Function Set-Valued Information Systems

Hongmei Chen, Tianrui Li, Chuan Luo, Junbo Zhang and Xinjiang Li

Abstract Set-valued Information System (SIS) aims at dealing with attributes' multi-values. It is an extension of the single-valued information system. In traditional SIS, the tolerance and dominance relations were introduced to deal with the relationship between objects. But the partial orders between the attributes values have not been taken into consideration. In this paper, a function set-valued equivalence relation is proposed first considering the applications of SIS. Then the properties of the function set-valued equivalence classes are analyzed. Furthermore, the definitions of approximations under the function SIS are presented and the properties w.r.t. approximations under the tolerance and dominance relations are investigated.

This work is supported by the National Science Foundation of China (Nos. 61175047, 61100117) and NSAF (No. U1230117), the Youth Social Science Foundation of the Chinese Education Commission (No. 11YJC630127), and the Fundamental Research Funds for the Central Universities (SWJTU11ZT08, SWJTU12CX091).

H. Chen (✉) · T. Li · C. Luo · J. Zhang · X. Li
School of Information Science and Technology, Southwest Jiaotong University, 610031 Chengdu, China
e-mail: hmchen@swjtu.edu.cn

T. Li
e-mail: trli@swjtu.edu.cn

C. Luo
e-mail: luochuan@my.swjtu.edu.cn

J. Zhang
e-mail: JunboZhang86@163.com

F. Sun et al. (eds.), *Knowledge Engineering and Management*,
Advances in Intelligent Systems and Computing 214,
DOI: 10.1007/978-3-642-37832-4_1,

1 Introduction

Granular Computing (GrC) proposed by Zadeh [1, 2] in 1989 has been successfully used in many areas, e.g., image processing, pattern recognition, and data mining. A granule is a chunk of knowledge made of different objects "drawn together by indistinguishability, similarity, proximity or functionality" [2]. Different levels of concepts or rules are induced in GrC. Rough set theory proposed by Pawlak in 1982 is considered as an important branch of GrC, which has been used to process inconsistent information [3].

In Traditional Rough Set theory (TRS), an attribute of one object in the information system can only have one value. But in real-life applications, attributes may have multi-values. For example, the degree of a person may be a combination of bachelor, master, and doctor. Set-valued systems are used to handle the case when objects' attribute have multi-values. In addition, there exists a lot of data in the information system which have uncertainty and incompleteness. Set-valued Information System (SIS) can also be used to process incomplete information. In this case, the missing value can be only one of the values in the value domain [4]. In [4], Guan defined the tolerance relation and the maximum tolerance block as well as the relative reduct based maximum tolerance block in SIS. Qian and Liang defined disjunctive and conjunctive SIS in [5] in view of different meaning of attributes' values and they studied the approach of attribute reduction and rule extraction in these two different SISs. Song and Zhang et al. defined the partly accordant reduction and the assignment reduction in an inconsistent set-valued decision information system [6, 7]. Huang et al. introduced a degree dominance relation to dominance set-valued intuitionist fuzzy decision tables and a degree dominance set-valued rough set model [8]. In [9, 10], Chen et al. introduced the probability rough sets into SIS and they studied the method for updating approximations incrementally under the variable precision set-valued ordered information while attributes' values coarsening and refining.

The equivalence relation is employed in TRS to deal with the relationship among objects and then equivalence classes induced by different equivalence relations form the elementary knowledge of the universe. Any set in the universe is described approximately by equivalence classes. Hence, a pair of certain sets, i.e., upper and lower approximations are used to describe the set approximately. The approximations partition the universe into three different regions, namely, positive, negative, and boundary regions. Then we can induce certain rules from the positive and negative regions and uncertain rules from the boundary region, i.e., a tree-way decision [11]. The SIS is a general model of single-valued information system. The equivalence relation is substituted by the tolerance relation and the dominance relation. Because the value of the missing data may be any value in the values' domain, the SIS has been used to process data in the incomplete information system. The tolerance and dominance relations in SIS consider multi-values of attributes. But the partial order among the attributes' values is not taken into consideration. In this paper, by analyzing the meaning of the tolerance and dominance relations, a new relationship in SIS is proposed, i.e., a function set-valued equivalence relation. Then the properties of the function set-valued equivalence relation are discussed. Furthermore, a more general relationship, a general function relation, is investigated and its properties are analyzed.

The paper is organized as follows. In Sect. 2, basic concepts in SIS are reviewed. In Sect. 3, a function SIS is proposed. A function equivalence relation and its properties are discussed. Then the general function relation is investigated. The paper ends with conclusions and further directions in Sect. 4.

