Axiomatic Approach of Knowledge Granulation in Information System

Jiye Liang[1] and Yuhua Qian[2]

Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education
School of Computer and Information Technology, Shanxi University
Taiyuan, 030006, People's Republic of China
[1]ljy@sxu.edu.cn,
[2]jinchengqyh@126.com

Abstract. Granular computing is potentially in knowledge discovery and data mining etc. In this paper, by introducing a partial relation $\preceq'$ with set size character to information system, an axiom definition of knowledge granulation for information system is presented, some existing the definitions of knowledge granulation become special forms. These results will be very helpful for understanding the essence of knowledge granulation and uncertainty measurement in information system.

Keywords: Information system, partial relation, knowledge granulation.

1 Introduction

The notion of information system (sometimes called data tables, attribute-value system, knowledge representation system, etc.), provides a convenient tool for the representation of objects in terms of their attribute values. Rough set theory has been introduced to deal with inexact, uncertain, or vague knowledge in information system. The use of the indiscernibility relation results in knowledge granulation.

According to whether or not there are missing data (null values), the information system can be classified into two categories: complete and incomplete. The knowledge granulation of an information system gives a measure of uncertainty about its actual structure[1-6]. In general, knowledge granulation can represent discernibility ability of knowledge in rough set theory, the smaller knowledge granulation is, the stronger its discernibility ability is [7-9]. Especially, several measures in information system closely associated with granular computing such as measure, information entropy, rough entropy and knowledge granulation and their relationships are discussed by Liang et al. in [6, 9]. In [7], combination granulation and combination entropy in information system are proposed, their gain function possesses intuitionistic knowledge content characteristic.

In this paper, we devote to the axiomatic approach of knowledge granulation in information system. In section 2, we review some basic concepts of information system, and proposed a new partial relation $\preceq'$ with set size character to an information system. In section 3, we give the axiom definition of knowledge

A. Sattar and B.H. Kang (Eds.): AI 2006, LNAI 4304, pp. 1074–1078, 2006.
© Springer-Verlag Berlin Heidelberg 2006

granulation to measure uncertainty of knowledge in formation system, prove that several existing knowledge granulations are all special instances of the definition. Finally, section 4 concludes the whole paper.

2 Preliminaries

An information system is a pair $S = (U, A)$, where,

(1) U is a non-empty finite set of objects;
(2) A is a non-empty finite set of attributes;
(3) for every $a \in A$, there is a mapping a, $a : U \to V_a$, where V_a is called the value set of a.

For an information system $S = (U, A)$, if $\forall a \in A$, every element in V_a is a definite value, then S is called a complete information system. If V_a contains a null value for at least one attribute $a \in A$, then S is called an incomplete information system, otherwise it is complete.

Let $S = (U, A)$ be an information system, $P, Q \subseteq A$. $K(P) = \{S_P(x_i) \mid x_i \in U\}$, $K(Q) = \{S_Q(x_i) \mid x_i \in U\}$. We define a partial relation $\preceq^{'}$ with set size character as follows: $P \preceq^{'} Q$ $(P, Q \in A)$, if and only if, for $K(P) = \{S_P(x_1), S_P(x_2), \cdots, S_P(x_{|U|})\}$, there exists a sequence $K^{'}(Q)$ of $K(Q)$, where $K^{'}(Q) = \{S_Q(x_1^{'}), S_Q(x_2^{'}), \cdots, S_Q(x_{|U|}^{'})\}$, such that $|S_P(x_i)| \leq |S_Q(x_i^{'})|$. If there exists a sequence $K^{'}(Q)$ of $K(Q)$ such that $|S_P(x_i)| < |S_Q(x_i^{'})|$, then we will call that P is strict granulation finer than Q, and denote it by $P \prec^{'} Q$.

3 Axiomatic Construction of Knowledge Granulation

In 1979, the problem of fuzzy information granule was introduced by Zadeh in [10]. Especially, several measures in an information system closely associated with granular computing such as granulation measure, information entropy, rough entropy and knowledge granulation and their relationships were discussed in [6,9]. However, there exists no unified description for knowledge granulation. In the following, an axiom definition of knowledge granulation is given.

