Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 13 (17): 3522-3527, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.3522.3527

Uncertainty Measures in Interval Ordered Information Systems

[1]Lv Jinlai, [1]Zhang Hui, [2]Fan Wenlei and [3]Du Xiaoping
[1]College of Computer Science and Technology, Taiyuan University of Technology,
[2]Center of Computer Technology, Taiyuan Audit Bureau,
[3]Shanxi Province Economic Information Center, Taiyuan, 030024, People's Republic of China

Abstract: Interval information systems are generalized models of single-valued information systems which is an important formal framework for the development of data mining. In this study, in terms of introducing dominance relations, a rough set approach in interval ordered information systems is first established. Then, the concept of dominance entropy and dominance combination entropy in interval ordered information systems is come up with. Finally, through calculating of dominance entropy and dominance combination entropy in interval information systems based on dominance relations, it is proved that in the wake of enhancement of knowledge discernment, dominance entropy and dominance combination entropy increases monotonously. These results give a kind of feasible approaches to discover and acquisition of knowledge in interval ordered information systems.

Key words: Interval information systems, dominance relations, entropy, measure

INTRODUCTION

Intelligent decision making and data mining are very important research issues in management science field (Cheng *et al*., 2006; Shi *et al*., 2002). Rough set theory, proposed by Pawlak and Skowron (2007), has been used by a tool which analyzes and calculates various types of data. In many real situations, the ordering of properties of the considered attributes plays a crucial role. However, the classical rough set theory does not take into account attributes with preference-ordered domains. In view of these considerations Greco *et al*. (2005) put forward an extension of rough set theory which called the Dominance-based Rough Set Approach (dRSA) (Greco *et al*., 1998, 2001, 2002). In recent years, the DRSA (Dembczynski *et al*., 2003; Shao and Zhang, 2005; Sai *et al*., 2001; Facchinetti *et al*., 1998) has lain in research. Interval-valued information system is an important type of data tables for the development of data mining. Most of them are basis of the concept of a possible degree between two interval numbers (Xu and Gu, 2002; Xu and Da, 2002).

Since Shannon put forward the concepts of information entropy in 1940s, they have been widely applied to characterize the information content in many diverse fields. Many authors (Liang and Qu, 2001; Liang and Xu, 2002; Slezak, 2002, 2005; Malyszko and Stepaniuk, 2008, 2010a, b; Hu *et al*., 2008, 2007; Yao, 2007; Yao and Zhao, 2008; Wei *et al*., 2010; Zhang *et al*., 2010; Liang *et al*., 2002; Yang *et al*., 2011) have made use of Shannon method to measure uncertainty in rough set theory. A new information entropy was put forward by Liang and Shi (2004) and Mi *et al*. (2005) gave a new fuzzy entropy and applied it for measuring the fuzziness of a fuzzy rough set based partition. A new combination entropy and combination granulation Qian and Liang (2006) was introduced to measure the uncertainty of an incomplete information system and the relationship between these two concepts was established. Up till now, however, how to measure Uncertainty by concept of entropy in interval information systems based on a dominance relation has not been reported.

In this study, interval ordered information systems is focused on. The main objective of this study is to put forward the concept of entropy in interval ordered information systems and investigate the problems of dominance entropy and dominance combination entropy in interval information systems based on dominance relations. One can learn that with enhancement of knowledge discernment, dominance entropy and dominance combination entropy in interval ordered information systems increases monotonously.

MATERIAL AND METHODS

Select an interval information system as research material: The concept of dominance entropy and

Corresponding Author: Lv Jinlai, College of Computer Science and Technology, Taiyuan University of Technology, Taiyuan, 030024, People's Republic of China

dominance combination entropy in interval ordered information systems is put forward. The numerical result of both is respectively calculated. More details about the method are described in the following sections.

Through using the method in interval information system, it is proved that with enhancement of knowledge discernment, dominance entropy and dominance combination entropy. The validity of this method is better illustrated.

RESULTS

In this study, through introducing dominance relations to interval ordered information system, a rough set approach is first established. Then, the concept of dominance entropy and dominance combination entropy in interval ordered information systems is put forward and the numerical result of both is respectively calculated.

