

# A Measure Method for Indiscernibility in Imperfect Information System

Jiye Liang Jifang Pang<sup>1</sup>

<sup>1</sup>Key Laboratory of Ministry of Education for Computation Intelligence and Chinese Information Processing  
School of Computer and Information Technology  
Shanxi University, Taiyuan 030006, China

## Abstract

Traditionally, the information system is assumed to be perfect, i.e. attribute values are not missing and supposed to be precise. In fact, imperfect information system is always existent. In this paper, based on imperfect information system (include missing data and imprecise data), the concepts of indiscernibility and discernibility are defined, their important properties are given, and the relationship between those concepts is established. These results will be helpful for measuring the indiscernibility of knowledge, and have instructive significance for studying for knowledge acquisition in imperfect information system.

**Keywords:** imperfect information system, tolerance relation, maximal consistent class, indiscernibility

## 1. Introduction

The entropy of a system as defined by Shannon gives a measure of uncertainty about its actual structure in [1]. In [2, 3], several authors have used Shannon's entropy and its variants to measure uncertainty in rough set theory. A new definition for information entropy in rough set theory is presented in [4]. Especially, M.J. Wierman presents a well justified measure of uncertainty, measure of granularity, along with an axiomatic derivation in [5]. Furthermore, the relationships among information entropy, rough entropy and knowledge granulation in complete information systems are established in [6].

However, imperfect information system is always existent. In [7], a generalization of the rough sets approach which deals with missing and imprecise data is presented. In [8], rough set and rough classification of imperfect information system are investigated. But researches on the measure of indiscernibility based on imperfect information system are relatively less.

In this paper, a measure method for indiscernibility in imperfect information system is proposed, some important properties are given, and the strict complementary relationship between indiscernibility

and discernibility is established. These results have instructive significance for studying for the measure of indiscernibility in imperfect information system.

## 2. Imperfect Information system

Let  $S = (U, A)$  be an information system, where

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

Each subset of attributes  $P \subseteq A$  determines a binary indiscernibility relation  $IND(P)$  on  $U$ ,

$$IND(P) = \{(x, y) \in U \times U \mid \forall a \in P, f_a(x) = f_a(y)\}.$$

The relation  $IND(P)$ ,  $P \subseteq A$ , is an equivalence relation and constructs a partition of  $U$ , which is denoted by  $U / IND(P)$ .

If  $V_a$  contains a null value for at least one attribute  $a \in A$ , then  $S$  is called an incomplete information system<sup>[9,10]</sup>.

By an incomplete information system we mean a system with missing data (null values).

If we change the mapping  $f_a$  to  $f_a : U \rightarrow (2^{V_a} - \emptyset)$ , where  $2^{V_a}$  is the power set of  $V_a$ ,  $\emptyset$  is empty set, then  $S$  is called an imperfect information system<sup>[8]</sup>.

By an imperfect information system we mean a system with missing data (null values) or imprecise data. Generally, a null value of attribute  $a$  may be any value in  $V_a$ , which is denoted by  $*$ .

Obviously, incomplete information system is a special instance of imperfect information system.

Let  $P \subseteq A$ , a tolerance relation (satisfies reflexivity and symmetry) is defined as follows<sup>[8]</sup>,

$$SIM(P) = \{(x, y) \in U \times U \mid \forall a \in P, f_a(x) \cap f_a(y) \neq \emptyset\}.$$

It is easily shown that  $SIM(P) = \bigcap_{a \in P} SIM(\{a\})$ .

Let  $U / SIM(P)$  denote the set of maximal consistent classes in terms of tolerance relation  $SIM(P)$ . By the maximal consistent class  $X \subseteq U$  we mean that any pair of objects in  $X$  are in  $SIM(P)$ , but if get an object from  $U - X$  to  $X$ , then at least two objects in  $X$  are not in  $SIM(P)$ . The maximal consistent classes in  $U / SIM(P)$  generally do not constitute a partition of  $U$  but a covering of  $U$ .

