A Continuous Information Attribute Reduction Algorithm Based on Hierarchical Granulation

Long Chen and Tengfei Zhang

College of Automation, Nanjing University of Posts and Telecommunications, Nanjing, China.
tfzhang@126.com

Abstract

The attribute reduction algorithms based on neighborhood approximation usually use the distance as the approximate metric. Algorithms could result in the loss of information with the same distance threshold to construct the neighborhood families of different dimension spaces. Thereby, an attribute reduction algorithm based on hierarchical granulation is proposed. This algorithm can reduce redundant attributes in the same granularity. Experimental results with UCI data sets show that the algorithm can improve the classification power, and reduce the loss of information.

Keywords: *Attribute Reduction; Hierarchical Granulation; Continuous Information; Granular Computing*

1. Introduction

Usually, there is some redundant information in large-scale data sets, many unnecessary attributes creates a barrier to the process of knowledge acquisition. Attribute reduction algorithm can effectively reduce the redundant attributes, which simplifies the process of knowledge acquisition. It is one of the hot issues of the knowledge acquisition.

The rough set theory proposed by Pawlak Professor in 1982, provides a powerful way for attribute reduction. It describes the indiscernibility of the data set by using known models to approach the target set. The classical form is lower approximate set and upper approximate set [1]. Many scholars have proposed a lot of reduction algorithms based on rough set theory. These reduction algorithms can be divided into the following categories:

1) Greedy reduction algorithm. Reduction algorithm selects or deletes the attributes one by one according to the change of the classification power of attribute set. The processing of these algorithms is simple, but the performance depends on the order of the searching.

2) Reduction algorithm based on heuristic information. These algorithms consider some heuristic information as an index for searching attributes, such as information entropy and so on. Relative to the greedy reduction algorithm, this method improves the efficiency, always looking for the solution set of optimal direction.

3) Algorithm based on discernibile matrix. Discernibile matrix marks differences between samples. These differences can count out the distinction performance between the properties. The discernibile matrix can effectively reduce the complexity of computation, and more directly reflect the process of attribute reduction.

These methods get a good reduction, but most only deal with discrete information. To continuous information attribute reduction, current approaches of the existing literature are as follows:

1) Depend on the expertise. For data sets in different fields, researchers can analysis datasets background with the expertise, so that the process of attribute reduction will be completed. Although this approach is effective, as the lack of universality, it is relatively difficult for the underlying application and spread of the algorithm.

2) Using the discretization. Many scholars have proposed discretization approach for continuous data. After discretization, the value ranges of continuous attributes are divided into a plurality of sections. Each interval is replaced by a uniform value. And then use the existing reduction algorithms for data reduction. Discretization usually results in the loss of useful information during the process of attribute reduction. And the reduction result largely depends on the effect of discretization.

3) Using neighborhood system. There is a correlation between the samples, such as data points are similar each other. These methods build neighborhood system with some similarity metric. It can directly mine relevance of data points, effectively reduce the loss of information and improve the precision of reduction set.

Currently, there are many kinds of reduction methods based on neighborhood relationship to deal with the continuous information. Hu [2] considers the distance metric as the neighborhood similarity. According to the approximation between neighborhood system and decision class, he proposes a greedy attribute reduction algorithm which defines the index of selecting attributes. And this method uses a variable precision to reduce the reduction error. Numerical attribute reduction algorithm based on neighborhood granulation and rough approximation avoids the loss of discretizing information and reduces the redundancy of attribute set. However, the reduction result always relies on the k-precision.

On the basis of the method, Hu [3] proposes a fast forward search method to solve the complexity issues. As attribute number increasing, it need recalculate the neighborhood lower approximate set to decision class, leading to the high calculate complexity. Fast forward search attribute reduction method decides the importance of reduction set by calculating the change of negative region of neighborhood system, after increasing the reduction attributes. This improvement reduces the amount of computation in attribute search to improve the efficiency of attribute reduction.

Feng [4] improves numerical attribute reduction algorithm based on neighborhood granulation and rough approximation with a new distance metric to build a neighborhood system. This distance metric consider the distribution of each attribute value. Using sample which can't be classified as the factor of approximate relationship decline, the distance metric has a higher sufficiency.

