



A Knowledge Innovation Algorithm Based on Granularity

Taishan Yan (Corresponding author)

School of Computer Science and Engineering, Xi'an University of Technology

Xi'an, Shaanxi 710048, China

E-mail: yantaishan163@tom.com

Duwu Cui

School of Computer Science and Engineering, Xi'an University of Technology

Xi'an, Shaanxi 710048, China

E-mail: Cuidw@xaut.edu.cn

The research is supported by the National Natural Science Foundation of China under Grant No. 60873035 and No.60743009, the Natural Science Foundation of Shaanxi of China under Grant No.2006F43.

Abstract

The structure of human knowledge is regarded as granule state by rough sets theory. Granularity is used to denote this structure of knowledge. Knowledge itself evolves ceaselessly as creatures. The mechanism of knowledge evolution includes the productive mechanism for new knowledge and the natural choice mechanism for the selection of the superior and the elimination of the inferior. Knowledge innovation is an important step of knowledge evolution course. Based on knowledge granularity, a knowledge innovation method is proposed in this paper. The main idea of this method is to constitute the partition granularity of knowledge base space ceaselessly depend on the measure consistency of attribute, till the sort of every granules in the granularity is only one. In the algorithm, the only one computation work is to measure the consistency of attributes in knowledge base space. So the numerical calculation is little and the time complexity is low. Experiments were taken on the imperfect knowledge base space of day weather classification by this algorithm. The working course of the algorithm was explained in the example. The successful results show that this algorithm is valid and feasible.

Keywords: Knowledge innovation, Imperfect knowledge base, Granule, Granularity, Consistency

1. Introduction

Knowledge is one of human being's own concept which belongs to epistemology. It is the accumulation or induction of experience and contains almost all the meaning of data, information, knowledge and intelligence (Zhu, 2000; Zhong, 2007). In the course of human evolution, knowledge plays a vital role. Humans have to obey the natural law, at the same time, formed a knowledge system gradually. The knowledge system had a great influence on human evolution and made human evolution rise rapidly. The famous science-philosopher Karl · Popper (Liu *et al.*, 2007; Shu *et al.*, 2005; Fu *et al.*, 2005) thought that the selection of scientific theory is similar to Darwin's natural selection. Our knowledge is always constituted by hypothesis, and passed in the struggle to survive down and showed its relative adaptability. The hypothesis which didn't adapt was eliminated in the competition. Karl · Popper put biological evolution together with scientific development, explained how the science evolve and created his knowledge evolutionism. Knowledge evolutionism tells us that knowledge can be regarded as a changing objective knowledge world, in which there is no eternal, immutable knowledge. That is to say, knowledge is incomplete, knowledge itself evolves ceaselessly. There are two mechanisms in knowledge evolution, one is the productive mechanism for new knowledge, and the other is the natural choice mechanism for the selection of the superior and the elimination of the inferior. Knowledge evolution has its own characteristics and rules. In the course of knowledge evolution, knowledge innovation is a very important aspect.

At present, the research on natural computing theory largely focused on the biological natural selection level ((Liu *et al.*,

2007; Fogel, 1994). However, the research on knowledge evolution theory and algorithm is bound to become a new research direction of natural computation. In order to provide a preliminary exploration for knowledge innovation and support the research on knowledge evolution algorithm, a knowledge innovation method based on knowledge granularity was proposed in this paper.

2. Granularity Concept

Knowledge is examined from a new view in Z Paw Lak's rough set theory (PAWLAK, 1991), and it is generally regarded as a granular structure which is expressed by granularity. Granularity is a very important conception in intelligent information processing. With the help of granularity, knowledge mining and knowledge processing can be carried through from a different perspective, knowledge innovation and knowledge evolution can be realized.

