A Multi-Threshold Granulation Model for Incomplete Decision Tables

Renpu Li
School of Information and Electrical Engineering, LuDong University, Yantai, China
Email: lrp0109@163.com

Tao Yu
School of Information and Electrical Engineering, LuDong University, Yantai, China
Email: tianxibing2008@163.com

Chunjie Zhou
School of Information and Electrical Engineering, LuDong University, Yantai, China
Email: lucyzcj@163.com

Hongbo Li
School of Information and Electrical Engineering, LuDong University, Yantai, China
Email: fast_run_man@126.com

Abstract—**How to establish basic granules of knowledge is a fundamental issue for data mining from incomplete decision tables. In the existing methods, basic granules under similarity relation contain too many objects and disturb the later knowledge mining, while granules under limited similarity relation, although simplifying the granules through introducing a limited threshold on two objects satisfying similarity relation, still have problems such as high computation and low prediction precision. In this paper, a multi-threshold model is presented to establish basic knowledge units of incomplete decision table based on the idea of granular computing, comparison experiments on the new model with two existing models show that the new model is superior to the other models on prediction precision, time cost and attribute reduction.**

Index Terms—**Incomplete decision tables, similarity relation, granular computing, multi-threshold**

1. INTRODUCTION

Granular computing is an efficient method to deal with inexact, uncertain or vague information [1][2]. There are two key tasks for applying a granular computing algorithm, the first one is selection and construction of basic granules and the second is granulation of specific problem [3].

Incomplete data exists widely in the real life, and knowledge mining from incomplete data has attracted a great deal of attentions [4] [5]. For complete data, by equivalence relation, all objects can be partitioned into some equivalence classes, which can be used as natural basic granules for knowledge reasoning and concept approximation. Different with the complete data, construction of basic granules in incomplete data is more complex. Because the equivalence relation needs to be extended in order to deal with the relations between objects in incomplete data.

Several extensions of equivalence relation, mainly belong to tolerance relation, have been proposed [6] [7]. However, these granules induced by extension relations have some defects, such as lack of clear semantic [8].

In addition, as a main characteristic of granular computing, the idea of hierarchy is used by many researchers to mining knowledge from incomplete data [9]. A hierarchical reduction method proposed in [10] converts a problem solving of single hierarchy and single granulation to a problem solving of multi-hierarchy and multi-granulation.

Hierarchical structures on multi-granulation spaces are investigated in [11]. Three different hierarchical structures are proposed on two types of the multi-granulation spaces, partition-based and covering-based, respectively. Properties about these hierarchical structures and the relationships between these hierarchical structures and the multi-granulation rough sets are deeply discussed.

Also based on the idea of hierarchy, [12] investigates the extended formulas and the formulation representation of granules and introduces some operations of granules in rough sets, and a conceptual framework of knowledge retrieval is proposed based on the granular structure model develop which enlarges the application areas of granular computing.

Manuscript received July 10, 2013; revised October 2, 2013; accepted October 25, 2013.

Corresponding author: Renpu Li; email: lrp0109@163.com;

This research was supported by the Project of Norway Government Scholarship 2013/2014 under Grant No. 229566/F1, etc.

doi:10.4304/jsw.9.7.1922-1929

[13] summarizes recent application developments of rough sets and granular computing in hierarchical learning, the general framework of rough set based hierarchical learning is presented, and some techniques for embedding the domain knowledge into the granular, layered learning process are also proposed in order to improve the quality of hierarchical classifiers.

The hierarchical idea of granular computing is also applied to the field of control engineering [14]. Using different fuzzy controls for coarse adjustment in the coarse granularity and adopting classical PID as fine control in the fine granularity, a supervisor which is a fuzzy controller is designed to switch between the different fuzzy controllers and PID control smoothly.

In this paper, a new granular computing model is presented for the knowledge mining from incomplete decision tables. The original incomplete decision table is firstly a set of complete objects and a set of incomplete objects. Secondly incomplete objects whose incompleteness degrees are limited by two thresholds are integrated according to the certain information provided by the set of complete objects. And then basic granules are created through a matching degree based on a limited tolerance relation.