2 Preliminary

In this section, Some basic concepts in SIS are briefly reviewed [4–6].

Definition 21 Let $S = (U, A, V, f)$ be a SIS, where $U = \{x_1, x_2, \ldots, x_n\}$ is a non-empty finite set of objects, called the universe. $A = \{a_1, a_2, \ldots, a_l\}$ is a non-empty finite set of attributes. The element in A is called an attribute. $A = C \cup D$, $C \cap D = \emptyset$, C is the set of condition attributes and D is the set of decision attributes. $V = \{V_{a_1}, V_{a_2}, \ldots, V_{a_l}, a_i \in A\}$, $V_{a_i}(i = 1, 2, \ldots, l)$ is the domain of attribute $a_i(a_i \in A)$. V_C is the domain of condition attributes. V_D is the domain of decision attributes. $V = V_C \cup V_D$ is the domain of all attributes. $f : U \times C \to 2^{V_C}$ is a set-valued mapping.$f : U \times D \to V_D$ is a single-valued mapping. $f(x_i, a_l)$ is the value of x_i on attribute a_l. An example of SIS is shown in Table 1.

The tolerance and dominance relation have been proposed to deal with the different applications in SIS. For example, if two persons have the tolerance relation on the attribute “language”, it means they can communicate through the common language. The tolerance relation is defined as follows.

Definition 22 Let $S = (U, A, V, f)$ be an SIS, for $B \subseteq C$, the tolerance relation in the SIS is defined as follows:

$$R_B^{\cap} = \{(y, x) \in U \times U | f(y, a) \cap f(x, a) \neq \emptyset (\forall a \in B)\} \tag{1}$$

Table 1 A set-valued information system

U	Price	Mileage	Size	Max-speed
1	{High, Medium}	High	Full	{High, Medium}
2	Low	{High, Medium}	Full	Medium
3	Medium	Low	Full	Medium
4	{Low, Medium}	{Low, Medium}	Compact	Low
5	High	{High, Medium}	Compact	{Low, Medium}
6	Medium	{Low, Medium}	Full	High
7	Low	High	Compact	High
8	High	Low	Compact	{Low, Medium}

The tolerance relation is reflexive and symmetric, but not transitive. Let $[x]_B^{\cap}$ denote the tolerance class of object x, where $[x]_B^{\cap} = \{y \in U \big| (y,x) \in R_B^{\cap}\} = \{y \in U | f(x,a_i) \cap f(y,a_i) \neq \emptyset, a_i \in B\}$.

Definition 23 Let $S = (U,A,V,f)$ be an SIS, $\forall X \subseteq U$, the approximations of X under the tolerance relation $R_C^{\cap}$ are defined as follows, respectively:

$$\underline{apr_B^T}(X) = \{x \in U \big| [x_i]_C^{\cap} \subseteq X\} \tag{2}$$

$$\overline{apr_B^T}(X) = \{x \in U \big| [x_i]_C^{\cap} \cap X \neq \emptyset\} \tag{3}$$

Definition 24 Given an SIS (U,A,V,f), for $B \supseteq C$, a dominance relation in the SIS is defined as follows:

$$\begin{aligned} R_B^{\subseteq} &= \{(y,x) \in U \times U | f(y,a) \supseteq f(x,a)(\forall a \in B)\} \\ &= \{(y,x) \in U \times U | y \succeq x\} \end{aligned} \tag{4}$$

$R_B^{\subseteq}$ is reflexive, dissymmetric, and transitive.

We denote $[x]_B^{\subseteq} = \{y \in U \big| (y,x) \in R_B^{\geq}\}$, $[x]_B^{\subseteq} = \{y \in U \big| (x,y) \in R_B^{\subseteq}\}$. $[x]_B^{\subseteq}$ $(x \in U)$ are granules of knowledge induced by the dominance relation, which are the set of objects dominating x. The dominance relation in SIS means in some degree, the more values in an attribute of one object the more the probability that it dominates the other objects. It considers the cardinality of the attributes' values.