Definition 1 A granule can be defined as any subset, class, object or cluster of a universe (Yao, 2008). A basic granule is considered to be indivisible or there is no need to divide. Elements in a granule are drawn together by indistinguishability, similarity, proximity, functionality, etc. Different fields have different units corresponding to granules (Yao, 2007; Yi *et al.*, 2008). That is to say the elements of a granule are regarded as a cluster, which is accordant with the basic viewpoint of rough set theory (PAWLAK, 1982,1991; Dun *et al.*, 2006; Yang *et al.*, 2006; QUINLAN, 1987).

Definition 2 Giving a granule set, and all the elements of a given field can be found in a element of the granule set, then the set can be called granularity of the given field (Dun *et al.*, 2006). Granularity calculation is knowledge-oriented, which is different from traditional numerical calculation. Traditional numerical calculation is data-oriented (Bernard *et al.*, 2006).

How to construct the granularity of a given knowledge base space? Combine the view of rough set and decision-making tree, the granularity of a given knowledge base space can be constructed according to the attribute value. Given a knowledge base space (U, C, D, V, f) , where, U is the set of objects, C is the set of condition attributes, D is the set of decision attributes. $V = \bigcup_{a \in A} V_a$, V_a is the value region of attributes. $f: U \times (C \cup D) \rightarrow V$ is a evaluating

function, it gives each attribute of all objects a value, namely, $\forall a \in C \cup D, x \in U, f(x, a) \in V_a$. Select attribute $a \in C \cup D$ randomly, $|a_v|$ delegates the set formed by the objects with attribute a whose value is v . If $v(a) = \{v_a^1, v_a^2, \dots, v_a^k\}$, then set $\{ |a_{v_a^i}| : i=1, 2, \dots, k \}$ forms a granularity of the knowledge base space.

For instance, table 1 denotes a Knowledge base space of weather classification (Shi, 1998; Miao *et al.*, 2002), in which weather is sorted N or P according to condition attribute set $\{a_1(\text{Outlook}), a_2(\text{Temperature}), a_3(\text{Humidity}), a_4(\text{Windy})\}$.

In table 1, the values of condition attribute $Outlook(a_1)$ are *Sunny*, *Overcast* or *Rain*, their corresponding granule is $|a1_{Sunny}| = \{1,2,8,9,11\}$, $|a1_{Overcast}| = \{3,7,12,13\}$, $|a1_{Rain}| = \{4,5,6,10,14\}$. The values of condition attribute $Temperature(a_2)$ are *Hot*, *Mild* or *Cool*, their corresponding granules are $|a2_{Hot}| = \{1,2,3,13\}$, $|a2_{Mild}| = \{4,8,10,11,12,14\}$, $|a2_{Cool}| = \{5,6,7,9\}$. The values of condition attribute $Humidity(a_3)$ are *High*, *Normal* or *Low*, their corresponding granules are $|a3_{High}| = \{1,2,3,4,8,12,14\}$, $|a3_{Normal}| = \{5,6,7,9,10,11,13\}$, $|a3_{Low}| = \{\}$. The values of condition attribute $Windy(a_4)$ are *True* or *False*, their corresponding granules are $|a4_{True}| = \{2,6,7,11,12,14\}$, $|a4_{False}| = \{1,3,4,5,8,9,10,13\}$. The values of decision-making attribute $Class(d)$ are *N* or *P*, their corresponding granules are $|d_N| = \{1,2,6,8,14\}$, $|d_P| = \{3,4,5,7,9,10,11,12,13\}$. So, the granularities formed by condition attribute $Outlook(a_1)$, $Temperature(a_2)$, $Humidity(a_3)$, $Windy(a_4)$ and decision-making attribute $Class(d)$ are:

$$G_{a_1} = \{ \{1,2,8,9,11\}, \{3,7,12,13\}, \{4,5,6,10,14\} \}$$

$$G_{a_2} = \{ \{1,2,3,13\}, \{4,8,10,11,12,14\}, \{5,6,7,9\} \}$$

$$G_{a_3} = \{ \{1,2,3,4,8,12,14\}, \{5,6,7,9,10,11,13\} \}$$

$$G_{a_4} = \{ \{1,3,4,5,8,9,10,13\}, \{2,6,7,11,12,14\} \}$$

$$G_d = \{ \{1,2,6,8,14\}, \{3,4,5,7,9,10,11,12,13\} \}$$

Definition 3 Given a knowledge base space (U, C, D, V, f) , $X \subseteq C$ is a sub-set of condition attribute. $CON(X \rightarrow D)$ denotes the consistency of $X \rightarrow D$. Its definition is:

$$CON(X \rightarrow D) = |X \cup D|/|X|$$

Where, $|X| = |ND(X)|$ denotes the base of $ND(X) \subseteq U \times U$.

$CON(X \rightarrow D)$ is a kind of measurement for expressing or forecasting D by X . Generally, $CON(X \rightarrow D)$ is more big, X is more excellent in the forecasting attribute set of D .

For instance, in table 1, the combined granularities of each condition attribute $a_i (1 \leq i \leq 4)$ and decision attribute d are:

$$G_{a_1 \cup d} = \{\{1,2,8\}, \{9,11\}, \{3,7,12,13\}, \{4,5,10\}, \{6,14\}\}$$

$$G_{a_2 \cup d} = \{\{1,2\}, \{3,13\}, \{4,10,11,12\}, \{8,14\}, \{5,7,9\}, \{6\}\}$$

$$G_{a_3 \cup d} = \{\{1,2,8,14\}, \{3,4,12\}, \{5,7,9,10,11,13\}, \{6\}\}$$

$$G_{a_4 \cup d} = \{\{1,8\}, \{3,4,5,9,10,13\}, \{2,6,14\}, \{7,11,12\}\}$$

And the consistency $CON(a_i \rightarrow d) (1 \leq i \leq 4)$ is:

$$CON(a_1 \rightarrow d) = |a_1 \cup d|/|a_1| = 0.6364$$

$$CON(a_2 \rightarrow d) = |a_2 \cup d|/|a_2| = 0.5588$$

$$CON(a_3 \rightarrow d) = |a_3 \cup d|/|a_3| = 0.6327$$

$$CON(a_4 \rightarrow d) = |a_4 \cup d|/|a_4| = 0.5800$$

3. Knowledge Innovation Algorithm Based on Granularity

3.1 Connotation of Knowledge Innovation

Knowledge is expressed by the following formula in the knowledge engineering field (Zhang *et al.*, 2005):

$$K = F + R + C$$

Where, K denotes knowledge, F denotes fact, R denotes rule, C denotes concept. Fact indicates the description for the state, attribute and feature of things. Rule indicates the causal relationship between premise and conclusions. Concept indicates the semantic description of facts and rules. Knowledge innovation indicates a new combination or a new transformation of knowledge factors. This is new knowledge production course using old knowledge fully. In other words, the emergence of new knowledge is not isolated, new knowledge is based on relevant existing knowledge.

3.2 Basic Idea of Algorithm

In knowledge base space (U, C, D, V, f) , if there is at least an object $x \in U$ and it has at least one attribute $a \in C \cup D$ with value $f(x, a)$ which is not in this knowledge base space, the knowledge base space can be called imperfect knowledge base space. Knowledge base space (U, C, D, V, f) may be abbreviated as (U, C, D) . Generally, knowledge base spaces are imperfect and perfect knowledge base space is only a special form of imperfect knowledge base space. Knowledge may be innovated and common knowledge base space can evolve.

For a given imperfect knowledge base space (U, C, D) , if a granularity G is consistent with the following condition: $\forall g \in G, |d(g)| = 1$. Namely, the values of decision-making attributes of every granule in granularity G are same. That means all granules in granularity G belong to a same class. So, after describing the granularity properly, the new knowledge rules can be gained according to the granularity. The algorithm uses consistency of attribute as a measure, constitute the partition granularity of knowledge base space ceaselessly, till the sort of every granules in the granularity is only one. This is a coarse-to-fine partition process for knowledge base space, and a tree structure is formed at last.

