Open Access

An Incomplete Data Filling Approach Based on a New Valued Tolerance Relation

Liu Ping[1],*, Qiu Taorong[1], xiong Xiaoxiao[1] and Yang Junping[2]

[1]*The school of Information Engineering, Nanchang University, Nanchang 330031, China*

[2]*Jiangxi University of Traditional Chinese Medicine, China*

Abstract: In real life there are many incomplete information system, However, the traditional rough set theory is not suitable for incomplete information system. A lot of extension of the rough sets theory have been proposed based on this. In these theories, the handling of null value or missing values is the key problem. In this paper a new valued tolerance and a concept of Tolerance Degree Vector are put forward at first; moreover a new incomplete data filling approach which based on new valued tolerance and tolerance degree vector(for short "NVT-TDV") is proposed. Subsequently, two series of experiments have been carried out, one was compared the classification accuracy with other extended rough set model, the other was the data filling accuracy test based on Waikato Environment for Knowledge Analysis platform . The experimental results show that it can be adopted as a pre-processing method in data mining.

1. INTRODUCTION

In our daily life there are various reasons lead to incomplete information system(IIS), such as the errors of data measuring, the limitations of acquiring data, some human factors, etc. For example, on the hospital information, the attributes are some clinical examinations and not all the examination results can be got, it leads to lots of null value of medical data. However the traditional rough set theory developed by Professor Pawlak which aimed at the complete information system, is not suitable to handle incomplete information system. Therefore, in recent years, many scholars and experts focus on how to manage incomplete information. At present, According to documents [1-16] acknowledged there are two methods to deal with incomplete information systems.

One is the direct method, model extension have put forward to deal with the incomplete information systems in rough set theory. Now a lot of generalized rough set models have been proposed. For example, Kryszkiewicz's tolerance relation [1], Stefanowski's non-symmetric similarity relation [2], and Guoyin Wang's limited tolerance relation [3], valued tolerance relation [4], On this basis, many other improved methods have proposed [5-15]. Compared with the other extension models, the valued tolerance relation has the following advantages. Firstly, it is based on the tolerance degree for judging the similarity between objects; Secondly, according to different incomplete information systems (IIS),

a pre-given threshold can be adjusted to determine the similarity between two objects. The disadvantage of valued tolerance relation is that a given threshold is difficult to determine. In recent years there are some research papers around the valued toleration relation, in article [7], an improved method to calculate the tolerance degree of two objects more objective is given; in article [8], the author puts forward a new calculation method of tolerance degree and the auto selection method of threshold which do not require any prior domain knowledge ;the objective calculation method of tolerance degree is not only founded on the basis of the statistical characteristics of attribute values in IIS, but also considers the effect of the number of known and same attribute values between objects.

The other is the indirect approach, it is called pretreatment of data, incomplete information system can be completed through data filling or deleting missing value. For example, using Bayesian model to fill the null value needs to know the probability density; using evidence theory requires evidence functions. In Paper [16], missing data are filled using indiscernibility relation of rough set (RS) , through the experiments demonstrate its effectiveness, but the complexity of computing discernibility matrix for vast data set is high.

In this paper, aiming at the defects of valued tolerance relation, a new valued tolerance relation is proposed at first, It can select the threshold value according to the characteristics of data set. In determining the tolerance relation between objects, consider the completeness of the object and whether the decision value are same. Then it is applied to the filling process and decision table rule derivation of incomplete data. In section 2,a new valued tolerance relation is defined;

In section 3, an improved algorithm of data filling for IIS is presented; In section 4, some of the algorithm accuracy testing has been done and the results of test are discussed. Conclusions are given in the last section.

2. A NEW VALUED TOLERATION RELATION FOR INCOMPLETE INFORMATION SYSTEM

In this section, some related concepts of valued tolerance relation will be first introduced, and then a new valued toleration relation based on rough set will be put forward.