**Example 2.1.** Here, we employ an example in [8] to illustrate the tolerance relation. Table 2.1 is an imperfect information system, where

$$\begin{aligned} U &= \{x_1, x_2, x_3, x_4, x_5, x_6\}, A = \{a, b, c, d\}, \\ V_a &= \{0, 1, 2\}, V_b = \{0, 1\}, V_c = \{0, 1\}, \\ V_d &= \{0, 1\}, f_a(x_4) = \{1, 2\}, \text{ and } f_d(x_3) = * = \{0, 1\}. \end{aligned}$$

$U$	$A$			
	$a$	$b$	$c$	$d$
$x_1$	0	1	0	1
$x_2$	1	0	1	0
$x_3$	0	1	0	*
$x_4$	1, 2	1	1	0
$x_5$	1	*	1	0
$x_6$	0, 2	0	0	1

Table 2.1: An imperfect information system [8].

By computing, we have that

$$U / SIM(A) = \{\{x_1, x_3\}, \{x_2, x_5\}, \{x_4, x_5\}, \{x_6\}\}.$$

### 3. Imperfect Information System

Let  $S = (U, A)$  be an imperfect information system,  $P \subseteq A$ . Based on the tolerance relation defined in [8], we denote the set of all maximal consistent classes of  $P$  which includes some object  $x$  as  $S_x(P)$ , i.e.

$$S_x(P) = \{X \in U / SIM(P) \mid x \in X\}.$$

**Definition 3.1.** Let  $S = (U, A)$  be an imperfect information system,  $P \subseteq A$ ,  $X \in U / SIM(P)$  and  $x, y \in X$ . The indiscernibility between  $x$  and  $y$  with respect to  $P$  is defined by

$$\begin{aligned} \rho_P(x, y) &= \frac{\prod_{a \in P} |f_a(x) \cap f_a(y)|}{\prod_{a \in P} |f_a(x)|} \cdot \frac{\prod_{a \in P} |f_a(x) \cap f_a(y)|}{\prod_{a \in P} |f_a(y)|} \\ &= \prod_{a \in P} \frac{|f_a(x) \cap f_a(y)|^2}{|f_a(x)| |f_a(y)|}. \end{aligned}$$

Let  $S = (U, A)$  be an imperfect information system,  $P \subseteq A$ ,  $S_x^*(P) = \bigcup \{X \in S_x(P)\}$ . Then the sum of the indiscernibility between  $x$  and all objects included in  $S_x^*(P)$  with respect to  $P$  is denoted by

$$\rho_x(S_x^*(P)) = \sum_{y \in S_x^*(P)} \rho_P(x, y).$$

We can easily deduce the following properties.

**Property 3.1.** Let  $S = (U, A)$  be an imperfect information system,  $P \subseteq A$ ,  $X \in U / SIM(P)$ . Then for  $\forall x, y \in X$ , we have that  $0 < \rho_P(x, y) \leq 1$ .

**Property 3.2.** Let  $S = (U, A)$  be an imperfect information system,  $P \subseteq A$ ,  $X \in U / SIM(P)$ . Then for  $\forall x, y \in X$ , we have that  $\rho_P(x, y) = \rho_P(y, x)$ .

**Property 3.3.** Let  $S = (U, A)$  be an imperfect information system,  $P \subseteq A$ . Then for  $\forall x \in U$ , we have that  $0 < \rho_x(S_x^*(P)) \leq |S_x^*(P)|$ .

**Theorem 3.1.** Let  $S = (U, A)$  be an imperfect information system,  $P, Q \subseteq A$  and  $Q \subseteq P$ . Then for  $\forall x \in U$ , we have that  $\rho_x(S_x^*(P)) \leq \rho_x(S_x^*(Q))$ .

### 3.1. The Indiscernibility of Information System

In this section we will introduce the indiscernibility of the imperfect information system  $S$  based on definition 3.1.

**Definition 3.2.** Let  $S = (U, A)$  be an imperfect information system,  $U = \{x_1, x_2, \dots, x_{|U|}\}$ . Then the indiscernibility of system  $S$  with respect to  $A$  is defined as

$$I(A) = \frac{1}{|U|^2} \sum_{i=1}^{|U|} \rho_{x_i}(S_{x_i}^*(A)).$$

For an imperfect information system  $S = (U, A)$ , we have that  $1/|U| \leq I(A) \leq 1$ .

**Theorem 3.2.** Let  $S = (U, A)$  be a complete information system,  $U = \{x_1, x_2, \dots, x_{|U|}\}$ ,  $U / IND(A) = \{Z_1, Z_2, \dots, Z_m\}$ . Then the indiscernibility of system  $S$  with respect to  $A$  degenerates into

$$G(A) = \frac{1}{|U|^2} \sum_{i=1}^m |Z_i|^2,$$

where  $G(A) = \frac{1}{|U|^2} \sum_{i=1}^m |Z_i|^2$  is the knowledge granulation of knowledge  $A$  defined in [6].