Comparison experiments show that the new model is superior to two existing models on prediction precision, time cost and attribute reduction.

This paper is organized as follows. Basic concepts related to granulation model are introduced in Section 2, In Section 3 a new multi-threshold granulation model is proposed and a demonstration example is given. Comparison experiments with two existing models are presented in Section4. Finally concluding remarks are given in Section 5.

II. Basic Concepts

2.1. Incomplete Decision Table

Definition 1 An information system can be expressed $S=(U,A,V,f)$ where U is a non-empty finite set of objects and A is a non-empty finite set of attributes, such that f: $U \to V_a$ for any $a \in A$, where V_a is called the value set of a.

If some attribute values of objects in an information system are missing, these values are called missing values, which will be denoted by symbol "*" in this paper, other values are called regular values. If an information system contains at least one missing value, it is called an incomplete information system, otherwise it is complete.

Decision table is a special information system $T=(U,C \cup \{d\},V,f)$ where d , $d \notin C$ *and* $* \notin V_d$, is a distinguished attribute called decision, and the elements of C are called condition attributes.

An example of incomplete decision table called S1 is presented in TABLE 1 where $C=\{C1,C2,C3,C4\}$, d is the decision attribute and there are 13 objects, some of them are incomplete objects. For example, x10 is an incomplete object having 3 missing values on condition attributes C2, C3 and C4 respectively.

TABLE 1.
INCOMPLETE DECISION TABLE *S1*

U	*C1*	*C2*	*C3*	*C4*	*D*
x1	3	2	1	0	A
x2	2	3	2	0	A
x3	2	3	2	0	B
x4	*	2	*	1	A
x5	*	2	*	1	B
x6	2	3	2	1	B
x7	3	*	*	3	A
x8	*	0	0	*	B
x9	3	2	1	3	B
x10	1	*	*	*	A
x11	*	2	*	*	B
x12	3	2	1	*	B
x13	3	2	1	0	A

2.2. Similarity Relation

Definition 2 Let $S=(U,A,V,f)$ be an incomplete information system. Each subset of attributes $P \subseteq A$ determines a binary similarity relation $SIM(P)$ on U :

$$SIM(P)=\{(x,y)\in U\times U \mid \forall a\in P, \quad f_a(x)=f_a(y) \;\; or \;\; f_a(x)=* \;\; or \;\; f_a(y)=*\}$$

Let $S_P(x)=\{y\in U \mid (x,y)\in SIM(P)\}$ called the similarity class of x . $S_P(x)$ can be used as a basic granule of an incomplete information and all similarity classes combine a granulation space which is a covering of the universe.

2.3. Incompleteness Rate and Matching Degree

As similarity relation is reflexive and symmetrical, the similarity classes induced by a similarity relation contain too redundancy [15], even some pairs of objects in a similarity classes have different values on some attributes. Aiming at this problem, adding some conditions to limit similarity relation is a natural choice [16] [17], where incompleteness rate and matching degree are two important measures [15].

Definition 3 Let $T=(U,C\cup\{d\},V,f)$ be an incomplete decision table and $x\in U$. Incompleteness rate of x is defined as follows:

$$IR(x)=\frac{|L_C(x)|}{|C|}$$

where |X| is the cardinality of set X.

In fact, the incompleteness rate of an object is the ratio of the number of condition attributes whose value is incomplete and the number of all condition attributes. The bigger the incompleteness rate of the object is, the more its missing values are. For an object with a big incompleteness rate, there is a big uncertainty even in describing the information of itself, so it is more infeasible to reasoning the relationships among objects based on it.