3.3 Steps and Working Flow of Algorithm

The working flow of knowledge innovation algorithm based on granularity can be described as follows.

Input- Imperfect knowledge base space (U, C, D)

Output-New knowledge which is not included in the old knowledge base space.

Step1: Calculate the consistency of each attribute $x_i \rightarrow D$ in knowledge base space (U, C, D) .

Step2: Choose the attribute with the biggest consistency value according to the result of Step1. Take this attribute as the

most important attribute a_m . If all attributes have a same consistency value, the coverage should be chosen as the most important attribute a_m .

Step3: Constitute the partition granularity of knowledge base space depend on the most important attribute a_m .

Step4: Judge if all the objects in a granule belong to a same class or not. For every granule, if all the objects belong to a same class, turn to Step7; otherwise turn to Step5.

Step5: Delete attribute a_m from attribute set.

Step6: Repeat the course of Step1~ Step4 till all the objects in a granule belong to a same class.

Step7: Describe every granule and output the new knowledge rules.

Step8: Stop.

The working flow can be described as Figure 1.

3.4 Time Complexity of Algorithm

In knowledge innovation algorithm based on granularity, the only one calculation object is attribute consistency. The time complexity is mainly decided by attributes combination in knowledge base space. In the worst case, the number of attributes in each calculation is $|C|, |C-1|, \dots, 1$. So, the most total calculation number is:

$$|C| + (|C|-1) + \dots + 1 = |C|(|C|+1)/2$$

If the influence of objects number is not considered, in the worst case, the time complexity of algorithm is $O(|C|^2)$.

4. Example

We take the imperfect knowledge base space of day weather classification shown in table 1 as an example, test the validity of above-mentioned knowledge innovation algorithm based on granularity.

The calculation results of example 2 tell us that consistency $CON(a_1 \rightarrow d)$ is the biggest one. So, attribute a_1 is selected to constitute the partition granularity of knowledge base space, as Figure 2 shows.

We can see from Figure 2 that there are two different conclusion attribute values in the first granule and the third granule. And there is only one conclusion attribute value in the second granule. So, it is only need to partition the first and the third granule again and not need to partition the second granule any more. After attribute a_1 is deleted, the objects in tow granules are shown in table 2 and table 3.

In the sub-space shown in table 2, each consistency is:

$$CON(a_2 \rightarrow d) = |a_2 \cup d|/|a_2| = 0.7778$$

$$CON(a_3 \rightarrow d) = |a_3 \cup d|/|a_3| = 1.0000$$

$$CON(a_4 \rightarrow d) = |a_4 \cup d|/|a_4| = 0.5385$$

As $CON(a_3 \rightarrow d) = 1.0000$ is the biggest one, so, attribute a_3 is selected to constitute the partition granularity of sub-space.

In the sub-space shown in table 3, each consistency is:

$$CON(a_2 \rightarrow d) = |a_2 \cup d|/|a_2| = 0.5385$$

$$CON(a_3 \rightarrow d) = |a_3 \cup d|/|a_3| = 0.5385$$

$$CON(a_4 \rightarrow d) = |a_4 \cup d|/|a_4| = 1.0000$$

As $CON(a_4 \rightarrow d) = 1.0000$ is the biggest one, so, attribute a_4 is selected to constitute the partition granularity of sub-space.

At last, a tree structure is formed, as Figure 3 shows.

In Figure 3, the leaf node set is the namely granularity with a knowledge space in which each granule only has one conclusion attribute value. This granularity is:

$$G = \{\{1,2,8\}, \{9,11\}, \{3,7,12,13\}, \{6,4\}, \{4,5,10\}\}$$

Then, describe all granules according to the path from base node to leaf nodes and the new knowledge rules can be gained. As table 4 shows.

The results in table 4 tell us that knowledge innovation is realized by the algorithm. The correctness and validity of algorithm is proved. There is only one computation namely the measure consistency of attribute in the algorithm, the time complexity is low. So, the feasibility of algorithm can be ensured.