The calculation results show that dominance combination entropy in interval ordered information systems along with fine of knowledge classification increases monotonously.

INTERVAL INFORMATION SYSTEMS BASED ON DOMINANCE RELATION

In this section, some basic concepts of interval information systems based on dominance relation are briefly reviewed and some of its important properties are obtained.

Definition 1: An Interval Information System (IIS) is a quadruple:

$$IIS = (U, AT, V, f)$$

where, U is a finite nonempty set of objects, AT is a finite nonempty set of attributes, V is equal to $U_{a\in C}V_a$ and V_a is a domain of attribute a, f is $U\times C\rightarrow V$ is function of an interval information system such that $f(x, a)\in V_a$ for ervey $a\in C$, $x\in U$ where, V_a is a set of interval numbers. Interval number of x under the attribute a is called:

$$f(x, a) = \langle a^{Lower}, a^{Upper}\rangle$$
$$\{q \mid a^{Lower}(x) \le q \le a^{Upper}(x)a^{Lower}(a), a^{Upper}(x) \in R\}$$

Definition 2: An interval information system is called an Interval Ordered Information System (IOIS) if all attributes are criterions.

It is assumed that domain of a criterion $a\in C$ is completely pre-ordered by an outranking relation $\succ a$; $x\succ_a y$ means that x is at least as good as (outranks) y with respect to criterion.

Definition 3: Let S = (U, AT, V, f) be an interval ordered information system, $A\subseteq AT$, then:

$$R_A^{\ge} = \{(y,x)\in U\times U \mid y\succ_A^x\}$$
$$= \{(y,x)\in U\times U \mid \ge a^{Lower}(y) \ge a^{Lower}(x)a^{Upper}(y)a^{Upper}(x)(\forall a\in A)\}$$

The dominance class induced by the dominance relation $R_A^{\ge}$ is the set of objects dominating x, i.e.,:

$$R_A^{\ge} = \{(y,x)\in U\times U \mid y\succ_A^x\}$$
$$= \{(y,x)\in U\times U \mid \ge a^{Lower}(y) \ge a^{Lower}(x)a^{Upper}(y)a^{Upper}(x)(\forall a\in A)\}$$

From the definition of $R_A^{\ge}$ and the following properties can be easily obtained.

Property 1: Let S = (U, AT, V, f) be an interval ordered information system and $A\subseteq AT$, then:

$$R_A^{\ge} = \bigcap_{a\in A} R_a^{\ge}$$

- **Proof:** Straightforward

Property 2: Let S = (U, AT, V, f) be an interval ordered information system and $A\subseteq AT$. Then:

- $R_A^{\ge}$ are reflexive
- $R_A^{\ge}$ are unsymmetric and
- $R_A^{\ge}$ are transitive
- $R_A^{\ge}$ is not a equivalence relation

Property 3: Let $R_A^{\ge}$ be an interval ordered information system, then:

- if $B\subseteq A\subseteq AT$, then $R_B^{\ge}\supseteq R_A^{\ge}\supseteq R_{AT}^{\ge}$
- if $B\subseteq A\subseteq AT$, then $[x]_B^{\ge}\supseteq[x]_A^{\ge}\supseteq[x]_{AT}^{\ge}$
- if $y\in[x]_A^{\ge}$, then $[y]_A^{\ge}\subseteq[x]_A^{\ge}$ and $[x]_A^{\ge}=\cup\{[y]_A^{\ge}: y\in[x]_A^{\ge}\}$

- **Proof:** Straightforward

When preference-orders of attributes domains (criterion) are to be taken into account, the classical rough set approach is incapable of solve the problem.

Definition 4: Let S = (U, AT, V, f) be a interval ordered information system. For any $X\subseteq U$ and $A\in AT$, the lower and upper approximations of X with respect to the dominance relation $R_A^{\ge}$ are defined as follows:

$$\underline{R_A^{\ge}}(X) = \{x\in U \mid [x]_{R_A^{\ge}} \subseteq X\}$$

$$\overline{R_A^{\geq}}(X) = \{x \in U \mid [x]_{R_A^{\geq}} \cap X \neq \phi\}$$

UNCERTAINTY MEASURES IN INTERVAL ORDERED INFORMATION SYSTEMS

In this section, the concepts of entropy in interval ordered information systems will be introduced, then uncertainty of interval ordered information systems is studied. In order to generalize Shannon's entropy in a natural way, a new formulation is given. In the new formulation, entropy are computed with the sum of uncertainty of single object.