Definition 4 Let $T=(U, C\cup\{d\}, V, f)$ be an incomplete decision table and $U=\{x_1, x_2, ..., x_n\}$. Matrix of matching degree of U is defined as follows:

$$MD(i,j)=\begin{cases} 0 & \begin{array}{l}(i=j)\\ \vee(\forall a\in C[f_a(x_i)=*\vee f_a(x_j)=*])\\ \vee(\exists a\in C[f_a(x_i)\neq *\wedge f_a(x_j)\neq *\\ \qquad \wedge f_a(x_i)\neq f_a(x_j)])\end{array}\\ \dfrac{|W(i,j)|}{|C|} & others\end{cases}$$

where

$$W(i,j)=\{a\in C\mid f_a(x_i)\neq *\wedge f_a(x_j)\neq * \wedge f_a(x_i)=f_a(x_j)\}$$

Obviously, MD is a symmetric matrix, that is, $MD(i,j)=MD(j,i)$ for any $1<i,j<n$. So in the real applications, one can only compute the upper triangular elements of MD.

III. A MULTI-THRESHOLD GRANULATION MODEL

3.1. The Idea of Integration

Missing values in an incomplete decision table bring uncertainty for data analysis and mining. In order to simplify the complexity of computation, many researchers divide the all objects into two sets: one consists of complete objects and the other consists of incomplete objects. However, time cost to deal with incomplete objects base on tolerance relation is still high.

From the principle of hierarchy, in this paper we think the objects with same attribute values contain same classification rules and some incomplete objects can be deleted before rule extraction. In other words, objects with different incompleteness rates provide different knowledge for classification. We can integrate those incomplete objects with very high or very low incompleteness rate refer to the certain information provided by the set of complete objects.

For this objective, three thresholds, called α, β and γ respectively, are used. The former two thresholds are used to assign an interval of the missing value ratio of incomplete data objects, and those objects do not belong this interval will be integrated.

This idea is consistent with the thinking habit of people. For example, in a decision table with dozens of condition attributes, for an incomplete object, except few missing values, has the same values with another complete object, and also have the same decision value, then it will be integrated to a complete object.

Definition 5 Let $T=(U, C\cup\{d\}, V, f)$ be an incomplete decision table. A partial order $\leq$ on C is defined as follows:

$$\prec_C=\{(x,y)\in U\times U\mid \forall c\in C,\quad f_c(x)=*\quad or\quad f_c(x)=f_c(y)\}$$

From the definition, it can be seen that if $x\prec_C y$, the certain information provided by x is included in y, that is, for any regular values of x, y has the same regular value.

Let $0<\alpha<\beta<1$, α, β be called the lower and upper bound of incompleteness rate respectively.

For an incomplete object x whose incomplete rate is less than or equal to α, if we can find a complete object y which has the same decision d with x satisfying $x\prec_C y$, x will be think as an object having same attribute value with y and be deleted. That means, the missing values of x are regarded as the corresponding values of y because the certain information on x is so consistent with y.

On the other hand, for an incomplete object x whose incomplete rate is more than or equal to β, if we can find a complete object y which has the same decision d with x such that $x\prec_C y$, also x can be think as an object having same attribute values with y and be deleted because the uncertainty on values of x is so large that we only integrate it with the few classification information of y for reducing computational complexity.

Obviously, two integration operations have different starting-points. The former is based on enough certain information from incomplete object itself and it's consistency with complete object, while the latter is a helpless choice because the incomplete object contains little certain information.

In addition, above integration operations are only done between the objects with the same decision values. For those objects with higher incompleteness rate should not be directly deleted because some important classification rules maybe contained in them.

For example, object x3 in TABLE 2 should not be deleted although it's incompleteness rate is 6/7 because it contains an obvious rule

$$c6=3\Rightarrow D=C\ .$$

TABLE 2
AN EXAMPLE

U	C1	C2	C3	C4	C5	C6	C7	D
x1	2	3	2	1	1	2	5	A
x2	2	3	2	*	*	2	5	B
x3	*	*	*	*	*	3	*	C
x4	*	3	2	*	1	1	5	A

Finally, for those incomplete objects whose incompleteness rates are more than α and less than β, a limited similarity relation is applied to granulate them with other complete or incomplete objects. Here the limited condition is given by third threshold γ, only when the matching degree [15] of two objects $W(i, j)$ is more than γ, they can be thought as at the same granule.