5. Conclusion

In this paper, the concept of granularity was expounded and a knowledge innovation method based on granularity was proposed. The numerical calculation work of this algorithm is little. It is entirely feasible. Examples show that the algorithm can realize knowledge innovation successfully using knowledge of imperfect knowledge base. So, the algorithm is correct and valid. Future work is to study how to evaluate and test the existing knowledge and new knowledge generated by the algorithm. The excellent new knowledge will be used to update the original knowledge base, and consequently realize knowledge evolution.

Acknowledgments

Support from the National Natural Science Foundation of China under Grant No.60873035 and No.60743009 and the Natural Science Foundation of Shaanxi of China under Grant No.2006F43 is gratefully acknowledged.

References

- Bernard, C., Tai, P.C., Harrison, R., & Pan, Y. (2006). *FIK Model: Novel Efficient Granular Computing Model for Protein Sequence Motifs and Structure Information Discovery*. Sixth IEEE Symposium on Bio-Informatics and Bio-Engineering, 27(2), 1-7.
- Dun, Y., Zhang, X., Sun, H., & Zhao, L. (2006). *A rules mining method based on granularity*. Journal of Lanzhou University of Technology, 32(1), 105-108.
- Fogel, D. B. (1994). An introduction to simulated evolutionary optimization. *IEEE Trans. on Neural networks*, 5(1), 3-14.
- Fu, J., Zhou, C., & Jiang, G. (2005). *Speculation and Refutation--Growth of Scientific Knowledge*. Shanghai: Shanghai Translation Publishing House.
- Liu, C., Yang S., & Zhang Y. (2007). Knowledge evolution strategies. *Systems Engineering and Electronics*, 29(6), 1017-1020.
- Miao, D., & Fan, S. (2002). *The Calculation of Knowledge Granulation and Its Application*. Systems Engineering Theory and Practice, 1, 48-55.
- PAWLAK, Z. (1982). *Rough sets*. Int J Comput Inform Sci, 11(5), 341-356.
- PAWLAK, Z. (1991). *Rough sets—theoretical aspects of reasoning about data*. Boston: Kluwer Academic Publishers.
- QUINLAN, J. R. (1987). *Simplifying decision trees*. International Journal of Man-Machine Studies, 27, 221-234.
- Shi, Z. (1998). *Advanced Artificial Intelligence*. Beijing: Science Press.
- Shu, W., & Zuo R. (2005). *Objective knowledge--Research on A Theory of Evolution*. Shanghai: Shanghai Translation Publishing House.
- Yang, W., Ye, X., Tang, Y., & Wei, P. (2006). *Granularity of Knowledge Computed by Genetic Algorithms Based on Rough Sets Theory*. Journal of Nanchang Institute of Technology, 25(2), 97-121.
- Yao, J.T. (2008). *Granular Computing: A New Paradigm in Information Processing*. North American Simulation Technology Conference 2008 (NASTEC'08), 5-6.
- Yao, Y.Y. (2007). *The art of granular computing*. Proceedings of the International Conference on Rough Sets and Emerging Intelligent Systems Paradigms, 101-112.
- Yi, Z., & Ning, Z. (2008). *On Granular knowledge Structures*. Progress of Advanced Intelligence: Proceedings of 2008 International Conference on Advanced Intelligence, 28-33.
- Zhang, F., & He, C. (2005). *The Theory and Path of Knowledge innovation*. Journal of Chinese Academy of Sciences Institute, 20(5), 389-394.
- Zhong, Y. (2007). Knowledge concepts and measures in views of information – knowledge - intelligence ecosystem. *Journal of Frontiers of Computer Science and Technology*, 1(2), 129-137.
- Zhu, Z. (2000). Research on the mechanism of knowledge evolution and knowledge innovation. *Journal Research and Development Management*, 12(6), 16-19.