Let $S = <U, C \cup D, V, f>$ be an information system(IS),Where $U = \{u_1, u_2, ..., u_{|U|}\}$ is a set of objects, it is a non-empty finite set of objects, called the universe, $A = C \cup D$ is a set of attributes. C as the condition attribute set, D as the decision attribute set. If some attribute values are unknown, the IS is IIS(incomplete information system), otherwise it is complete information system(CIS). In IIS, the unknown attribute value is denoted by " * " . In article [8], the probability $P_{\{b\}}(x, y)$ that x is similar to y on b [8].

$$P_{\{b\}}(x,y) = \begin{cases} 1 & b(x) = b(y) \land b(x) \neq * \land b(y) \neq * \\ 0 & b(x) \neq b(y) \land b(x) \neq * \land b(y) \neq * \\ k_b^i / \sum_{j=1}^{m} k_b^j & (b(x) = b_i \land b(y) = *) \lor (b(x) = * \land b(y) = b_i) \\ \sum_{i=1}^{m} (k_b^i / \sum_{j=1}^{m} k_b^j)^2 & b(x) = * \land b(y) = * \end{cases} \quad (1)$$

In the formula, $IIS = <U, C \cup D, V, f>$, $B \subseteq C$, $\exists b \in B$, $V_b' = \{k_b^1 b_1, k_b^2 b_2, \cdots k_b^m b_m\}$, where $b_1, b_2, \cdots b_m$ is all possible known values of b, and k_b^i denotes the cardinality of the set $\{x \in U \mid b(x) = b_i\}$, For any $x \in U$, the probability that $b(x) = b_i$ is $k_b^i / (k_b^1 + k_b^2 + \cdots + k_b^m)$. x is similar to y on B is shown as the probability $P_B(x, y)$[8];

$$P_B(x, y) = \prod_{b \in B} p_{\{b\}}(x, y) \quad (2)$$

The selection of threshold is a key problem for valued tolerance relation, For different data sets, using a compromise approach to select the threshold, the average toleration degree was chosen as the threshold of each data set. Next, we give the definition .

Definition 1 : The average value of toleration degree is defined as $\lambda = \frac{\sum P_B(x, y)}{C_{|U|}^2}$, x, y are two arbitrary objects in U , $x, y \in U$; $|U|$ is cardinality of the set U . $C_{|U|}^2$

represents permutation and combination , it shows that select any two from $|U|$ objects.

Definition 2: suppose $S = <U, C \cup D, V, f>$ is a incomplete information system(IIS), In the subset $B \subseteq C$,The completeness of the object $x_i \in U$ is defined as: $\gamma_B(x_i) = |T(x_i)| / |B|$, In the formula: $T(x_i) = \{b \mid b \in B \land b(x_i) \neq *\}$, $|B|$ represents the cardinality of B set, Then the completeness of IIS in the B is defined as follows

$$\gamma_B = \sum_{i=1}^{|u|} \gamma_B(x_i) / |U|$$

For the purpose of acquiring potential and usable rules from the decision table, it should keep the initial decision table's features, and potential laws as possible. So for the incomplete decision table, objects with the same values on condition attributes are same as values on decision attributes as possible. That is to say we should keep the consistence of the decision table. So in the following definition, we consider whether the decision value are same.

Definition 3 : New Valued Tolerance relation

$IIS = <U, C \cup D, V, f>$, $B \subseteq C$, $\exists b \in B$, New Valued Tolerance relation is defined as

$$NVT_B^\lambda = \{(x, y) \in U \times U \mid (P_B(x, y) > \lambda) \\ \land (\frac{\gamma(x) + \gamma(y)}{2} \geq \gamma_B) \land (f(x, D) = f(y, D))\} \bigcup I_U \quad (3)$$

$$I_U = \{<x, x> \mid x \in U\}$$

Definition 4 : New valued tolerance class :

$$NVT_B^\lambda(x) = \{y \mid (x, y) \in NVT_B^\lambda, y \in U\} \quad (4)$$

Definition 5: $IIS = <U, C \cup D, V, f>$, if $x_i \in U$, then the missing attribute set $MAS(x_i)$ of object x_i , missing object set MOS of IIS are defined as:

$$MAS(x_i) = \{k \mid a_k(x_i) = *, k = 1, 2, ..., m\}$$

$$MOS = \{t \mid MAS_t \neq \varphi, t = 1, 2, ..., n\}$$

Definition 6: Let S be an incomplete information system in which $\forall B \subseteq C$, then $\forall X \in U$,the lower approximation and upper lower approximation of X in terms of NVT_B^λ are denoted by $\underline{NVT_B^\lambda(X)}$ and $\overline{NVT_B^\lambda(X)}$, respectively, where,

$$\underline{NVT_B^\lambda} = \{x \in U : NVT_B^\lambda(x) \subseteq X\} \quad ;$$

$$\overline{NVT_B^\lambda(x)} = \{x \in U : NVT_B^\lambda(x) \cap X \neq \varphi\}$$