Let $[x]_{R_A^{\geq}}$ denote the dominance class in interval ordered information systems induced by objects x_i and attribute set A. The uncertainty of object x_i is computed as:

$$-\log \frac{\left|[x_i]_{R_A^{\geq}}\right|}{|U|}$$

Definition 5: Let S = (U, AT, V, f) be an interval ordered information system and A⊆AT then dominance entropy of the set U with respect to attribute A is defined as follows:

$$H_{R_A^{\geq}}(U) = -\frac{1}{|U|} \cdot \sum_{i=1}^{|U|} \log\left(\frac{\left|[x_i]_{R_A^{\geq}}\right|}{|U|}\right)$$

where, $H_{R_A^{\geq}}(U)$ be the average uncertainty of U, also called dominance entropy of U with respect to A.

Example 1: An interval information system is presented in Table 1, where U = {x1, x2, x3, x4, x5, x6, x7, x8, x9, x10}, AT = {a1, a2, a3, a4, a5}.

A an interval ordered information system is presented in Table 1, where U = {x1, x2, x3, x4, x5, x6, x7, x8, x9, x10}, AT = {a1, a2, a3, a4, a5}:

$$[x_1]_{R_{AT}^{\geq}} = \{x_1, x_5, x_7, x_8\}$$

Table1: An interval-valued table

U	a_1	a_2	a_3	a_4	a_5
x_1	1	[0, 1]	2	1	[1, 2]
x_2	[0, 1]	0	[1, 2]	0	1
x_3	[0, 1]	0	[1, 2]	1	1
x_4	0	0	1	0	1
x_5	2	[1, 2]	3	[1, 2]	[2, 3]
x_6	[10, 2]	[1, 2]	[1, 3]	[1, 2]	[2, 3]
x_7	1	1	2	1	2
x_8	[1, 2]	[1, 2]	[2, 3]	2	[2, 3]
x_9	[1, 2]	2	[2, 3]	[0, 2]	3
x_{10}	2	2	3	[0, 1]	3

$x_1, x_2, \ldots, x_{10}$ represent objects of domain. $a_1, a_2, \ldots, a_4$ represent attributes of objects

$$[x_2]_{R_{AT}^{\geq}} = \{x_1, x_2, x_3, x_5, x_6, x_7, x_8, x_9, x_{10}\}$$

$$[x_3]_{R_{AT}^{\geq}} = \{x_1, x_3, x_5, x_6, x_7, x_8\}$$

$$[x_4]_{R_{AT}^{\geq}} = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}$$

$$[x_5]_{R_{AT}^{\geq}} = \{x_5\},\ [x_6]_{R_{AT}^{\geq}} = \{x_5, x_6, x_8\}$$

$$[x_7]_{R_{AT}^{\geq}} = \{x_5, x_7, x_8\}$$

$$[x_8]_{R_{AT}^{\geq}} = \{x_8\},\ [x_9]_{R_{AT}^{\geq}} = \{x_9\},\ [x_{10}]_{R_{AT}^{\geq}} = \{x_{10}\}$$

Let A = $\{a_1, a_2\}$ then:

$$[x_1]_{R_A^{\geq}} = \{x_1, x_5, x_7, x_8, x_9, x_{10}\}$$

$$[x_2]_{R_A^{\geq}} = \{x_1, x_2, x_3, x_5, x_6, x_7, x_8, x_9, x_{10}\}$$

$$[x_3]_{R_A^{\geq}} = \{x_1, x_2, x_3, x_5, x_6, x_7, x_8, x_9, x_{10}\}$$

$$[x_4]_{R_A^{\geq}} = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}$$

$$[x_5]_{R_A^{\geq}} = \{x_5, x_{10}\},\ [x_6]_{R_A^{\geq}} = \{x_5, x_6, x_8, x_9, x_{10}\}$$