All of these methods treat distance metric as the index of the approximate relationship. As reduction set increases a new attribute, algorithm will recalculate the neighborhood system again. But these methods using the same distance threshold to measure spaces of different dimensionality, may select the attributes which has lower classification ability during searching reduction attributes. It possibly brings the error of neighborhood granulation.

To solve this issue, this study proposes a method based on the network sequence hierarchical granulation model, which directly calculates neighborhood system of the higher dimension space with the property of the model. It avoids the loss of information which is resulted by distance calculation. And using the classification power of each attribute, the method can simplify the searching process.

2. Hierarchical Granulation Model

2.1. Neighborhood Granulation

The study uses the distance metric as the main metric to construct the neighborhood model.

Definition 1. Given an N dimension space O, and two samples x_i, $x_j \in O$. If the distance metric is δ, x_j and x_i must satisfy the following properties:

(1) $\delta(x_i, x_j) \geq 0$, *if and only if* $i = j$, $\delta(x_i, x_j) = 0$;

(2) $\delta(x_i, x_j) = \delta(x_j, x_i)$;

(3) $\delta(x_i, x_j) + \delta(x_j, x_k) \geq \delta(x_i, x_k)$;

All of three properties are reflexivity, symmetry, non-transitive of neighborhood relationship.

Definition 2. Given an N dimension space O, x_i, $x_j \in O$, if the distance threshold is ε, and

$$\delta(x_i, x_j) < \varepsilon \tag{1}$$

It is called that x_j satisfies some neighborhood relationship S about x_i, denoted by $x_i S x_j$.

The distance formula usually uses the p-norm:

$$\delta(x_i, x_j) = \left(\sum_{k=1}^{N} \left(x_i^k - x_j^k \right)^p \right)^{\frac{1}{p}} \tag{2}$$

In the formula, x_i^k means the k-th attribute value of the sample x_i. This study takes the Euclidean distance, that p equals 2.

Definition 3. Given an N dimension non-empty finite universe U. To an arbitrary $x_i \in U$, there is a set

$$N_S(x_i) = \left\{ x_j \middle| x_i S x_j, x_j \in U \right\} \tag{3}$$

It is called that $N_S(x_i)$ is the neighborhood set of x_i [5].

Theorem 1. Given an N dimension non-empty finite universe U. To arbitrary x_i, $x_j \in U$, there is a neighborhood relationship S, and $N_S(x_i) = N_S(x_j)$. Then x_i equals to x_j [6].

According to theorem 1, neighborhood relationship S can be turn to equivalence relation E. So that algorithm can construct the network sequence hierarchical granulation system.

Definition 4. Given an N dimension non-empty finite universe U. To an arbitrary $x_i \in U$, there is a set

$$R_E(x_i) = \{ x_j \middle| \delta(x_i) = \delta(x_j), x_j \in U \} \tag{4}$$

$R_E(x_i)$ is called the equivalent granule of x_i.

Definition 5. Given an N dimension non-empty finite universe U. To an arbitrary set $X \in U$, its upper and lower approximate sets are respectively defined as

$$\overline{R}_E X = \{ x_i \middle| x_i \in X, R_E(x_i) \bigcap \mathrm{X} \neq \phi \} \tag{5}$$

$$\underline{R}_E X = \{x_i \big| x_i \in X, R_E(x_i) \subseteq \mathrm{X}\} \quad (6)$$

Upper and lower approximate sets respectively describe the concept of destination set in different ways. In other words, it's a process of approaching the unknown set with known concept. Lower approximate set is a subset of *X*, which reflects the determined concept of *X*. However, the intersection between the upper approximate set and *X* is non-empty, which reflects the uncertainty of the boundary of *X*.

2.2. Neighborhood Granulation

YY. Yao has proposed several hierarchical granulation models in literature [7]. The structure of the network sequence hierarchical granulation model is shown in Figure 1.

