Journal of Software Engineering

ISSN 1819-4311

Academic
Journals Inc.

www.academicjournals.com

Journal of Software Engineering 10 (1): 129-137, 2016
ISSN 1819-4311 / DOI: 10.3923/jse.2016.129.137

An Evaluation Study on Information System Based on Rough Set and Condition Information Entropy

Sha Fu, Guang Sun, Hangjun Zhou and Feng Yan
Department of Information Management, Hunan University of Finance and Economics, 410205, People's Republic of China

Corresponding Author: Sha Fu, Department of Information Management, Hunan University of Finance and Economics, 410205, People's Republic of China Tel: 13875982792/24849322

ABSTRACT

Based on the incompleteness and conceptual uncertainty of information in management decision making and evaluation, the rough set theory and condition information entropy were introduced to build a comprehensive evaluation model based on the rough set condition information entropy, so as to present the tendency of experts' experience and knowledge towards index importance. To solve the problem of index weight acquisition in that system, the decision table was partitioned according by analyzing and evaluating characteristics of the small and medium sample data according to the factual condition, so as to obtain its weight value in an objective manner through hierarchical calculation method from the aspect of information entropy and eventually obtain the comprehensive evaluation result of the information system. Through real case analysis, the feasibility and effectiveness of the rough set intelligent evaluation model were verified.

Key words: Rough set, condition information entropy, information system, index system, index weight

INTRODUCTION

With progression of the global informationization, the coverage of information has reached various aspects of the country and society. People are now subject to unprecedented impact and control while enjoying the convenience of information. As the information technology develops rapidly, the scale of information system with network as the carrier and information resources as the core keeps expanding. At the same time, the safety risk and threat also increasing. To formulate an effective scheme to avoid risk in an early stage, it is urgently needed to evaluate the information system effectively. During evaluation, weight is usually used to present the relative importance of an index to the object being evaluated. The method to determine weight has been a hot topic of study in the field of decision making and evaluation and its reasonableness and effectiveness pose a direct impact on the scientific property and reliability of evaluation. Therefore, the way to determine weight reasonably is the key issue with the information system evaluation. Currently, the main methods used for weight determination include expert grading method, entropy weight method, fuzzy statistical method, binary ordered correlation method and the grey correlation method (Leung *et al.*, 2008). These methods aim at studying the weight from different aspects but all have certain scope of application and limitedness. In all, both the subjective weighting method which reflects the subjective judgment and experience of the decision maker and the objective weighting method which is based on the comparatively complete mathematical theory and method are too heavily relying on the experience and knowledge of the decision maker (Danielson, 2005). Grimsley and Meehan (2007) developed an evaluative design framework for e-Government projects that complements traditional approaches to information systems evaluation. The framework is based upon Moor's concept of public value. Irani (2002) proposed a set of conjectures. These were tested within a case study to analyze the investment justification process of a manufacturing information systems investment. Information systems make it possible to improve organizational efficiency and effectiveness which can provide competitive advantage. Xu *et al*. (2014) studied the method to determine weight using the rough set and information entropy theory and has established the multi-effect size mixed evaluation model for dam safety using from the aspect of probability theory and information theory. Chen *et al*. (2013) applied the information entropy in the information theory to the field of abnormal data mining and proposed the abnormal data mining theory and method based on information entropy in the framework of granular computation. Irani *et al*. (2005) attempted to highlight those extant components of knowledge that contribute to the overall information systems engineering process. This is achieved via an analysis of case study data against the well-known knowledge transformation model proposed by Nonaka and Takeuchi (1995). Mao *et al*. (2013) considered the factors that pose an impact on the information resource allocation effect of local government and established as set of related evaluation index system and set up an intelligent evaluation model based on the rough set theory with the foundation he has already laid. By now, the study on application of rough set theory and condition information entropy to information system evaluation is not mature. This study aims at establishing a scientific and reasonable evaluation model and discussing similar issues with other researchers.