$$[x_7]_{R_A^{\geq}} = \{x_5, x_7, x_8, x_9, x_{10}\}$$

$$[x_8]_{R_A^{\geq}} = \{x_5, x_8, x_9, x_{10}\},\ [x_9]_{R_A^{\geq}} = \{x_9, x_{10}\}$$

$$[x_{10}]_{R_A^{\geq}} = \{x_{10}\}$$

From the above calculation, dominance entropy of the set U in an interval ordered information system in respect to attributes, $\{a_1, a_2\}$, AT are obtained as follows:

$$= -\frac{1}{10}\log\frac{4}{10}\log - \frac{9}{10}\log\frac{1}{10}\log - \frac{6}{10}\log\frac{1}{10}\log - \frac{10}{10}\log\frac{1}{10}\log\frac{1}{10}$$

$$= -\frac{1}{10}\log\frac{3}{10}\log - \frac{1}{10}\log\frac{3}{10}\log - \frac{1}{10}\log\frac{1}{10}\log - \frac{1}{10}\log\frac{1}{10}\log\frac{1}{10} = 0.57103$$

$$H_{R_{\{a_1,a_2\}}^{\geq}}(U) = -\frac{1}{|U|} \cdot \sum_{i=1}^{|U|} \log\left(\frac{\left|[x_i]_{R_{\{a_1,a_2\}}^{\geq}}\right|}{|U|}\right)$$

$$= -\frac{1}{10}\log\frac{6}{10} - \frac{1}{10}\log\frac{9}{10} - \frac{1}{10}\log\frac{9}{10} - \frac{1}{10}\log\frac{10}{10} - \frac{1}{10}\log\frac{2}{10}$$

$$=-\frac{1}{10}\log\frac{5}{10}-\frac{1}{10}\log\frac{5}{10}-\frac{1}{10}\log\frac{4}{10}-\frac{1}{10}\log\frac{2}{10}-\frac{1}{10}\log\frac{1}{10}=0.37095$$

From example 1, it may be obtained as follows.

$$H_{R_{AT}^{\geq}}(U)>H_{R_{\{a_1,a_2\}}^{\geq}}(U)$$

Moreover, one may learn that as knowledge partition becomes finer, dominance entropy of the set U in an interval ordered information system monotonously increases.

In general, the elements in an dominance relation class in interval ordered information systems cannot be distinguished each other but the elements in different dominance relation classes can be distinguished each other in rough set theory. In a sense, the knowledge content of an approximation space K = (U, $R_A^{\geq}$) is the whole number of distinguishable pairs of the elements on the universe U in interval ordered information systems . With this consideration, the dominance combination entropy to discourse the uncertainty of interval ordered information systems is introduced.

Definition 6: Let S = (U, AT,V, F) be an interval ordered information system and $A \subseteq AT$ then dominance combination entropy of the set U with respect to attribute A is defined as follows:

$$CH_{R_A^{\geq}}(U)=\frac{1}{|U|}\cdot\sum_{i=1}^{|U|}\left(1-\frac{C^2_{\left|[x_i]_{R_A^{\geq}}\right|}}{C^2_{|U|}}\right)$$

Example 2 (continue): Dominance combination entropy of the set U in an interval ordered information system in respect to attributes, $\{a_1, a_2\}$, AT as follows may be calculated:

$$CH_{R_{AT}^{\geq}}(U)=\frac{1}{|U|}\cdot\sum_{i=1}^{|U|}\left(1-\frac{C^2_{\left|[x_i]_{R_{AT}^{\geq}}\right|}}{C^2_{|U|}}\right)$$

$$=\frac{1}{10}\Big[(1-\frac{6}{45})+(1-\frac{36}{45})+(1-\frac{15}{45})+(1-\frac{45}{45})+(1-\frac{0}{45})$$

$$+(1-\frac{3}{45})+(1-\frac{3}{45})+(1-\frac{0}{45})+(1-\frac{0}{45})+(1-\frac{0}{45})\Big]=0.76$$

$$CH_{R_{\{a_1,a_2\}}^{\geq}}(U)=\frac{1}{|U|}\cdot\sum_{i=1}^{|U|}\left(1-\frac{C^2_{\left|[x_i]_{R_{\{a_1,a_2\}}^{\geq}}\right|}}{C^2_{|U|}}\right)$$

$$=\frac{1}{10}\Big[(1-\frac{15}{45})+(1-\frac{36}{45})+(1-\frac{36}{45})+(1-\frac{45}{45})+(1-\frac{1}{45})$$

$$+(1-\frac{10}{45})+(1-\frac{10}{45})+(1-\frac{6}{45})+(1-\frac{1}{45})+(1-\frac{0}{45})\Big]=0.64$$