Definition 7: Toleration degree vector

$$IIS = < U, C \cup D, V, f >, \quad B \subseteq C, \quad \forall x, y \in U,$$

$$b_i \in B(i = 1, 2, \dots n), \ n = |B|,$$

the toleration degree vector of objects x, y on B is defined:

$$V_B(x, y) = (P_{b1}(x, y), P_{b2}(x, y), \cdots P_{bn}(x, y)) \tag{5}$$

Definition 8: norm of Toleration degree vector

$IIS = < U, C \cup D, V, f >, \quad B \subseteq C, \quad \forall x, y \in U$, norm of toleration degree vector is defined as:

$$\tau_B(x, y) = \sqrt{P_{b1}^{\,2}(x, y) + P_{b2}^{\,2}(x, y) + \dots + P_{bn}^{\,2}(x, y)} \tag{6}$$

3. A NEW INCOMPLETE DATA ANALYSIS APPROACH BASED ON ROUGH SET

Based on the above description, we next discuss a data filling method according to incomplete decision table.

Algorithm 1 : a completed method based on New Valued Tolerance relation (NVT)

Input: Incomplete decision table $S^0 = < U, C \cup D, V, f^0 >$

Output: Complete information table

Step 1: for $S^0 = < U, C \cup D, V^0, f^0 >$, compute the MOS^0;

Step 2: Aiming at the object which have missing values, calculate the tolerance class.

for $x \in MOS^0$ DO

compute $NVT_B^{\lambda}(x) = \{y \mid (x, y) \in NVT_B^{\lambda}, y \in U\}$.

$$NVT_B^{\lambda} = \{(x, y) \in U \times U \mid (P_B(x, y) > \lambda) \wedge \left(\frac{\gamma(x) + \gamma(y)}{2} \geq \gamma_B \right) \wedge (f(x, D) = f(y, D)) \} \cup I_U$$

Step 3: Filling missing values in incomplete information system.

for $x_i \in MOS^0$ DO

compute $MAS(x_i)$, based on $NVT_B^{\lambda}(x_i)$, $b_k \in B(1 \leq k \leq |B|)$

Initialization: r=0;

While ($NVT_B^{\lambda}(x_i^r) \neq \varphi$)

{(1) In the object sets which are compatible with x_i, Look for the pair who have the Maximum tolerance degree or have the Maximum completeness degree.

find $\gamma(y_{max}) = \max\{\gamma(y_j)\} \vee P_B(x_i, y_{max}) = \max\{P_B(x_i, y_j)\}$

$1 \leq \max \leq |NVT_B^{\lambda}(x_i)|; \ 1 \leq j \leq |NVT_B^{\lambda}(x_i)|; \ y_j \in NVT_B^{\lambda}(x_i)$

(2) Filling the missing data.

$$b_k(x_i^{r+1}) = \begin{cases} b_k(y_{max}), & b_k(x_i^r) = * \\ b_k(x_i^r), & b_k(x_i^r) \neq * \end{cases}, k = 1, 2, \dots |B|;$$

(3) if ($MAS(x_i^{r+1}) \neq \varphi$)

{ $NVT_B^{\lambda}(x_i^{r+1}) = NVT_B^{\lambda}(x_i^{r+1}) - \{y_{max}\}$;

r++;

}

Else

{ $MOS = MOS - \{x_i\}$;

break;}// Exit the loop

}//end while

Through the above analysis, if there are still missing values in the information system, combination completeness approach(Algorithm 2) is adopted for further process.

Algorithm 2: Completeness approach based on norm of Toleration degree vector (TDV)

Input: $x_i \in MOS$; $MOS \neq \varphi$; $SIM(x_i) = \varphi$; $B \subseteq C$; $1 \leq i \leq |MOS|$;

Output: x_i'

for $x_i \in MOS$ do

{ step1: find the object which decision value is equal that of x_i ;

for $y_j \in U$ do $// 1 \leq j \leq |U|$

If $d(x_i) = d(y_j)$

$SIM(x_i) = SIM(x_i) + \{y_j\}$;

Step2: choose the object z whose norm of Toleration degree vector is the maximum

For $z_t \in SIM(x_i)$ $(1 \leq t \leq |SIM(x_i)|)$

// Based formula 5,6

{Compute $V_B(x_i, z_t) = (P_{b1}(x_i, z_t), P_{b2}(x_i, z_t), \cdots P_{bn}(x_i, z_t))$;