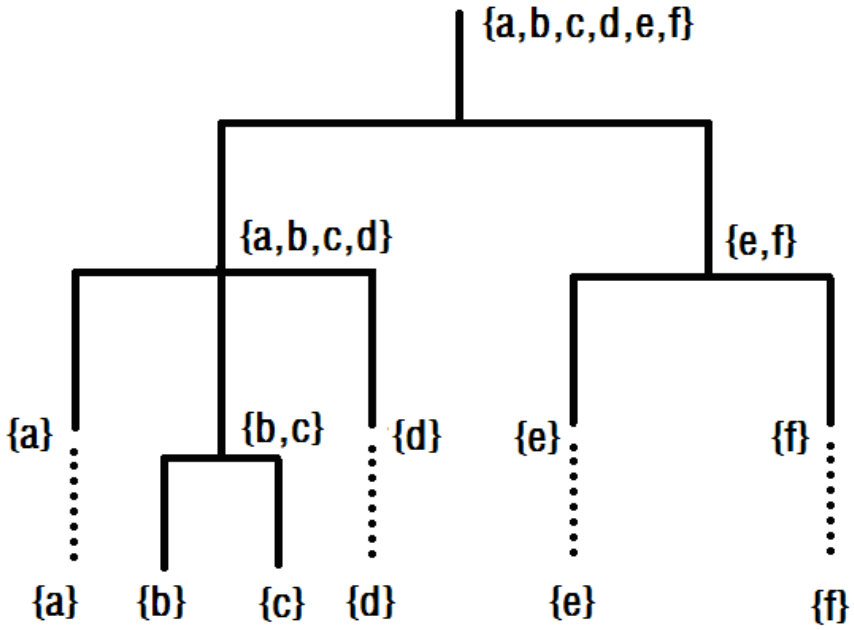

Figure 1. Network Sequence Structure

In Figure 1, the model has 4 levels. The domain $U=\{a,b,c,d,e,f\}$, the attribute set $A=\{A_1, A_2, A_3, A_4\}$. There is equivalence relationship $U \times U=E_{\{\}} \subset E_{A1} \subset E_{A1}E_{A2} \subset E_{A1}E_{A2}E_{A3}E_{A4}$ from higher level to lower level. Obviously as the level decreasing, the equivalence granules get smaller and the partitions get more meticulous. The attribute set of high level is the intersection of attribute sets in low level. And the equivalence granules of low level are always the intersection of the equivalence granules in high level [8-9].

Similarly, the granulation model can be constructed as shown in Figure 2.

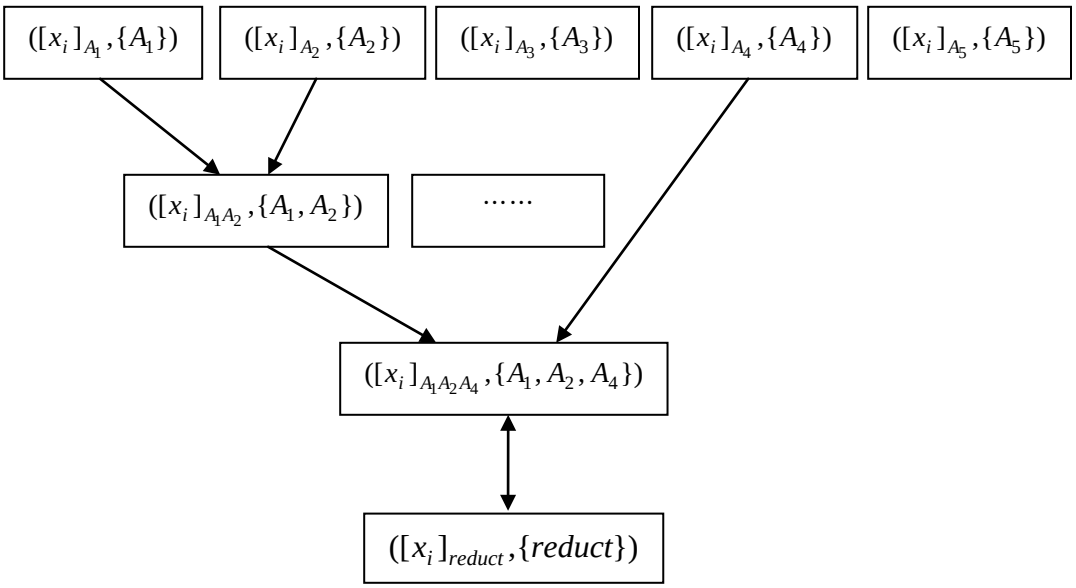

Figure 2. The Granulation Structure of Attribute Reduction

In Figure 2, each level means a kind of granulation degree. To the samples in the same level, their neighborhood granulation is the same. The process of getting the attribute reduction will deepen the granulation.