Figures and tables

Table 1. Knowledge base space of weather classification

No	Condition attributes				Class (d)
	a ₁	a ₂	a ₃	a ₄	
1	Sunny	Hot	High	False	N
2	Sunny	Hot	High	True	N
3	Overcast	Hot	High	False	P
4	Rain	Mild	High	False	P
5	Rain	Cool	Normal	False	P
6	Rain	Cool	Normal	True	N
7	Overcast	Cool	Normal	True	P
8	Sunny	Mild	High	False	N
9	Sunny	Cool	Normal	False	P
10	Rain	Mild	Normal	False	P
11	Sunny	Mild	Normal	True	P
12	Overcast	Mild	High	True	P
13	Overcast	Hot	Normal	False	P
14	Rain	Mild	High	True	N

Table 2. Subspace(*value*(Outlook)=*Sunny*)

No	Condition attributes			Class (d)
	a ₂	a ₃	a ₄	
1	Hot	High	False	N
2	Hot	High	False	N
8	Mild	High	False	N
9	Cool	Normal	False	P
11	Mild	Normal	True	P

Table 3. Subspace(*value*(Outlook)=*Rain*)

No	Condition attributes			Class (d)
	a ₂	a ₃	a ₄	
4	Mild	High	False	P
5	Cool	Normal	False	P
6	Cool	Normal	True	N
10	Mild	Normal	False	P
14	Mild	High	True	N

Table 4. New knowledge rules

No	Condition attributes				Class (d)
	a_1	a_2	a_3	a_4	
1	Sunny	Random	High	Random	N
2	Sunny	Random	Normal	Random	P
3	Overcast	Random	Random	Random	P
4	Rain	Random	Random	True	N
5	Rain	Random	Random	False	P

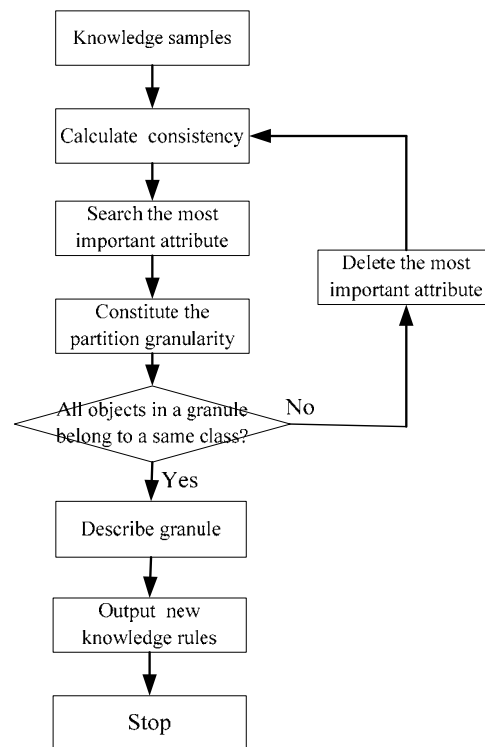
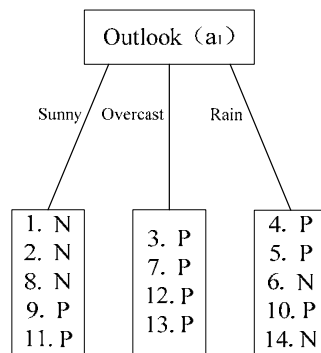


Figure 1. Working flow of Knowledge Innovation Algorithm Based on Granularity

Figure 2. Granularity of knowledge space formed by a_1

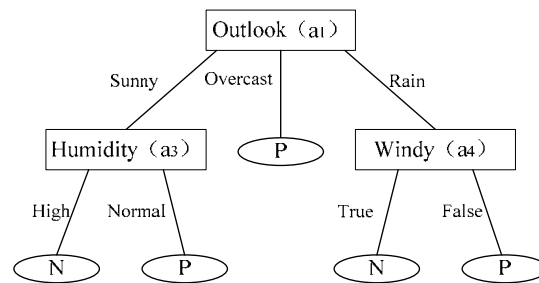


Figure 3. A tree structure gained by granularity calculation