Table 1. Five complete data sets in UCI.s

Data Sets	No. of Objects	No.of Condition Attributes	No. of Decision Attributes
Balance	625	4	1
Car	1728	6	1
Chess	3196	36	1
Promoters	106	58	1
Tic-Tac-Toe	958	9	1

Compute

$$\tau(x_i,z_t) = \sqrt{P_{b1}^2(x_i,z_t) + P_{b2}^2(x_i,z_t) + ... + P_{bn}^2(x_i,z_t)};$$
$$b1,b2,\cdots bn \in B$$

}//end for

Step3: The filling process

$$TR = \{\tau(x_i,z_1),\tau(x_i,z_2),...\tau(x_i,z_t)\}$$

r=0;

While $(TR \neq \varphi)$

{Find $\tau_{max} = \tau(x_i,z_{max})$ // $\tau_{max} \in TR$

$$b_k(x_i^{r+1}) = \begin{cases} b_k(z_{max}), & b_k(x_i^r) = * \\ b_k(x_i^r), & b_k(x_i^r) \neq * \end{cases}, k=1,2,...|B|$$

If $((MAX(x_i) \neq \varphi))\{$

$TR = TR - \{\tau_{max}\};$

r++;

Else break;

}//end while

4. EXPERIMENTS AND RESULT ANALYSIS

4.1. Experiment 1: The Test of Classification Accuracy about New Valued Tolerance Relation

For New Valued Tolerance relation, we compared its classification accuracy with that of tolerance relation, limited tolerance relation, statistical valued tolerance relation.

Statistical valued tolerance relation is defined by the formulas (1),(2),(7)

$$VT_B^\lambda = \{<x,y> \in U^2 \mid P_B(x,y) \geq \lambda\} \cup I_U \qquad (7)$$

limited tolerance relation is defined by formula (8).

$$L(x,y) = \{(x,y) \in U \times U \mid$$
$$\forall a \in C(a(x) = a(y) = *) \vee$$
$$[p(x) \cap p(y) \neq \varphi \wedge \forall a \in C((a(x) \neq *) \qquad (8)$$
$$\wedge (a(y) \neq *) \rightarrow a(x) = a(y)))]$$
$$p(x) = \{a \in C \mid a(x) \neq *\}$$

In our experiments, five complete data sets (Balance, Tic-Tac-Toe, Chess, Car, Promoters,)in UCI have been used. For each data set, randomly selected from 5%, 10%, 30% data for missing values, so 15 incomplete data sets (Balance-5%, Balance-10%, Balance-30%, Tic-Tac-Toe-5%, Tic-Tac-Toe-10%, Tic-Tac-Toe-30%, Chess-5%, Chess-10%, Chess-30%, Car-5%, Car-10%, Car-30%, Promoters-5%, Promoters -10%, Promoters -30%) have been got.

We use the definition of classification accuracy in paper [8].

$$\mu_R = \begin{cases} \dfrac{\sum\limits_{x \in U} \dfrac{|[x]_E \cap R(x)|}{|[x]_E \cup R(x)|}}{|U|} & (9) \\[4ex] \dfrac{\sum\limits_{x \in U} \dfrac{|[x]_E \cap (K_R(x) \cup K_R^{-1}(x))|}{|[x]_E \cup (K_R(x) \cup K_R^{-1}(x))|} k}{|U|} & (10) \end{cases}$$

If R is symmetric, use formula (9) to compute μ_R, else use formula (10) to compute μ_R.

The experiment results are shown in Table **2**, where TR, LTR, SVT, NDVT denote respectively tolerance relation, limited tolerance relation, statistical valued tolerance relation and new valued tolerance relation.

From Table **2**, we can see that the classification accuracy of NVT is obviously higher than that of TR and LTR. When the missing data rate is lower than 30%, the classification of NVT is higher than that of SVT; while the missing data rate is higher than 30%, the classification of SVT is slightly better than NVT.

Table 2. Classification accuracies of four generalized indiscernibility relations.