3. Attribute Reduction Based on Hierarchical Granulation

Attribute reduction using the structure of Figure 2 need calculate the granules of different levels. The process considers the classification power as the index to decide whether an attribute set is a reduction set or not. This study proposes the classification precision of rough set as the heuristic information.

Definition 6. Given an N dimension non-empty finite universe U. To the attribute set $A=C\cup D$, C is the condition attribute set, and D is the decision attribute set. D partitions the domain to n equivalence classes, denoted by $\{D_i|i=1, 2,\ldots,n\}$. If there is equivalence relationship $E\in R$, and a condition attribute c, its classification precision is defined as [10-12]

$$\gamma_c(D)=\sum_{i=1}^{n}\frac{card\left(\underline{R}_c(D_i)\right)}{card(U)} \tag{7}$$

$\gamma_c(D)$ reflects the classification effect of E relative to the decision classes. $card(\underline{R}_E(D_i))$ means the number of the samples whose lower approximate set is a subset of D_i. The larger the number, the better the classification power of E. The larger the number, the more important the attribute c. $\gamma_c(D)$ is an important index to search the attribute reduction.

Definition 7. Given an N dimension non-empty finite universe U. To the attribute set $A=C\cup D$, C is the condition attribute set, and D is the decision attribute set. If *reduct* is a reduction set of A, it will satisfy

(1) $\forall a\in reduct$, $\gamma_{reduct}(D)>\gamma_{reduct-a}(D)$;

(2) $\gamma_{reduct}(D) = \gamma_C(D)$.

From definition 7, it can be summarized that every one of the attribute reduction set is necessary, and the classification power of reduction set is the same as the classification power of the attribute set.

Although $\gamma_c(D)$ can pick out all the attributes which has high classification precision, there is overlap of classification power among these attributes. Not all the attributes which has high classification precision will be contained in the reduction set. It can define the significance of an attribute which measures the influence of classification power of the new increasing attribute to decide whether select it or not.

Definition 8. Given an N dimension non-empty finite universe U. To the attribute set $A=C \cup D$, C is the condition attribute set, and D is the decision attribute set. Given an attribute set T, and an attribute $a \in C-T$. Then the significance of a relative to T is

$$SIG(a,T,D) = \gamma_{T \cup a}(D) - \gamma_T(D) \tag{8}$$

It is worthy to note that in the past literatures distance of different space is measured with the same threshold ε. It will lead to the error of the reduction result, as being shown in Figure 3.