Theorem 3.2 states that the knowledge granulation in complete information system is a special instance of the indiscernibility of imperfect information system.

**Theorem 3.3.** Let  $S = (U, A)$  be an imperfect information system,  $U = \{x_1, x_2, \dots, x_{|U|}\}$ ,  $P, Q \subseteq A$ . If  $Q \subseteq P$ , then we have that  $I(P) \leq I(Q)$ .

Theorem 3.3 states that the indiscernibility of system  $S$  with respect to  $Q$  decreases by adding more attributes to the attribute set  $Q$ .

### 3.2. The Discernibility of Information System

In this section we give a measure of discernibility in the imperfect information system.

**Definition 3.3.** Let  $S = (U, A)$  be an imperfect information system,  $U = \{x_1, x_2, \dots, x_{|U|}\}$ . Then the discernibility of system  $S$  with respect to  $A$  is defined as

$$\begin{aligned} D(A) &= \sum_{i=1}^{|U|} \frac{1}{|U|} \left[ 1 - \frac{\rho_{x_i}(S_{x_i}^*(A))}{|U|} \right] \\ &= \frac{1}{|U|^2} \sum_{i=1}^{|U|} [|U| - \rho_{x_i}(S_{x_i}^*(A))], \end{aligned}$$

where  $|U| - \rho_{x_i}(S_{x_i}^*(A))$  represents the sum of the discernibility between  $x_i$  and all objects within the universe  $U$  with respect to  $A$ .

For an imperfect information system  $S = (U, A)$ , we have that  $0 \leq D(A) \leq 1 - 1/|U|$ .

**Theorem 3.4.** Let  $S = (U, A)$  be a complete information system,  $U = \{x_1, x_2, \dots, x_{|U|}\}$ ,  $U / IND(A) = \{Z_1, Z_2, \dots, Z_m\}$ . Then the discernibility of system  $S$  with respect to  $A$  degenerates into

$$E(A) = \sum_{i=1}^m \frac{|Z_i|}{|U|} \left( 1 - \frac{|Z_i|}{|U|} \right),$$

where  $E(A) = \sum_{i=1}^m \frac{|Z_i|}{|U|} \left( 1 - \frac{|Z_i|}{|U|} \right)$  is the information entropy of knowledge  $A$  defined in [6].

Theorem 3.4 states that the information entropy in complete information system is a special instance of the discernibility of imperfect information system.

**Theorem 3.5.** Let  $S = (U, A)$  be an imperfect information system,  $U = \{x_1, x_2, \dots, x_{|U|}\}$ ,  $P, Q \subseteq A$ . If  $Q \subseteq P$ , then we have that  $D(Q) \leq D(P)$ .

Theorem 3.5 states that the discernibility of system  $S$  with respect to  $Q$  increases by adding more attributes to the attribute set  $Q$ .

### 3.3. Relationship Between the Indiscernibility and the Discernibility of Information System

In this section the relationship between the indiscernibility and the discernibility of system  $S$  is established, which is proved to be a strict complementary relationship.

**Theorem 3.6.** For an imperfect information system  $S = (U, A)$ , we have that

$$I(A) + D(A) = 1.$$

**Proof.** Let  $S = (U, A)$  be an imperfect information system,  $U = \{x_1, x_2, \dots, x_{|U|}\}$ . By the definition 3.2 and 3.3, we have that

$$\begin{aligned} D(A) &= \sum_{i=1}^{|U|} \frac{1}{|U|} \left[ 1 - \frac{\rho_{x_i}(S_{x_i}^*(A))}{|U|} \right] \\ &= \sum_{i=1}^{|U|} \frac{1}{|U|} - \sum_{i=1}^{|U|} \frac{\rho_{x_i}(S_{x_i}^*(A))}{|U|^2} \\ &= 1 - I(A). \end{aligned}$$

It follows that  $I(A) + D(A) = 1$ . This completes the proof.