Data Sets	TR	LTR	SVT(0.05)	SVT(0.1)	NVT
Balance-5%	0.650	0.650	0.639	0.639	1.000
Balance-10%	0.395	0.395	0.505	0.505	1.000
Balance-30%	0.056	0.065	0.667	0.671	0.322
Car-5%	0.581	0.581	0.533	0.647	1.000
Car-10%	0.308	0.308	0.316	0.553	0.709
Car-30%	0.032	0.033	0.476	0.803	0.250
Chess-5%	0.815	0.815	0.837	0.866	0.943
Chess-10%	0.629	0.629	0.778	0.857	0.825
Chess-30%	0.114	0.114	0.985	0.997	0.408
Promoters-5%	1.0	1.0	1.000	1.000	1.000
Promoters-10%	1.0	1.0	1.000	1.000	1.000
Promoters-30%	1.0	1.0	1.000	1.000	1.000
Tic-Tac-Toe-5%	0.930	0.930	0.950	0.976	1.000
Tic-Tac-Toe-10%	0.741	0.741	0.887	0.957	1.000
Tic-Tac-Toe-30%	0.105	0.105	0.968	0.988	0.464

Table 3. Classification accuracies of new filling algorithm(NVT-TDV) and SVT.

		Raw data			NVT-TDV			SVT--0.1			SVT--0.2		
		Accuracy	Error Rate	Unclassi-fied	Accuracy	Arror Rate	Unclassi-fied	Accuracy	Error Rate	Unclas-sified	Accura-cy	Error Rate	Unclas-sified
Balance	5%	0.3808	0.0704	0.5488	0.5621	0.1053	0.3326	0.4859	0.1771	0.3372	0.4835	0.1749	0.3416
	10%				0.6590	0.1312	0.2098	0.5337	0.2507	0.2156	0.5469	0.2378	0.2153
	30%				0.8498	0.09248	0.05776	0.7427	0.2213	0.0360	0.7750	0.1861	0.0389
Car	5%	0.8935	0.3530	0.07118	0.8969	0.05272	0.05035	0.7845	0.1318	0.08374	0.7924	0.1279	0.07974
	10%				0.9056	0.05856	0.03582	0.7731	0.1599	0.06694	0.7933	0.1530	0.05365
	30%				0.9425	0.04456	0.01290	0.8710	0.1144	0.01458	0.9175	0.0713	0.01169
Chess	5%	0.9969	0.0031	0	0.9912	0.0088	0.0	0.9570	0.0429	0.00007	0.9678	0.0322	0
	10%				0.9730	0.02697	0.00003	0.9449	0.0504	0.0047	0.9599	0.0399	0.0002
	30%				0.9414	0.05845	0.00125	0.9249	0.0795	0.0003	0.9258	0.0740	0.0028

Table **3**. contd…

		Raw Data			NVT-TDV			SVT--0.1			SVT--0.2		
		Accuracy	Error Rate	Unclassi-fied	Accuracy	Arror Rate	Unclassi-fied	Accuracy	Error Rate	Unclas-sified	Accura-cy	Error Rate	Unclas-sified
Promoters	5%	0.7641	0.1981	0.0378	0.7113	0.2377	0.05188	0.7113	0.2368	0.05189	0.7113	0.2368	0.05189
	10%				0.7226	0.1960	0.08132	0.7226	0.1960	0.0813	0.7264	0.1960	0.0813
	30%				0.8057	0.1632	0.0311	0.8057	0.1632	0.03113	0.8056	0.1632	0.0311
Tic-Tac-Toe	5%	0.8319	0.1388	0.0293	0.8219	0.1420	0.03685	0.8132	0.1487	0.0381	0.8212	0.1420	0.0368
	10%				0.8461	0.1227	0.03110	0.8237	0.1291	0.04715	0.8461	0.1227	0.03110
	30%				0.9031	0.07714	0.01973	0.8949	0.0850	0.0200	0.8987	0.0820	0.01920

For promoters data set, we can see that the result is not consistent with other data set. In Table **2**, all the result about the classification of promoters is equal to 1.The reason lies in the characteristics of this data set is multi-attribute, there are 58 condition attributes and multi attribute values for each species, so for missing data objects, the tolerance classes which can be found only its own.

4.2. Experiment 2: The Experiment About Accuracy of New Filling Algorithm (NVT-TDV)

The comparison of filling accuracy has been implemented with the help of Waikato Environment for Knowledge Analysis platform. Testing method was: First an incomplete data set was transformed into completed using the algorithm of this article, then selected 70% from it as training set to extract decision rules, using these rules to do classifying test on the rest data set, the average classification accuracy has been acquired after test 10 times for every incomplete data set. The test result is shown in Table **3**.