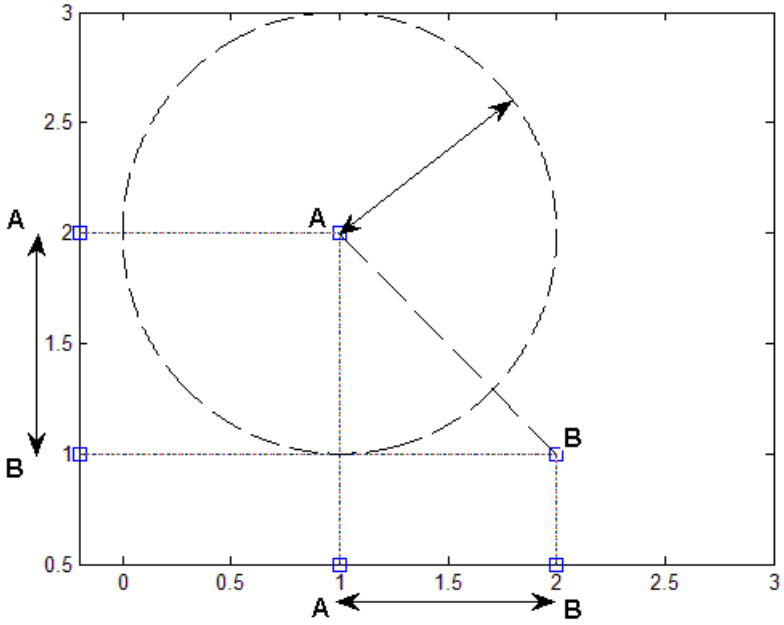

Figure 3. The Distance of Different Spaces

When the distance threshold takes 1, points A and B are similar each other in the one dimension space, but separated in the two dimension space. The range of the distance value increases. So it is not proper to take a same threshold. The issue can be solved with the property of hierarchical granulation model.

Theorem 2. Given an N dimension non-empty finite universe U, the equivalence relationship E_1,E_2, and attributes c_1, $c_2 \subseteq C$. If $E_1=\{c_1\}$, $E_2=\{c_2\}$, then $R_{E1}(x_i) \cap R_{E2}(x_i) = R_{E1 \cup E2}(x_i)$.

Proof. $\because$ $c_1 \subseteq c_1 \cup c_2$ and $c_2 \subseteq c_1 \cup c_2$, $\therefore$ $R_{c1 \cup c2}(x_i) \subseteq R_{c1}(x_i)$ and $R_{c1 \cup c2}(x_i) \subseteq R_{c2}(x_i)$. According to the definition of the equivalence relationship, $R_{c1}(x_i) \cap R_{c2}(x_i) = R_{c1 \cup c2}(x_i)$.

Using theorem 2, it can construct the equivalence granules system of high level space without calculating the distance. It makes the process can reduce the loss of information.

The study designs a continuous information attribute reduction algorithm based on hierarchical granulation. The algorithm steps are as follows:

Input: the domain U, the condition attribute set $C=\{a_i|i=1,\ldots,m\}$, the decision class $D=\{D_j|j=1,\ldots,n\}$.

Output: the attribute reduction set *reduct*.

(1) $reduct=\{\}$, and to arbitrary $a_i \in C$, calculate $R_{ai}(x_i)$, construct granulation model.

(2) To arbitrary $a_i \in C$, calculate $\gamma_{ai}(D)$. If $\gamma_{ai}(D) > 0$, then $a_i \rightarrow C'$.

(3) Pick out the attribute a_i which has $\max(\gamma_{ai}(D))$ from C', then $C'\text{-}a_i \rightarrow C'$.

(4) Calculate $SIG(a_i,reduct,D)$. If $SIG(a_i,reduct,D) > 0$, then $a_i \rightarrow reduct$; otherwise, delete this attribute.

(5) If $C'=\{\}$ or $\gamma_{reduct}(D)=\gamma_c(D)$, then stop; otherwise, turn to step (3).

4. Simulation and Analysis

The experiment selects several UCI data sets to test the performance of algorithm. Table 1 contains some characters of data sets.

Before the reduction, algorithm must take some pretreatment for the data sets. To eliminate the influence of different attribute value range, it needs to normalize every attribute.

To different data sets, the thresholds are different. If the value is too small, it will lead to sharp increase of the classification precision of single attribute, and the result of algorithm will have too many redundant attributes. If the threshold is too large, the result of algorithm will have a low classification precision.

By revising the threshold, the reduction set can satisfy the classification precision of decision classes. Table 2 is the reduction results of several UCI data sets. According to Table 2, the algorithm has a good reduction result without losing the classification precision.

Table 1. UCI Database

No.	Name	Number of samples	Number of attributes	Number of classes
1	Ionosphere	351	34	2
2	Sonar	208	60	2
3	Soybean	47	35	4
4	Wine	178	13	3

Table 2. The Results of Attribute Reduction based on Hierarchical Granulation

No.	Name	Threshold	Number of reduction /Number of attributes	Classification Precision
1	Ionosphere	0.00249	33/34	100%
2	Sonar	0.0201	15/60	99.038%
3	Soybean	0.02625	11/35	100%
4	Wine	0.01228	9/13	98.315%

In order to directly observe and analyze the results of above algorithm, 10 test threshold points are selected for each data set respectively. Figure 4-7 show the relationship between the threshold selection and reduction effect, classification power.