Definition 25 Let $S = (U,A,V,f)$ be an SIS, $\forall X \subseteq U$, the approximations of X under the dominance relation $R_C^{\supseteq}$ are defined as follows, respectively:

$$\underline{apr_B^D}(X) = \{x \in U \big| [x_i]_C^{\supseteq} \subseteq X\} \tag{5}$$

$$\overline{apr_B^D}(X) = \{x \in U \big| [x_i]_C^{\supseteq} \cap X \neq \emptyset\} \tag{6}$$

3 Function Set-Valued Information Systems

Multi-values of attributes in SIS give more information than single-valued information system. In real-life applications, a partial order may exist in the attribute values, i.e., for $V_{a_i} = \{v_{a_1}^1, v_{a_1}^2, \cdots, v_{a_1}^k\}$ $(k = |V_{a_i}|)$, $v_{a_1}^1 \preceq v_{a_1}^2 \preceq \cdots \preceq v_{a_1}^k$ is the partial order among the attribute values on the attribute a_i. In Definitions 22 and 24, the multi-value property of attributes is taken into consideration but the partial order in the attribute values is neglected. Then, we present a new definition, namely, a function set-valued equivalence relation.

Definition 31 Let $S = (U, A, V, f)$ be a SIS, $V_{a_i} = \{v_{a_1}^1, v_{a_1}^2, \ldots, v_{a_1}^k\}(k = |V_{a_i}|)$ is the attribute value domain of attribute a_i, where $v_{a_1}^i(1 \leq i \leq k)$ $s.t.$ the $v_{a_1}^1 \preceq v_{a_1}^2 \preceq \ldots \preceq v_{a_1}^k$. Then an attribute function F_{a_i} is defined as $F_{a_i} : V_{a_i} \rightarrow P(V_{a_i})$, which is a mapping from V_{a_i} to the power set of V_{a_i}.

The definition of F_{a_i} is given by the requirement of real-life applications. Usually, F_{a_i} may be maximum, minimum, average of the set-valued. For example, the degree a person awarded may be bachelor, master, doctor; the employer in universities may be more interesting in the person with the highest degree. Let $F(f(x, a_i))$ denote the value of the function F_{a_i} on the attribute value $f(x_i, a_l)$. Then the function equivalence relation is defined as follows:

Definition 32 Let $S = (U, A, V, f)$ be an SIS, the function equivalence relation is defined as follows:

$$R_B^F = \{(y, x) \in U \times U | F(f(x, a_i)) = F(f(y, a_i)), a_i \in B\} \tag{7}$$

Let $[x]_B^F$ denote the function equivalence class of object x, where $[x]_B^F = \{y \in U \big| (y, x) \in R_B^F\}$.

Property 31 For R_B^F, the following hold:

1. Reflexive: $\forall x \in U$, $xR_B^F x$;
2. Symmetric: $\forall x, y \in U$, if $xR_B^F y$, then $yR_B^F x$;
3. Transitive: $\forall x, y, z \in U$, if $xR_B^F y$ and $yR_B^F z$, then $xR_B^F z$;

$\bigcup_{i=1}^{|U|} [x_i]_B^F = U$ and $[x_i]_B^F \cap [x_j]_B^F = \emptyset$, i.e., U/R_B^F forms a partition of the university.

For any $B \subseteq C$, $[x_i]_C^F \subseteq [x_i]_B^F$. The tolerance, dominance, and function equivalence relation meet the different requirements of real-life applications. In the following, we compare these three different relations defined in SIS. Let $[x_i]_B^*$ denote the classes induced by different relations in a SIS and $\widetilde{I^*} = \{[x_i]_B^* \big| x_i \in U\}$ denote a classes cluster set, where the superscript * may be replaced by $\cap$, $\supseteq$, F and $\widetilde{I^\cap}$, $\widetilde{I^\supseteq}$, $\widetilde{I^F}$ denote tolerance, dominance, and function equivalence classes, respectively. Liang et al. defined the granularity of knowledge under the tolerance relation in incomplete information system [12]. Similarly, the granularity and the rough entropy of B in an SIS is defined as follows:

Definition 33 Let $S = (U, A, V, f)$ be a SIS, $[x_i]_B^*(x_i \in U, i = 1, 2, \ldots, |U|)$ is the granule induced by different relation in a set-valued information system, $\forall B \subseteq C \subseteq A$, where the superscript $*$ may be substituted by $\cap$, $\supseteq$, F and the granularity of B is defined as follows:

$$GK^*(B) = -\sum_{i=1}^{|U|} \frac{1}{|U|} \log_2 \frac{\left|[x_i]_B^*\right|}{|U|} \tag{8}$$

Definition 34 Let $S = (U, A, V, f)$ be an SIS, $[x_i]_B^*(x_i \in U, i = 1, 2, \ldots, |U|)$ is granule induced by different relation in a set-valued information system, $\forall P \subseteq C \subseteq A$. Then the rough entropy of P is defined as follows:

$$E_r^*(P) = -\sum_{i=1}^{|U|} \frac{1}{|U|} \log_2 \frac{1}{\left|[x_i]_B^*\right|} \tag{9}$$

For convenience, we define a partial order binary relation $\preceq$ first.