3.2. The Algorithm of Multi-threshold Granulation Model

Based on the above integration idea, an algorithm MTGM (Multi-Threshold Granulation Model) for incomplete decision tables is proposed, which is outlined below.

Input: an incomplete decision table $T=(U, C \cup \{d\}, V, f)$, $C=\{c_1, c_2, ..., c_s\}$ and three thresholds α , β , γ .

Output: a granulation space of T, denoted as GRS and the new incomplete decision table T'.

(1) divide the objects of U into two sets, the first set includes complete objects denoted as $SET_1 == \{x_1, x_2, ..., x_m\}$, and the second set includes incomplete objects as $SET_2 = \{y_1, y_2, ...y_n\}$;

(2) Merge the same objects of SET_1 ;

(3) compute the incompleteness rates of all objects of SET_2 ;

(4) For i=1 to n

if ($IR(y_i) \le \alpha \quad or \quad IR(y_i) \ge \beta$)

and ($\exists x_i$ such that $y_i \prec x_j$)

delete y_i ;

(5) Let $U' = SET_1 \cup SET_2 = \{x_1, x_2, ..., x_p\}$;

(6) construct matrix of matching degree of U';

(7) get the granulation space GRS of T with the third threshold γ based on the method provided in [15] ;

(8) obtain the new incomplete decision table T' according to GRS.

It should be noted that in the step 6, the matrix obtained is a transformation of matrix, rather than a true matrix. More specific, for every incomplete object, only two objects that have the largest matching degree are saved in the matrix, which will improve greatly the efficiency of computation.

3.3. An Example

In this section, TABLE 1 is used as an example to illustrate the granulation process of the multi-threshold granulation model MTGM.

Assume α＝30%, β＝70%, γ＝20%, the granulation process of GBMD on TABLE 1 is as follows:

(1) divide the objects of U into two sets, obtain

$$SET_1 == \{x1, x2, x3, x6, x9, x13\},$$

and $SET_2 = \{x4, x5, x7, x8, x10, x11, x12\}$;

(2) Merge the same objects of SET_1 , obtain

$$SET_1 == \{x1, x2, x3, x6, x9\};$$

(3) compute the incompleteness rate of all objects of SET_2 , obtain

IR(x4)=1/2,
IR(x5)=1/2,
IR(x7)=1/2,
IR(x8)=1/2,
IR(x10)=3/4,
IR(x11)=3/4,
IR(x12)=1/4;

(4) because $IR(x11) \ge \beta$, $IR(x12) \le \alpha$, $x11 \prec x9$ and $x12 \prec x9$, $x11$ and $x12$ are deleted from SET_2 .

So we have

$$SET_2 = \{x4, x5, x7, x8, x10\}$$

(5) then we got

$$U' = SET_1 \cup SET_2 = \{x1, x2, x3, x6, x9, x4, x5, x7, x8, x10\}$$

(6) construct matrix of matching degree of U';

(7) get the granulation space GRS of T with the third threshold γ based on the method provided in [15] ;

(8) obtain the new incomplete decision table T' according to GRS.

It can be seen that the lower approximation of TABLE 1 obtained by the new algorithm is

{a1, a7, a10, a12, a6, a8, a9, a11}.

While the lower approximation obtained by the limited tolerance relation based on membership degree [8] ($\alpha_1 = 0.5$, $\alpha_2 = 0.5$) is

{a1, a6, a8, a12}.

Even if the two objects a10 and a11 deleted in the limited tolerance relation based on membership degree can be defined separately the lower approximation is obtained as

{a1, a6, a8, a10, a11, a12}.

Compared the results of two methods, the lower approximation obtained by new algorithm is a superset of that obtained by the limited tolerance relation based on membership degree. It shows that the algorithm proposed in this paper can achieve better approximation result than traditional method.

In addition, the granulation spaces obtained by traditional granulation methods are usually a covering of the universe, which is larger than original object space. Meanwhile the granulation space obtained by our algorithm is a partition of the universe, which is a smaller and simpler space than a covering.