In Table **3**, Raw data is on behalf of original UCI data set. The obtaining method of incomplete information systems is same as Experiment 1.

From the theoretical analysis, the NVT-TDV method takes account of the consistent decision values in filling data, in search of compatible class requires the decision value consensus. In Table **3**, we can see that the average classification accuracy of NVT-TDV is significantly higher than that of SVT. The experimental results are consistent with theoretical deduction. See from experimental result, the average classification accuracy of NVT-TDV is close to the original UCI data set, it is fully shows the correctness of our filling method.

CONCLUSION

In this paper, a new valued tolerance relation is developed based on the idea of data-driven data mining. This definition of new valued toleration relation has the following advantages: one is to adopt a method of the paper [8] in the calculation of the tolerance degree; two is in the search for valued tolerance class fully considering the completeness of objects and the consistency of decision value.

From the theoretical analysis and experimental data show that the new method is advanced, provides a good solution for pretreatment process of incomplete information system.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Kryszkiewicz M., "Rough set approach to incomplete information system," Information Sciences, vol.112, issues 1-4,pp.39-49, 1998.

[2]   Stefanowski J. and Tsouki A., "Incomplete information tables and rough classification," Computational Intelligence, 17, pp.545-566, 2001.

[3]   G. Y.Wang, "Extension of rough set under incomplete information systems ," Journal of Computer Research and Development, 39(10), 2002, 1238-1243

[4]   Stefanowski J. and Tsouki A.., "On the extension of rough sets under incomplete information," Lecture Notes in ArtificialIntelligence 1711,pp.73-81, 1999.

[5]   Zaimei Zhang, Renfa Li, Zhongsheng Li, "An Incomplete Data Analysis Approach Based on the Rough Set Theory and Divide-and-Conquer Idea," Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007).

[6]   Mingfen Wu,Ting-Liang Wang,"AN ALGORITHM: FILLING THE MISSING VALUES IN THE INCOMPLETE DECISION TABLE ," Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, Kunming, 12-15 July 2008

[7]   JIXia,LILongshu, CHEN Shengbing ,XU Yi," A New Extended Rough Set Model Based on The Dynamic Tolerance Relation and Improved Tolerance Degree Calculating Method," 2009 Asia-Pacific Conference on Information Processing, pp.94-96

[8]   Guoyin Wang, Lihe Guan, "Data-Driven Valued Tolerance Relation," RSKT 2012,LNAI 7414,PP.11-19,2012; Springer-Verlag Berlin Heidelberg 2012

[9]   Nguyen Do Van, Koichi Yamada,Muneyuki Unehara, "Rough Set Model Based on Parameterized Probabilistic Similarity Relation in Incomplete Decision Tables," SCIS-ISIS 2012, Kobe, Japan, November 20-24, 2012.

[10]   Junyi Zhou, Xibei Yang," Study on characteristic consistent blocks based rough set approach to incomplete system," 2012 IEEE International Conference on Granular Computing

[11]   Yu-qin GAO, Guo-hua FANG, Ya-qin LIU," θ-improved limited tolerance relation model of incomplete information system for evaluation of water conservancy project management modernization," Water Science and Engineering, 2013,6(4):469-477 doi:10.3882/j.issn.1674-2370.2013.04.010

[12]   Yongqiang Bai, Wenzhong Zha , Jie Chen, Zhihong Peng," A Refined Classification Method with Tolerance Relation-Based Rough Sets for Incomplete Decision Systems," 2013 IEEE International Conference on Systems, Man, and Cybernetics,pp.69-73

[13]   Do Van Nguyen, Koichi Yamada, and Muneyuki Unehara," On Probability of Matching in Probability Based Rough Set Definitions," 2013 IEEE International Conference on Systems, Man, and Cybernetics, DOI 10.1109/SMC.2013.82,pp.449-454

[14]   Jianhua Dai," Rough set approach to incomplete numerical data," Information Sciences 241 (2013) 43‑57

[15]   Shucheng Huang,, Mingxing Li," Limited and Variable Precision Rough Set Model," Journal of Information & Computational Science 11:10 (2014) 3493‑3501

[16]   Weihua Zhou, Wei Zhang, Yunqing Fu. "An incomplete data analysis approach using rough set theory". Intelligent Mechatronics and Automation.2004, pp.332-338.