$I^* \preceq I^{\Delta} \Leftrightarrow$ *if* $\forall x_i \in U, \exists [x_i]_B^* \subseteq [x_i]_B^{\Delta}$, where $[x_i]_B^* \in \widetilde{I^*}$ and $[x_i]_B^{\Delta} \in \widetilde{I^{\Delta}}$.

Note: The superscript Δ may also be replaced by $\cap$, $\supseteq$, F. Then the following properties hold:

Property 32 For $[x_i]_B^*$ and $\widetilde{I^*}$, the following properties hold:

1. For any $B_1 \subset B_2 \subseteq C$, $[x_i]_{B_2}^F \subseteq [x_i]_{B_1}^F \subseteq [x_i]_C^F$;
2. For any $B \subseteq C$, $[x_i]_B^F \subseteq [x_i]_B^{\cap}$;
3. For any $B \subseteq C$, $I^F \preceq \widetilde{I^{\cap}}$;
4. For any $B \subseteq C$, $GK^{\supseteq}(B) \leq GK^{\cap}(B)$;
5. For any $B \subseteq C$, $E_r^{\cap}(P) \leq E_r^{\supseteq}(P)$.

Proof Proof 1. $\forall x_j \in [x_i]_{B_2}^F$, $x_j R_{B_2}^F x_i$, i.e., $F(f(x_i, a_i)) = F(f(x_j, a_i)), \forall a_i \in B_2$. Since $B_1 \subset B_2$, then $x_j \in [x_i]_{B_1}^F$. On the other hand, if $\forall x_j \in [x_i]_{B_1}^F$, $x_j R_{B_1}^F x_i$, i.e., $F(f(x_i, a_i)) = F(f(x_j, a_i))$, $\forall a_i \in B_1$. Because $B_1 \subset B_2$, then if $\exists F(f(x_i, a_i)) \neq F(f(x_j, a_i))$, $\forall a_i \in B_2 - B_1$. Then $x_j \not R_{B_2}^F x_i$, i.e., $x_j \notin [x_i]_{B_1}^F$. Therefore, we have $[x_i]_{B_2}^F \subseteq [x_i]_{B_1}^F$. Consequently, for any $B_1 \subset B_2 \subseteq C$, $[x_i]_C^F \subseteq [x_i]_{B_2}^F \subseteq [x_i]_{B_1}^F$. 2. $\forall x_j \in [x_i]_C^F$, $F(f(x_i, a_i)) = F(f(x_j, a_i)), a_i \in B$. Because $F_{a_i} : V_{a_i} \rightarrow P(V_{a_i})$, then $f(x_i, a_i) \cap f(x_j, a_i) \neq \emptyset$, i.e., $\forall x_j \in [x_i]_C^{\cap}$. $\forall x_j \in [x_i]_C^{\cap}$, we have $f(x_i, a_i) \cap f(x_j, a_i) \neq \emptyset$, if $\exists F(f(x_i, a_i)) \neq F(f(x_j, a_i)), a_i \in B$, $x_j \notin [x_i]_C^F$. Therefore, $[x_i]_B^F \subseteq [x_i]_B^{\cap}$. Consequently, for any $B \subseteq C$, $[x_i]_B^F \subseteq [x_i]_B^{\cap}$. Since 2 is always true, then 3, 4, and 5 hold.

The tolerance, dominance, and function equivalence relations are used to deal with different cases in a SIS. In most cases, three kinds of relations may coexist on different attributes. In the following, a general relation is given.

Definition 35 Let $S = (U, A, V, f)$ be a SIS, where $B_T \subseteq C$, $B_D \subseteq C$, $B_F \subseteq C$, $B_T \cap B_D \cap B_F = \emptyset$, $B_T \cup B_D \cup B_F = B \subseteq C$, the general function relation is defined as follows:

$$\begin{aligned} R_B^G = \{(y,x) \in U \times U | f(x, a_t) \cap f(y, a_t) \neq \emptyset \wedge f(x, a_d) \subseteq f(y, a_d) \\ \wedge F(f(x, a_f)) = F(f(y, a_f)), a_t \in B_T, a_d \in B_D, a_f \in B_F\} \end{aligned} \quad (10)$$

Let $[x]_C^G$ denote the function equivalence class of object x, where $[x]_B^G = \{y \in U \big| (y,x) \in R_B^G\}$.

Property 33 For R_B^G,

1. Reflexive: $\forall x \in U$, $xR_B^G x$;
2. Dissymmetric: $\forall x, y \in U$, if $xR_B^F y$, then $\overline{yR_B^G x}$;
3. Non-transitive.