IV COMPARISON EXPERIMENTS

In order to test the performance of the multi-threshold granulation (MTGM) model, a series of data mining experiments are designed for comparison with other two granulation models based on tolerance relation [6] (GMTR) and variable precision tolerance relation [17] (GMVPTR).

Firstly algorithms are edited based on three models respectively, and then the results of reduct are obtained through simulation experiments, finally the classification accuracies are outputted from a data mining software Orange [19].

Tolerance relation think the missing value has the possibility of equivalent to any value, so in this relation, two objects should be classified into different granules only when they have different values on at least one attribute, otherwise they should be classified into the same granule.

In general, this relation is very loose for the requirements of two objects in a granule and will cause the result that many objects with little equivalence possibility are classified into a granule, decrease the classification ability and produce few rules.

Variable precision tolerance relation introduces parameters to control the similarity degree of attribute values of two objects, compared to tolerance relation, this relation gives a more flexible and more accurate method to express the similarity degree of two objects, and could produce better granulation results. However, the computation efficiency is a serious problem for those algorithms based on this relation.

Experimental data include 6 data sets from UCI machine learning repository [18] whose information is listed in TABLE 3, through randomly deleting some attribute values all data sets are preprocessed to the different level incompleteness rates such as 10%, 20%, 30% or 40% respectively.

In these experiments, each incomplete data set is randomly separated into two parts: three fourth as training set and the rest one fourth as testing set.

Secondly the training set is transformed by the three granulation models respectively and the transformed data are mined by the same attribute reduction algorithm and same rule extraction algorithm.

Finally the test accuracies are obtained by applying the rules mined to the test set. The detail steps of experiments base on the software Orange [19] are shown in Fig. 1 where C4.5 algorithm is selected as classifier for rule extraction.

Firstly a UCI data set is saved as an excel file and preprocessed for the next step, for example the serial number of objects, which is irrelevant with data mining, should be deleted.

TABLE 3
INFORMATION OF EXPERIMENTAL DATA

Data sets	Number of objects	Number of attributes	Complete?	Rate of missing
Balance Scale (BS)	625	5	yes	0.0%
Car Evaluation (CE)	1728	7	yes	0.0%
Pittsburgh Bds (PB)	108	12	no	7.1%
Mammographic (Mp)	961	6	no	3.4%
Chess （Cs）	3196	37	yes	0.0%
Breast Cancer (BC)	699	10	no	0.3%