From example 2, it may be obtained as follows:

$$CH_{R_{AT}^{\geq}}(U)\geq CH_{R_{\{a_1,a_2\}}^{\geq}}(U)$$

The calculation results show that dominance combination entropy in interval ordered information systems along with fine of knowledge classification increases monotonously.

Proposition 1: Let $R_A^{\geq}$ and $R_B^{\geq}$ be two dominance relations on a nonempty and finite set U in interval ordered information systems:

$$R_A^{\geq}\in R_B^{\geq}\Rightarrow CH_{R_A^{\geq}}(U)\geq CH_{R_B^{\geq}}(U)$$

Obviously, $R_A^{\geq}\in R_B^{\geq}$ as above is known, that is classification of $R_A^{\geq}$ is finer that of $R_B^{\geq}$.

CONCLUSION

To recapitulate, interval information systems are an important type of data tables, which are generalized models of single-valued information systems. In this study, through introducing dominance relations to interval ordered information system, a rough set approach is first established. Then, the concept of dominance entropy and dominance combination entropy in interval ordered information systems is put forward and the numerical result of both is respectively calculated. It is proved that with enhancement of knowledge discernment, dominance entropy and dominance combination entropy in interval information systems based on dominance relations increases monotonously. These results give a kind of feasible approaches to discover and acquisition of knowledge in interval ordered information systems.

ACKNOWLEDGMENT

This study was supported by the Shanxi Province Science Research Project (Grant No. 20110311039-2).

REFERENCES

Cheng, S.W., R.W. Dai, W.X. Xu and Y. Shi, 2006. Research on data mining and knowledge management and its applications in China's economic development: Significance and trend. Int. J. Inform. Technol. Decis. Making, 5: 585-596.

Dembczynski, K., R. Pindur and R. Susmaga, 2003. Generation of exhaustive set of rules within dominance-based rough set approach. Electron. Notes Theor. Comput. Sci., 82: 96-107.

Facchinetti, G., R.G. Ricci and S. Muzzioli, 1998. Note on ranking fuzzy triangular numbers. Int. J. Intell. Syst., 13: 613-622.

Greco, S., B. Matarazzo and R. Slowinski, 1998. A new rough set approach to multicriteria and multiattribute classification. Proceedings of the 1st International Conference on Rough Sets and Current Trends in Computing, June 22-26, 1998, Warsaw, Poland, pp: 60-67.

Greco, S., B. Matarazzo and R. Slowinski, 2001. Rough sets theory for multicriteria decision analysis. Eur. J. Oper. Res., 129: 1-47.

Greco, S., B. Matarazzo and R. Slowinski, 2002. Rough sets methodology for sorting problems in presence of multiple attributes and criteria. Eur. J. Operat. Res., 8: 247-259.

Greco, S., B. Matarazzo, R. Slowinski and J. Stefanowski, 2005. An algorithm for induction of decision rules consistent with the dominance principle. Proceedings of the 2nd International Conference on Rough Sets and Current Trends in Computing, October 16-19, 2000, Banff, Canada, pp: 304-313.

Hu, Q., Z. Xie and D. Yu, 2007. Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation. Pattern Recognit., 40: 3509-3521.

Hu, Q.H., D, Yu, J.F. Liu and C.X. Wu, 2008. Neighborhood rough set based heterogeneous feature subset selection. Inform. Sci., 178: 3577-3594.

Liang, J., K.S. Chin, C. Dang and R.C. Yam, 2002. A new method for measuring uncertainty and fuzziness in rough set theory. Int. J. Gen. Syst., 31: 331-342.

Liang, J.Y. and K.S. Qu, 2001. Information measures of roughness of knowledge and rough sets in incomplete information systems. J. Syst. Sci. Syst. Eng., 24: 544-547.