Definition 1. *Let $S = (U, A)$ be an information system, G be a mapping from the power set of A to the set of real numbers. We say that G is a knowledge granulation in an information system if G satisfies the following conditions:*

(1) $G(P) \geq 0$ for any $P \subseteq A$ (Non-negativity);

(2) $G(P) = G(Q)$ for any $P, Q \in A$ if there is a bijective mapping function $f : K(P) \to K(Q)$ such that $|S_P(u_i)| = |f(S_P(u_i))|$ $(\forall i \in \{1, 2, \cdots, |U|\})$, where $K(P) = \{S_P(x_i) \mid x_i \in U\}$ and $K(Q) = \{S_Q(x_i) \mid x_i \in U\}$ (Invariability);

(3) $G(P) < G(Q)$ for any $P, Q \in A$ with $P \preceq^{'} Q$ (Monotonicity).

Theorem 1 (Extremum). *Let $S = (U, A)$ be an information system, $\forall P \in A$, then if $U/P = \omega = \{\{u\} \mid u \in U\}$, $G(P)$ achieves minimum value; if $U/P = \delta = \{U \mid u \in U\}$, $G(P)$ achieves maximum value; where ω denotes identity relation, δ denotes universal relation.*

In [14,15,17,18], some different kinds of knowledge granulations were given, we can prove that these knowledge granulations are all special forms under definition 1.

Definition 2. [17] *Let* $S=(U,A)$ *be a complete information system,* $U/IND(A)=\{P_1,P_2,\cdots,P_m\}$*. Knowledge granulation of A is defined by*

$$GK(A)=\frac{1}{|U|^2}\sum_{i=1}^{m}|P_i|^2. \tag{1}$$

Theorem 2. *GK in definition 2 is a knowledge granulation under definition 1.*

Proof. (1) Obviously, it is non-negative;

(2) Let $P,Q\subseteq A$, then $U/IND(P)=\{P_1,P_2,\cdots,P_m\}$ and $U/IND(Q)=\{Q_1,Q_2,\cdots,Q_n\}$ in complete information system can be uniformly denoted by $U/SIM(P)=\{S_P(u_1),\ S_P(u_2),\cdots,S_P(u_{|U|})\}$ and $U/SIM(Q)=\{S_Q(u_1),\ S_Q(u_2),\cdots,S_Q(u_{|U|})\}$.

Suppose that there be a bijective mapping function f: $U/SIM(P)\rightarrow U/SIM(Q)$ such that $|S_P(u_i)|=|f(S_P(u_i))|$ $(i\in\{1,2,\cdots,|U|\})$ and $f(S_P(u_i))=S_Q(u_{j_i})$ $(j_i\in\{1,2,\cdots,|U|\})$, then we have that

$$\begin{aligned}GK(P)&=\tfrac{1}{|U|^2}\sum_{i=1}^{m}|P_i|^2=\tfrac{1}{|U|^2}\sum_{i=1}^{|U|}|S_P(u_i)|\\&=\tfrac{1}{|U|^2}\sum_{i=1}^{|U|}|S_Q(u_{j_i})|=\tfrac{1}{|U|^2}\sum_{i=1}^{|U|}|S_Q(u_i)|\\&=\tfrac{1}{|U|^2}\sum_{j=1}^{n}|Q_j|^2=GK(Q);\end{aligned}$$

(3) Let $P,Q\subseteq A$, $U/SIM(P)=\{S_P(u_1),S_P(u_2),\cdots,S_P(u_{|U|})\}$, $U/SIM(Q)=S_Q(u_1),S_Q(u_2),\cdots,S_Q(u_{|U|})$ and $P\prec' Q$, then for arbitrary $S_P(u_i)(i\leq|U|)$, there exists a sequence $\{S_Q(u_1^{'}),S_Q(u_2^{'}),\cdots,S_Q(u_{|U|}^{'})\}$ such that $|S_P(u_i)|<|S_Q(u_i^{'})|(i=1,2,\cdots,|U|)$. Hence, we obtain that

$$\begin{aligned}GK(P)&=\tfrac{1}{|U|^2}\sum_{i=1}^{m}|P_i|^2=\tfrac{1}{|U|^2}\sum_{i=1}^{|U|}|S_P(u_i)|\\&<\tfrac{1}{|U|^2}\sum_{i=1}^{|U|}|S_Q(u_i^{'})|=\tfrac{1}{|U|^2}\sum_{i=1}^{|U|}|S_Q(u_i)|\\&=\tfrac{1}{|U|^2}\sum_{j=1}^{n}|Q_j|^2=GK(Q).\end{aligned}$$