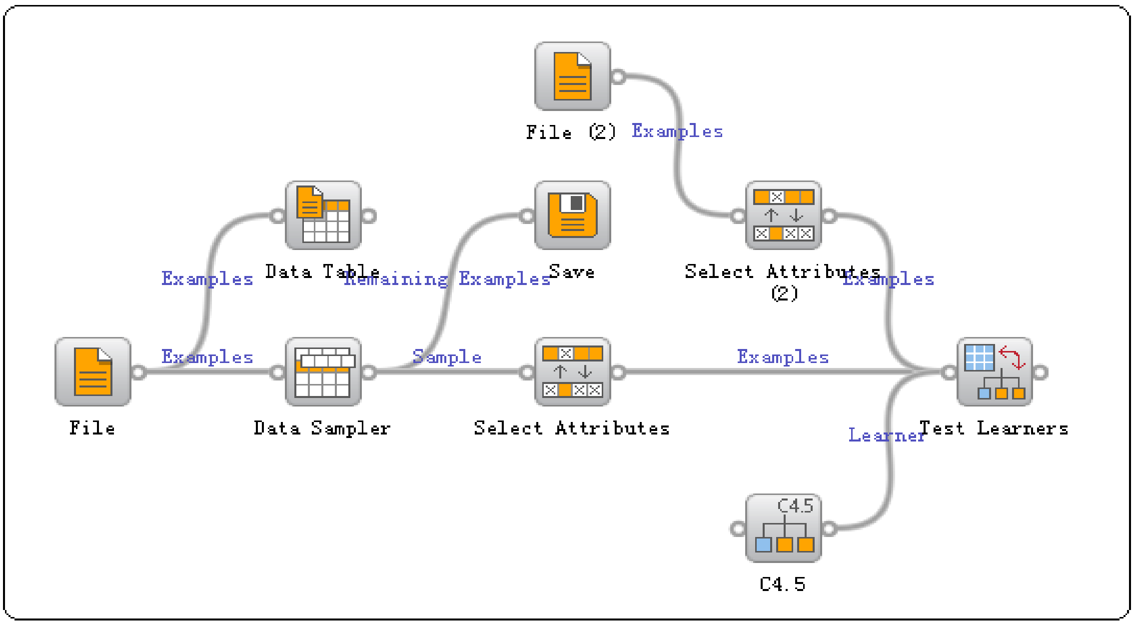

Fig. 1 Detail steps of data mining experiments

And then algorithm programming will be made based on the granulation models under the Java integrated development environment called NetBeans IDE6.7.1, mainly on the attribute reduction algorithm and preprocess of attribute values.

TABLE 4.
COMPARISON OF THREE GRANULATION MODELS ON ACCURACY (%)

	MTGM	GMTR	GMVPTR
BS	53.20%	52.52%	50.04%
CE	71.64%	69.27%	43.92%
PB	56.00%	40.67%	54.62%
Mp	61.10%	60.00%	59.57%
Cs	66.01%	69.75%	70.13%
BC	88.75%	94.86%	82.10%

Finally the reduced data set is divided into two parts, training set and testing set, as the input data of C4.5 algorithm.

Three measures including accuracy of rule (Accuracy), the number of attributes in a reduct (No. of attributes) and the running time of mining algorithm (Time) are used as the measures of evaluation of the granulation models. Results of three granulation models on six data sets with incompleteness rate 30% on three measures are listed in TABLE 4, TABLE 5 and TABLE 6 respectively.

The results in TABLE 4 show that MTGM is superior to GMTR and GMVPTR on accuracy in general. In the three granulation models, MTGM got the best accuracies on four data sets except Cs and BC. The reason can be seen from TABLE 5, in this table the number of attributes in the reduct of MTGM on Cs 16 is less than the numbers of attributes of GMTR and GMVPTR, 35 and 36 respectively, and the number of MTGM on BC is 5, also the minimal in three models.

It is obvious in TABLE 5 that the average number of attributes in the reduct obtained by MTGM is less than GMTR and GMVPTR. For GMTR, tolerance relation is loose for the classification of objects in data sets, so the granules produced by GMTR contain too many uncertain objects, and then only few attributes are reduced in the process of attribute reduction.

TABLE 5.
COMPARISON OF THREE GRANULATION MODELS ON NO. OF ATTRIBUTES (%) IN A REDUCT

	MTGM	GMTR	GMVPTR
BS	4	4	3
CE	6	6	4
PB	4	11	7
Mp	5	5	4
Cs	16	35	36
BC	5	9	5

Comparatively, for the GMVPTR, variable precision tolerance relation has the strict condition to ensure the similarity property of objects in a granule, which is effective to get a smaller reduct. For example, the number of attribute in a reduct obtained by GMVPTR is minimal on four data sets except PB and Cs.

TABLE 6 shows the running times of three granulation models on six data sets. The running time can be divided into two parts, one is the time of object classification in the process of granulation and the other is the time of attribute reduction. The final reduct obtained is just one of all possible reducts, not the minimal reduct. It can be seen that the running time of MTGM is greatly shorter than that of the other two models.

TABLE 6.
COMPARISON OF THREE GRANULATION MODELS ON RUNNING TIME (s)

	MTGM	GMTR	GMVPTR
BS	13.82	11.69	23.12
CE	25.51	71.46	289.10
PB	1.79	2.33	23.55
Mp	22.36	70.19	191.95
Cs	542.71	2659.3	11939
BC	17.60	118.15	211.09

In order to demonstrate the performance of MTGM on data with different level incompleteness rates, four data sets including CE, PB, Mp and BC are selected in a comparison experiment. Accuracies of MTGM on data sets with four incompleteness rates, 10%, 20%, 30% and 40% respectively are listed in TABLE 7.