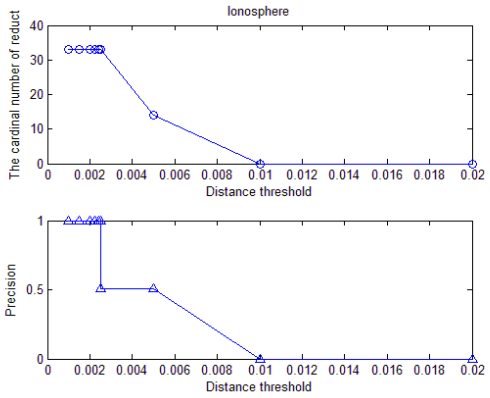

Figure 4. The Influence of the Distance Threshold of Lonosphere

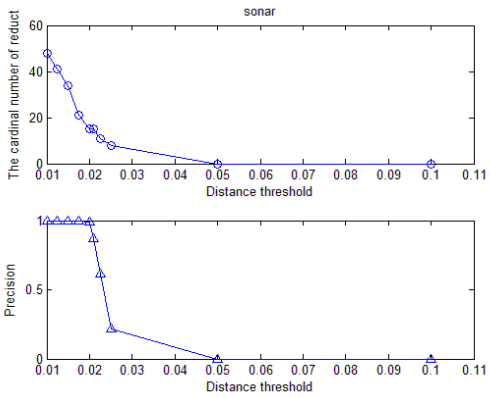

Figure 5. The Influence of the Distance Threshold of Sonar

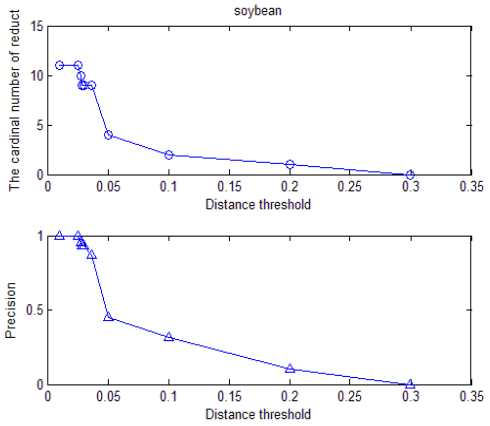

Figure 6. The Influence of the Distance Threshold of Soybean

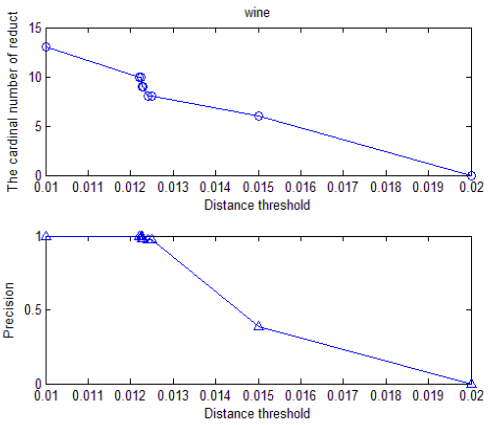

Figure 7. The Influence of the Distance Threshold of Wine

From Figure 4-7, it can be summarized that the distance threshold directly decides the reduction result. If the threshold is too large, the classification ability will decline, even leading to the reduction result is empty. If the threshold is too small, the reduction set will contain the redundant attributes. To different data sets, the influence range of the threshold is also different. As the set soybean, the range is wide. However, other ranges are narrower. It reflects that the approximation of each data set is different, and there is indiscernibility among the continuous information samples.

The experiments use the data sets to clustering after reduction which respectively uses the methods in literature [2] and in this study. Table 3 records the precision and running time of clustering. And algorithm one uses the algorithm of literature [2]. Algorithm two uses the algorithm of this study.