**Example 3.1.** Continued from example 2.1, by computing, we have that

$$\rho_{x_1}(S_{x_1}^*(A)) = \rho_{x_1}(\{x_1, x_3\}) = \frac{3}{2},$$

$$\rho_{x_2}(S_{x_2}^*(A)) = \rho_{x_2}(\{x_2, x_5\}) = \frac{3}{2},$$

$$\rho_{x_3}(S_{x_3}^*(A)) = \rho_{x_3}(\{x_1, x_3\}) = \frac{3}{2},$$

$$\rho_{x_4}(S_{x_4}^*(A)) = \rho_{x_4}(\{x_4, x_5\}) = \frac{5}{4},$$

$$\rho_{x_5}(S_{x_5}^*(A)) = \rho_{x_5}(\{x_2, x_4, x_5\}) = \frac{7}{4},$$

$$\rho_{x_6}(S_{x_6}^*(A)) = \rho_{x_6}(\{x_6\}) = 1.$$

$$\begin{aligned} I(A) &= \frac{1}{|U|^2} \sum_{i=1}^{|U|} \rho_{x_i}(S_{x_i}^*(A)) \\ &= \frac{1}{36} \left( \frac{3}{2} + \frac{3}{2} + \frac{3}{2} + \frac{5}{4} + \frac{7}{4} + 1 \right) \\ &= \frac{17}{72}. \end{aligned}$$

$$\begin{aligned} D(A) &= \sum_{i=1}^{|U|} \frac{1}{|U|} \left[ 1 - \frac{\rho_{x_i}(S_{x_i}^*(A))}{|U|} \right] \\ &= \frac{1}{6} \left[ \left( 1 - \frac{3}{12} \right) + \left( 1 - \frac{3}{12} \right) + \left( 1 - \frac{3}{12} \right) \right. \\ &\quad \left. + \left( 1 - \frac{5}{24} \right) + \left( 1 - \frac{7}{24} \right) + \left( 1 - \frac{1}{6} \right) \right] \\ &= \frac{55}{72}. \end{aligned}$$

It is clear that  $I(A) + D(A) = 1$ .

## 4. Conclusions

In this paper, the concepts of the indiscernibility and the discernibility in imperfect information system are proposed, their important properties are given, and the relationship between the indiscernibility and the discernibility of information system is established, it can be expressed as  $I(A) + D(A) = 1$ . These results provide new study view for the measure of the indiscernibility of knowledge, and can be further used in many fields such as knowledge discovery and data mining etc.

## Acknowledgements

This work was supported by the national natural science foundation of China (No.70471003, 60275019), the doctor authorization foundation of the ministry of education (No.20050108004), the top

scholar foundation of Shanxi, China, and key project of science and technology research of the ministry of education of china.

## 5. References

- [1] C.E. Shannon, "The Mathematical Theory of Communication," *The Bell System Technical Journal*, 27 (3 and 4): pp. 373-423, 1948.
- [2] T. Beaubouef, F.E. Petry and G. Arora, "Information-Theoretic Measures of Uncertainty for Rough Sets and Rough Relational Databases," *Information Sciences*, 109: pp. 535-563, 1998.
- [3] I. Duntsch, G. Gediga, "Uncertainty Measures of Rough Set Prediction," *Artificial Intelligence*, 106: pp. 109-137, 1998.
- [4] J.Y. Liang, K.S. Chin, C.Y. Dang and C.M.YAM. Richard, "A New Method for Measuring Uncertainty and Fuzziness in Rough Set Theory," *International Journal of General Systems*, 31 (4): pp. 331-342, 2002.
- [5] M.J. Wierman, "Measuring Uncertainty in Rough Set Theory," *International Journal of General Systems*, 28 (4): pp. 283-297, 1999.
- [6] J.Y. Liang, Z.Z. Shi, "The Information Entropy, Rough Entropy and Knowledge Granulation in Rough Set Theory," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 12 (1): pp. 37-46, 2004.
- [7] R. Slowinski, J. Stefanowski, "Rough Classification in Incomplete Information Systems," *Mathematical Computing Modelling*, 12 (10/11): pp. 1347-1357, 1989.
- [8] J. Wang, J.Y. Liang, "Rough Set and Rough Classification Based on Imperfect Information Systems," *Proceedings of the 4th world congress on intelligent control and automation*, pp. 437-440, 2002.
- [9] M. Kryszkiewicz, "Rough Set Approach to Incomplete Information Systems," *Information Sciences*, 112: pp. 39-49, 1998.
- [10] M. Kryszkiewicz, "Rules in Incomplete Information Systems," *Information Sciences*, 113: pp. 271-292, 1999.