TABLE 7.
ACCURACIES OF MTGM ON DATA SETS WITH DIFFERENT INCOMPLETENESS RATE

Incompleteness rate	CE	PB	Mp	BC
40%	69.30%	41.33%	74.17%	76.57%
30%	75.65%	56.92%	81.56%	79.80%
20%	73.85%	55.33%	77.95%	90.40%
10%	71.30%	36.67%	76.41%	91.02%

It can be seen that under the different level incompleteness rate, the accuracies of rules remain relatively stable, which indicates the robustness of MTGM on mining incomplete data.

TABLE 8.
OTHER RESULTS CORRESPONGING TO TABLE 7.

Three thresholds	Cardinality of reduct/sum of objects integraed			
	40%	30%	20%	10%
CE(0.15/0.85/0.5)	6/8	6/3	6/0	6/0
PB(0.10/0.90/0.4)	6/0	6/0	6/0	5/0
Mp(0.15/0.80/0.5)	5/1	5/8	5/39	5/145
BC(0.15/0.95/0.4)	6/0	5/126	6/76	4/182

It is obvious that the change tendencies of sum of objects integrated on four data sets are different. For example, with the increase of incomplete rates, the sum of objects integrated of CE increases, while the sum of objects integrated of Mp and BC decrease. The reason is related to the data distribution of different data sets.

Also for these comparison experiments, TABLE 8 provides the results of cardinality of reduct and sum of objects integrated of four data sets on different four incomplete rates.

From above experiments, it can be summarized that tolerance relation is very loose to define the similarity of two objects. So many objects with little similarity will be classified into same granule. In general, the algorithms based on this relation have lower ability on attribute reduction and produce fewer rules.

Variable precision tolerance relation uses different parameters to control the process of granulation and get a suitable way to express similarity. In many experiments, algorithms based on this relation could get better reduction effect and more accurate rules, however its computation complexity is high, which restricts its application.

Multi-threshold granulation model proposed in this paper introduce the idea of hierarchy into the process of granulation, uses different thresholds to control phases of data granulation and attribute reduction, not only get the better reducts and rules, but decrease the computation cost.

V. Conclusion

In this paper, a multi-threshold model is presented to granulation of an incomplete decision table. Two thresholds for the incompleteness rate of object are used to select the objects which are integrated to forming the basic knowledge units and the third threshold for the matching degree of two objects is applied to decide whether the two objects should be integrated or not.

Comparison experiments on the new model with two existing models show that the new model is superior to the other models and exhibit good robustness on mining data with different level incompleteness rates.

Acknowledgement

This work is supported by the National Natural Science Foundation of China under Grant No. 61202111, the Project of Norway Government Scholarship 2013/2014 under Grant No. 229566/F1, the Natural Science Foundation of Shandong Province China under Grant No. ZR2011GQ001 and ZR2012FQ029, the Project Sponsored by the Scientific Research Foundation for the Returned Overseas Chinese Scholars under Grant No. 43 and the Project of Shandong Province Higher Educational Science and Technology Program under Grant No. J12LN05.

References

[1] M. Luo, Granular Computing and Its Current Status of Research [J], Computer and Modernization, 2007, 11(1): 1-5.

[2] Daoguo Li, Zhaoxia Chen, Jie Yin, A New Attribute Reduction Recursive Algorithm Based On Granular Computing [J], Journal of Computers, 2013, 8(3): 630-637.

[3] G. Wang, Q. Hu, Uncertainty of rough sets in different knowledge granularities [J]. Chinese Journal of Computers, 2008, 31(9): 1588-1988.

[4] Shibao Sun, Ruijuan Zheng, Qingtao Wu, Tianrui Li, VPRS-Based Knowledge Discovery Approach in Incomplete Information System [J], Journal of Computers, 2010, 5(1): 110-116.