$\bigcup_{i=1}^{|U|} [x_i]_B^G = U$and $[x_i]_B^G \cap [x_j]_B^G = \emptyset$, i.e., U/R_B^G forms a covering of the university.

Then, the approximations based on the general function relation are defined as follows:

Definition 36 Let $S = (U, A, V, f)$ be a SIS, $\forall X \subseteq U$, the lower and the upper approximations of X under the general function relation R_C^G are defined as follows, respectively:

$$\underline{apr_B^G}(X) = \{x \in U \big| [x_i]_C^G \subseteq X\} \quad (11)$$

$$\overline{apr_B^G}(X) = \{x \in U \big| [x_i]_C^G \cap X \neq \emptyset\} \quad (12)$$

Based on approximations of X, one can partition the universe U into three disjoint regions, i.e., the *positive region* $POS_B^G(X)$, the *boundary region* $BNR_B^G(X)$, and the *negative region* $NEG_B^G(X)$:

$$POS^G(X) = \underline{apr_B^G}(X); \quad (13)$$

$$BNR^G(X) = \overline{apr_B^G}(X) - \underline{apr_B^G}(X); \quad (14)$$

$$NEG^G(X) = U - \overline{apr_B^G}(X). \quad (15)$$

We can extract certain rules from the position and negative regions and possible rules from the boundary region too.

Property 34 For $\underline{apr}_B^G(X)$ and $\overline{apr}_B^G(X)$, the following hold:

$$\underline{apr}_B^T(X) \subseteq \underline{apr}_B^G(X);$$

$$\overline{apr}_B^D(X) \subseteq \overline{apr}_B^G(X) \subseteq \overline{apr}_B^T(X);$$

$$If\ A \subseteq B \subseteq C, then\ \underline{apr}_A^G(X) \subseteq \underline{apr}_B^G(X);$$

$$If\ A \subseteq B \subseteq C, then\ \overline{apr}_B^G(X \subseteq \overline{apr}_A^G(X);$$

Proof (1) From Property 32, for any $B \subseteq C$, $[x_i]_B^F \subseteq [x_i]_B^\cap$, then $[x_i]_B^G \subseteq [x_i]_B^\cap$. Therefore, we have $\underline{apr}_B^T(X) \subseteq \underline{apr}_B^G(X)$.

4 Conclusions

In this paper, the partial order among the multi-values in an SIS is taken into consideration. The function equivalence relation is proposed. Then the properties of function equivalence classes are analyzed. Furthermore, a general function relation is defined in the SIS. Then, approximations and its properties under the general function relation are discussed. In the future work, we will study the attribute reduction under the general function relation in the SIS.

References

1. Zadeh LA (1997) Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets Syst 19(1):111–127
2. Zadeh LA (2008) Is there a need for fuzzy logic? Inf Sci 178:2751–2779
3. Pawlak Z, Skowron A (2007) Rudiments of rough sets. Inf Sci 177:3–27
4. Guan YY, Wang HK (2006) Set-valued information systems. Inf Sci 176(17):2507–2525
5. Qian YH, Dang CY, Liang JY, Tang DW (2009) Set-valued ordered information systems. Inf Sci 179(16):2809–2832
6. Song XX, Zhang WX (2009) Knowledge reduction in inconsistent set-valued decision information system. Comput Eng Appl 45(1):33–35 (In chinese)
7. Song XX, Zhang WX (2006) Knowledge reduction in set-valued decision information system. In: Proceedings of 5th international conference on rough sets and current trends in computing (RSCTC 2006), 4259 LNAI, pp 348–357
8. Huang B, Li HX, Wei DK (2011) Degree dominance set-valued relation-based RSM in intuitionistic fuzzy decision tables. In: Proceedings of the international conference on internet computing and information services, pp 37–40
9. Chen HM, Li TR, Zhang JB (2010) A method for incremental updating approximations based on variable precision set-valued ordered information systems. In: Proceedings of the 2010 IEEE international conference on granular computing (GrC2010), pp 96–101

10. Zou WL, Li TR, Chen HM, Ji XL (2009) Approaches for incrementally updating approximations based on set-valued information systems whileattribute values' coarsening and refining. In: Proceedings of the 2009 IEEE international conference on granular computing (GrC2009), pp 824–829
11. Yao YY (2010) Three-way decisions with probabilistic rough sets. Inf Sci 180(7):341–353
12. Liang JY, Shi ZZ (2004) The information entropy, rough entropy and knowledge granulation in rough set theory. Int J Uncertainty Fuzziness Knowl Based Syst 12(1):37–46