Thus, GK in definition 2 is the knowledge granulation under definition 1.

Definition 3. [14] *Let* $S=(U,A)$ *be an incomplete information system,* $U/SIM(A)=\{S_A(u_1),S_A(u_2),\cdots,S_A(u_{|U|})\}$*. Knowledge granulation of A is defined by*

$$GK(A)=\frac{1}{|U|^2}\sum_{i=1}^{m}|S_A(u_i)|. \tag{2}$$

Theorem 3. *GK in definition 3 is a knowledge granulation under definition 1.*

Proof. Similar to theorem 2, it can be proved.

Definition 4. [18] *Let* $S=(U,A)$ *be a complete information system,* $U/IND(A)=\{P_1,P_2,\cdots,P_m\}$. *Combination granulation of* A *is defined by*

$$CG(A)=\sum_{i=1}^{m}\frac{|P_i|}{|U|}\frac{C_{|P_i|}^2}{C_{|U|}^2}. \tag{3}$$

Theorem 4. *CG in definition 4 is a knowledge granulation under definition 1.*

Proof. (1) Obviously, it is non-negative;

(2) Let $P,Q\subseteq A$, then $U/IND(P)=\{P_1,P_2,\cdots,P_m\}$ and $U/IND(Q)=\{Q_1,Q_2,\cdots,Q_n\}$ in complete information system can be uniformly denoted by $U/SIM(P)=\{S_P(u_1),\ S_P(u_2),\cdots,S_P(u_{|U|})\}$ and $U/SIM(Q)=\{S_Q(u_1),\ S_Q(u_2),\cdots,S_Q(u_{|U|})\}$.

Suppose that there be a bijective mapping function f: $U/SIM(P)\rightarrow U/SIM(Q)$ such that $|S_P(u_i)|=|f(S_P(u_i))|$ $(i\in\{1,2,\cdots,|U|\})$ and $f(S_P(u_i))=S_Q(u_{j_i})$ $(j_i\in\{1,2,\cdots,|U|\})$, then we have that

$$\begin{aligned}CG(P)&=\sum_{i=1}^{m}\frac{|P_i|}{|U|}\frac{C_{|P_i|}^2}{C_{|U|}^2}=\sum_{i=1}^{|U|}\frac{|S_P(u_i)|}{|U|}\frac{C_{|S_P(u_i)|}^2}{C_{|U|}^2}\\&=\sum_{i=1}^{|U|}\frac{|S_Q(u_{j_i})|}{|U|}\frac{C_{|S_Q(u_{j_i})|}^2}{C_{|U|}^2}=\sum_{i=1}^{|U|}\frac{|S_Q(u_i)|}{|U|}\frac{C_{|S_Q(u_i)|}^2}{C_{|U|}^2}\\&=\sum_{j=1}^{n}\frac{|Q_j|}{|U|}\frac{C_{|Q_j|}^2}{C_{|U|}^2}=CG(Q);\end{aligned}$$