[5] Xu E, Yuqiang Yang, Yongchang Ren, A New Method of Attribute Reduction Based On Information Quantity in An Incomplete System [J], Journal of Software, 2012, 7(8): 1881-1888.

[6] L. Guan, Processing incomplete information methods based on rough set [J], Journal of Chongqing University of Posts and Telecommunication, 2009, (4): 461-466.

[7] G.Wang, Calculation methods for core attributes of decision table [J], Chinese Journal of Computers，2003, (5): 612-615.

[8] R. Li, Y. Yao, Indiscernibility and Similarity in an Incomplete Information Table [J], Lecture Notes in Computer Science, 2010, 6401:110-117.

[9] Xueyong Li, Jiaxia Sun, Guohong Gao, Junhui Fu, Research of Hierarchical Clustering Based on Dynamic Granular Computing [J], Journal of Computers, 2011, 6(12): 2526-2533.

[10] B. Qiao, H. Zheng, et al., The hierarchical reduction approach to incomplete information system [J], Journal of Circuits and Systems, 2001, 6(2): 78-82.

[11] XB Yang, YH Qian, JY Yang. Hierarchical structures on multi-granulation spaces. Journal of Computer and Technology, 2012, 27(6): 1169-1183.

[12] Lin Sun, Jiucheng Xu, Chuan Wang, Tianhe Xu and Jinyu Ren, Granular Computing-based Granular Structure Model and its Application in Knowledge Retrieval [J]. Information Technology Journal, 2012, 11: 1714-1721.

[13] Sinh Hoa Nguyen, Tuan Trung Nguyen, Marcin Szczuka, Hung Son Nguyen, An Approach to Pattern Recognition Based on Hierarchical. Granular Computing [J], Fundamenta Informaticae, 2013, 127: 369-384

[14] Pottem Sai Tharun, M. J. Nigam, Hierarchical fuzzy pid controller using granular computing for inverted pendulum, in Proceedings of SARC-IRAJ International Conference, June 2013, pp. 141-145

[15] Tao Yu, Renpu Li, Xiaohui Chen, Granulation based on matching degree for incomplete decision tables [J], Journal of Computational Information Systems, 2010, 6(10): 67-74

[16] G.Wang, Extension of rough set under incomplete information systems [J], Journal of Computer Research and Development, 2002, Vol.10, pp.1238-1243.

[17] Xia Ji, Shupeng Li. Extended rough set model based on variable precision dynamic tolerance relation [J]. Journal of System Simulation, 2009, 21(18): 5731-5734.

[18] UCI Machine Learning Repository, http: //archive.ics.uci.edu/ml/datasets.html.

[19] Orange 2.0b for Windows, http://orange.biolab.si/.

Renpu Li, professor, born in 1976, received the BS.c. in electrical engineering from Institute of Shandong Engineering in 1997, the MS.c. in computer application technology from Tianjin Polytechnic University in 2000, and the Ph.D. in management science and engineering from Tianjin University, China, in 2003. He was a postdoctor at East China University of Science and Technology from 2003 to 2005. Dr. Li is currently a professor in School of Information and Electrical Engineering, Ludong University, China. His current research interests include data mining, rough sets, soft computing etc. He is a member of the China Computer Federation and the ACM.

Tao Yu, born in 1981, received his BS.c. degree in computer science from Shandong University, China, in 2008. He is currently a postgraduate of Ludong University. His research interests include data mining and rough sets.

Chunjie Zhou, lecturer, born in 1981, received the BS.c. and MS.c. in computer science and technology from Hunan University of Science and Technology, China, and received Ph.D. in computer software and theory from Renmin University of China in 2010. Now she is a lecturer in School of Information and Electrical Engineering, Ludong University. Her current research interests include Ubiquitous computing, mobile data management, pattern mining etc. Dr. Zhou is the program chair of WMSC 2013, and the PC member of SCA 2012.

Hongbo Li, associate professor, born in 1970, received his MS.c. in electronical information engineering from Dalian University of Technology, China, in 2000. Now he is an associate professor in School of Information and Electrical Engineering, Ludong University. His current research interests include business intelligence and information system development.