Table 3. Comparison of Reduction Results

Name	Algorithm One		Algorithm Two	
	Precision(%)	Time(s)	Precision(%)	Time(s)
Wine	73%	0.187447	73%	0.20862
Soybean	100%	0.0249	100%	0.007695
Ionosphere	63.38%	0.173628	68.93%	0.159007
Sonar	58.99%	0.11481	61.51%	0.051381

From Table 3, it can be analyzed that the clustering precision which uses the method proposed in the study is better than another. And the run time of the latter is also less than the former. All of these prove attribute reduction algorithm based on hierarchical granulation has a higher precision, and reduces the loss of information.

5. Conclusion

The conventional algorithms based on neighborhood system usually measure the similarity of different dimension spaces with the same distance threshold witch will product the loss of information inevitably. This study proposes an improved algorithm based on hierarchical granulation for continuous information attribute reduction. It calculates the granulation system with the properties of hierarchical granulation model. The experiment with UCI data proves that the algorithm can effectively achieve continuous information of attribute reduction.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (61105082), Natural Science Foundation of Jiangsu Province (BK2012470), Natural Science Foundation of Jiangsu Province Education Commission (11KJB120001), 'Climbing Project' Science Foundation of Nanjing University of Posts and Telecommunications (NY212093) and Qing Lan Project of Jiangsu Province.

References

[1] Z. Pawlak, "Rough Set", International Journal of Computer and Information Science, vol. 11, no. 5, **(1982)**.
[2] H. Qinghua, Y. Daren and X. Zongxia, "Numerical Attribute Reduction Based on Neighborhood Granulation and Rough Approximation", Journal of Software, vol. 19, no. 3, **(2008)**.
[3] H. Qinghua, Z. Hui and Y. Daren. "Efficient Symbolic and Numerical Attribute Reduction with Neighborhood Rough Sets", PR & AI, vol. 21, no. 6, **(2008)**.
[4] F. Lin, Y. Yongle, G. Shirong, Y. Jun and S. Li, "An approach for real domain rough sets model and attribute reduction", Control and Decision, vol. 27, no. 4, **(2012)**.
[5] Y. Y. Yao, "Neighborhood systems and approximate retrieval", Information Sciences, vol. 176, **(2006)**.
[6] Y.Y. Yao, "Generalized rough set models", Rough Sets in Knowledge Discovery", Polkowski, L. and Skowron, A. (Eds.), Physica-Verlag, Heidelberg, **(1998)**.
[7] Y.Y. Yao, "Information granulation and rough set approximation", International Journal of Intelligent Systems, vol. 16, no. 1, **(2001)**.
[8] Y.Y. Yao and J. T. Yao, "Granular Computing as a Basis for Consistent Classification Problems", Proceedings of PAKDD'02 Workshop on Toward the Foundation of Data Mining, **(2002)**.
[9] Y.Y. Yao, "Rough Sets, Neighborhood Systems, and Granular Computing", Proceedings of the 1999 lEEE Canadian Conference on Electrical and Computer Engineering, Shaw Conference Center, Edmonton, Alberta, Canada, **(1999)** May 9-12.
[10] B. Walczak, D.L. Massart, "Rough sets theory", Chemometrics and Intelligent Laboratory Systems, vol. 47, **(1999)**.

[11] Q. Hua, L. Zhang, D. Chen, W. Pedrycz and D. Yu, "Gaussian kernel based fuzzy rough sets: Model, uncertainty measures and applications", International Journal of Approximate Reasoning, vol. 51, **(2010)**.
[12] Q. Hu, S. An and D. Yu, "Soft fuzzy rough sets for robust feature evaluation and selection", Information Sciences, vol. 180, **(2010)**.

Authors

Long Chen received the ME degree in Pattern Recognition and Intelligent System from Nanjing University of Posts and Telecommunications. His research interests include data mining, rough set theory, pattern recognition and intelligent computation.

Tengfei Zhang received bachelor degree from Henan University in 2002. And he respectively received master and Ph.D degree from Shanghai Maritime University in 2004 and 2007. He is currently an associate professor and supervisor for master student at Nanjing University of Posts and Telecommunications. His research interests include intelligent information processing and pattern recognition.