Liang, J.Y. and Z.B. Xu, 2002. The algorithm on knowledge reduction in incomplete information systems. Int. J. Uncertainty Fuzziness Knowl. Based Syst., 10: 95-103.

Liang, J.Y. and Z.Z. Shi, 2004. The information entropy, rough entropy and knowledge granulation in rough set theory. Int. J. Uncertainty Fuzziness Knowl. Based Syst., 12: 37-46.

Malyszko, D. and J. Stepaniuk, 2008. Standard and fuzzy rough entropy clustering algorithms in image segmentation. Proceedings of the 6th International Conference on Rough Sets and Current Trends in Computing, Volume 5306, October 23-25, 2008, Akron, OH., USA., pp: 409-418.

Malyszko, D. and J. Stepaniuk, 2010a. Adaptive multilevel rough entropy evolutionary thresholding. Inform. Sci., 180: 1138-1158.

Malyszko, D. and J. Stepaniuk, 2010b. Adaptive rough entropy clustering algorithms in image segmentation. Fund. Informaticae, 98: 199-231.

Mi, J.S., Y. Leung and W.Z. Wu, 2005. An uncertainty measure in partition-based fuzzy rough sets. Int. J. Gen. Syst., 34: 77-90.

Pawlak, Z. and A. Skowron, 2007. Rudiments of rough sets. Inform. Sci., 177: 3-27.

Qian, Y. and J. Liang, 2006. Combination entropy and combination granulation in incomplete information system. Proceedings of the 1st International Conference on Rough Sets and Knowledge Technology, Volume 4062, July 24-26, 2006, Chongquing, China, pp: 184-190.

Sai, Y., Y.Y. Yao and N. Zhong, 2001. Data analysis and mining in ordered information tables. Proceedings of the IEEE International Conference on Data Mining, November 29- December 2, 2001, San Jose, CA., USA., pp: 497-504.

Shao, M.W. and W.X. Zhang, 2005. Dominance relation and rules in an incomplete ordered information system. Int. J. Intell. Syst., 20: 13-27.

Shi, Y., Y. Peng, W.X. Xu and X.W. Tang, 2002. Data mining via multiple criteria linear programming: Applications in credit card portfolio management. Int. J. Inform. Technol. Decis. Making, 1: 131-151.

Slezak, D., 2002. Approximate entropy reducts. Fund. Inform., 53: 365-390.

Slezak, D., 2005. Association reducts: A framework for mining multi-attribute dependencies. Proceedings of the 15th International Symposium on Foundations of Intelligent Systems, Volume 3488, May 25-28, 2005, Saratoga Springs, NY., USA., pp: 354-363.

Wei, W., J.Y. Liang, Y.H. Qian, F. Wang and C.Y. Dang, 2010. Comparative study of decision performance of decision tables induced by attribute reductions. Int. J. Gen. Syst., 39: 813-838.

Xu, Z.S. and H.F. Gu, 2002. An approach to uncertain multi-attribute decision making with preference information on alternatives. Proceedings of the 9th Bellman Continuum International Workshop on Uncertain Systems and Soft Computing, July 24-27, 2002, Beijing, China, pp: 89-95.

Xu, Z.S. and Q.L. Da, 2002. The uncertain OWA operator. Int. J. Intell. Syst., 17: 569-575.

Yang, X.B., M. Zhang, H.L. Dou and J.Y. Yang, 2011. Neighborhood systems-based rough sets in incomplete information system. Knowledge-Based Syst., 24: 858-867.

Yao, Y. and Y. Zhao, 2008. Attribute reduction in decision-theoretic rough set models. Inform. Sci., 178: 3356-3373.

Yao, Y.Y., 2007. Decision-Theoretic Rough Set Models. In: Rough Sets and Knowledge Technology, Yao, J.T., P. Lingras, W.Z. Wu, M. Szczuka, N.J. Cercone and D. Slezak (Eds.). Springer, Berlin Heidelberg, Germany, pp: 1-12.

Zhang, J.B., T.R. Li and D. Liu, 2010. An approach for incremental updating approximations in Variable precision rough sets while attribute generalized. Proceedings of the 2010 IEEE International Conference on Intelligent Systems and Knowledge Engineering, November 15-16, 2010, Hangzhou, China, pp: 77-81.