, S.-L. Wang, Learning rules from incomplete training examples by rough sets, Expert Systems with Applications 22 (2002) 285–293.

[14] Y. Leung, D.Y. Li, Maximal consistent block technique for rule acquisition in incomplete information systems, Information Sciences 153 (2003) 85–106.

[15] Y. Leung, W.-Z. Wu, W.-X. Zhang, Knowledge acquisition in incomplete information systems: a rough set approach, European Journal of Operational Research 168 (2006) 164–180.

[16] M.-W. Shao, W.-X. Zhang, Dominance relation and rules in an incomplete ordered information system, International Journal of Intelligent Systems 20 (2005) 13–27.

[17] J. Stefanowski, A. Tsoukias, Incomplete information tables and rough classification, Computational Intelligence 17 (2001) 545–566.

[18] W.-X. Zhang, J.-S. Mi, Incomplete information system and its optimal selections, Computers and Mathematics with Applications 48(2004) 691–698.

[19] .R..Slowinski,. J.Stefanowski., Rough classification and incomplete information system, Mathematical and Computer Modeling, 12(1989)1347-1357.

[20] I. Kononenko, I. Bratko, E. Roskar, Experiments in automatic learning of medical diagnostic rules, Technical Report, Jozef Stefan Institute, Ljubljana, Yugoslavia, 1984.

[21] J.R. Quinlan, Induction of decision trees, in: J.W. Shavlik, T.G. Dietterich (Eds.), Readings in Machine Learning,

Morgan Kaufmann, Los Altos, CA, (1990), 57–69.
[22] .M.R.Chmielewski. J.W.Grzymala-Busse, N.W.Peterson, S.Than,. The rule induction system LERS-A version for personal computers, Found.Comput.Decision Science 18(1993)181-212.
[23] .P.J. Lingras, Y.Y. Yao, Data Mining using extensions of the Rough Set model, Journal of the American society for information science. 49(5)(1998)415-422.
[24] .P.J. Lingras, Plausibilistic rule extraction from incomplete databases using non-transitive rough set model, Proceedings of the twenty-third Computer Science Conference (CSC'95) orkshop on Rough sets and Database Mining(1995)..
[25] M. Kryszkiewicz., Rough set approach to incomplete information systems, Information Sciences ,112(1998)39-49.
[26] T.Q. Deng, Y.M. Chen, W.L. Xu, Q.H. Dai, A novel approach to fuzzy rough sets based on a fuzzy covering, Information Sciences 177 (2007) 2308–2326
[27] T.P. Hong, T.T. Wang, S.L. Wang, Mining fuzzy b-certain and b-possible rules from quantitative data based on the variable precision rough-set model, Expert Systems with Applications 32 (2007) 223–232.
[28] T.P. Hong, T.T. Wang, S.L. Wang, Knowledge acquisition from quantitative data using the rough-set theory, Intelligent DataAnalysis 4 (2000) 289–304.
[29] R. Jensen, Q.A. Shen, Fuzzy-rough sets assisted attribute selection, IEEE Transactions on Fuzzy Systems 15 (2007) 73–89.
[30] X. Wang, E.C.C. Tsang, S. Zhao, D. Chen, D.S. Yeung, Learning fuzzy rules from fuzzy samples based on rough set technique,
Information Sciences 177 (2007) 4493–4514.
[31] M.Kryszkiewicz,.Rules incomplete information systems," Information Sciences,113(1999)39-49
[32] B.Walczak. D.L. Massart.. Rough sets theory, Chemometrics and intelligent laboratory systems, 47(1999)1-16.
[33] Pawlak, Z., 1991. Rough Sets: Theoretical Aspects of Reasoning About Data, 9, Kluwer Boston.