(3) Let $P,Q\subseteq A$, $U/SIM(P)=\{S_P(u_1),S_P(u_2),\cdots,S_P(u_{|U|})\}$, $U/SIM(Q)=S_Q(u_1),S_Q(u_2),\cdots,S_Q(u_{|U|})$ and $P\prec' Q$, then for arbitrary $S_P(u_i)(i\leq|U|)$, there exists a sequence $\{S_Q(u_1'),S_Q(u_2'),\cdots,S_Q(u_{|U|}')\}$ such that $|S_P(u_i)|<|S_Q(u_i')|(i=1,2,\cdots,|U|)$. Hence, we obtain that

$$\begin{aligned}CG(P)&=\sum_{i=1}^{m}\frac{|P_i|}{|U|}\frac{C_{|P_i|}^2}{C_{|U|}^2}=\sum_{i=1}^{|U|}\frac{|S_P(u_i)|}{|U|}\frac{C_{|S_P(u_i)|}^2}{C_{|U|}^2}\\&<\sum_{i=1}^{|U|}\frac{S_Q(u_i')}{|U|}\frac{C_{|S_Q(u_i')|}^2}{C_{|U|}^2}=\sum_{i=1}^{|U|}\frac{S_Q(u_i)}{|U|}\frac{C_{|S_Q(u_i)|}^2}{C_{|U|}^2}\\&=\sum_{j=1}^{n}\frac{|Q_j|}{|U|}\frac{C_{|Q_j|}^2}{C_{|U|}^2}=CG(Q).\end{aligned}$$

Thus, CG in definition 4 is the knowledge granulation under definition 1.

Definition 5. [15] *Let* $S=(U,A)$ *be an incomplete information system,* $U/SIM(A)=\{S_A(u_1),S_A(u_2),\cdots,S_A(u_{|U|})\}$. *Combination granulation of* A *is defined by*

$$CG(A)=\frac{1}{|U|}\sum_{i=1}^{|U|}\frac{C_{|S_A(u_i)|}^2}{C_{|U|}^2}. \tag{4}$$

Theorem 5. *CG in definition 5 is a knowledge granulation under definition 1.*

Proof. Similar to theorem 4, it can be proved.

4 Conclusions

In the present research, the concept of partial relation $\preceq'$ with set size character in information system is introduced. Furthermore, we give the axiom definition of knowledge granulation to measure uncertainty of knowledge in formation system, prove that several existing knowledge granulations are all special instances of the definition. Present conclusions appears to be well suited for measuring uncertainty of knowledge in information system.

Acknowledgements. This work was supported by the national natural science foundation of China (No. 70471003, No. 60573074, No. 60275019), the foundation of doctoral program research of the ministry of education of China (No. 20050108004), the top scholar foundation of Shanxi, China, and key project of science and technology research of the ministry of education of China.

References

1. Chakik, F.E., Shahine, A., Jaam, J. and Hasnah, A.: An approach for constructing complex discriminating surfaces based on bayesian interference of the maximum entropy. *Information Sciences*. 163 (2004) 275-291
2. Düntsch, I. and Gediga, G.: Uncertainty measures of rough set prediction. *Artificial Intelligence*. 106 (1998) 109-137
3. Liang, J.Y., Chin, K.S., Dang, C.Y. and Richard C.M.Yam.: A new method for measuring uncertainty and fuzziness in rough set theory. *International Journal of General Systems*. 31 (4) (2002) 331-342
4. Kryszkiewicz, M.: Rules in incomplete information systems. *Information systems*. 113 (1999) 271-292
5. Liang, J.Y., Xu, Z.B.: The algorithm on knowledge reduction in incomplete information systems. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. 24 (1) (2002) 95-103
6. Liang, J.Y., Shi, Z.Z., Li, D.Y. and Wierman, M.J.: The information entropy, rough entropy and knowledge granulation in incomplete information system. *International Journal of General Systems*. (to appear)
7. Qian, Y.H., Liang, J.Y.: Combination entropy and combination granulation in incomplete information system. *Lecture Notes in Artificial Intelligence*. 4062 (2006) 184-190
8. Leung, Y., Li, D.Y.: Maximal consistent block technique for rule acquisition in incomplete information systems. *Information Sciences*. 153 (2003) 85-106
9. Liang, J.Y., Shi, Z.Z.: The information entropy, rough entropy and knowledge granulation in rough set theory. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. 12 (1) (2004) 37-46
10. Zadeh, L.A.: Fuzzy sets and information granularity. In: Gupta, M. and Yager, R. (Eds): Advances in Fuzzy Set Theory and Application. North-Holland, Amsterdam. (1979) 3-18