MATERIALS AND METHODS

Preliminaries

Information entropy theory: The entropy concept was firstly proposed by Mr. Rudolf Clausius, a German physicist, in 1865 and was then introduced into the information field by the father of information theory C.E. Shannon in 1948 via utilizing the “Information entropy” as a measure of the disorder degree of information (Liang and Qian, 2008). Shannon systematically presented the measuring method for information going as used entropy to measure uncertainty or information quantity of a random event with probability statistics which laid the scientific theoretical basis for modern information theory and extended the quantified application of entropy into studies on uncertainty and random quantification of the system (Ding *et al*., 2010):

$$H(X) = H(p_1, p_2, \ \ldots \ , p_n) = -k\sum_{i=1}^{n} P_i \log P_i$$

Information entropy is an abstract concept in mathematics and it refers to the probability of occurrence of certain information and is used to represent the uncertainty of information system. The high the uncertainty of an information system, the lower the information entropy will be and vice versa. In different systems, entropy can be used to measure state confusion or disorder, uncertainty or lack of information, in homogeneities or richness and etc (Dzazali and Zolait, 2012). Therefore, appraising order degree of the system architecture with extensive application of the theory on information entropy became a new approach for today's system architecture evaluation.

Rough set theory: Rough set theory was proposed by Pawlak (1982), a Polish mathematician, in 1982 as a new theory method to express, study and induce the knowledge and data that are not certain and processed completely. The most distinct difference between this theory and the probability statistics, fuzzy set theory and evidence theory is that it needs no prior information other than the data set to be processed.

Definition 1: Assume S = (U, A, V, f) as a decision table, where $U = \{u_1, u_2, \ldots, u_n\}$ is a non-empty and limited set and is also called the domain of disclosure; A is the non-empty and limited set of the attribute; V is the attribute range:

$$V = \bigcup_{\alpha \in A} V_\alpha$$

where, V_α is the range of attribute a; f:U×A→V is regarded as an information function it gives an attribute value to each attribute of each object. When the attribute in the information system:

$$A = C \cup D$$

where, C is the condition attribute set and D is the decision attribute set (Yao, 2010).

Definition 2: Given decision table S, if B⊆A, then the indiscernible relation on the definition attribute set B IND(B) should be:

$$IND(B) = \{(u_i, u_j) \in U^2 \mid \forall \alpha \in B, f(u_i, \alpha) = f(u_j, \alpha)\}$$

U/IND(B) constitutes a partition of U and is regarded as knowledge of U. Each equivalence class is regarded as a knowledge grain.

Definition 3: From the aspect of the probability theory, the equation S = (U, A, V, f) is essentially a stochastic system. Assume X as a random variable with attribute partition characteristics on the non-empty and limited domain of disclosure U, $X = \{X_1, X_2, \ldots, X_n\}$ and then its probability measure distribution could be determined as follows (Zhen and Shi, 2011):

$$[X : p] = \begin{bmatrix} X_1 & X_2 & \cdots & X_n \\ p(X_1) & p(X_2) & \cdots & p(X_n) \end{bmatrix} \tag{1}$$

Where:

$$p(X_i) = \frac{|X_i|}{|U|}$$

where, i = 1, 2, ..., n; $|X_i|$ is the cardinal number of set X_i.

Definition 4: From the aspect of information, assume S = (U, A, V, f) as an information system and $U/A = \{X_1, X_2, \ldots, X_n\}$, then the entropy function in the information entropy theory could be introduced. Information entropy of A should be:

$$\begin{aligned} H(A) &= H(p(X_1), p(X_2), \cdots, p(X_n)) \\ &= -k\sum_{i=1}^{n} p(X_i)\log p(X_i) \end{aligned} \quad (2)$$

Definition 5: In the decision table S, if U/IND(C) = $\{X_1, X_2, ..., X_q\}$, U/IND(D) = $\{Y_1, Y_2, ..., Y_p\}$, then the information entropy of object set U under the condition attribute C corresponding to the decision attribute D should be defined as follows (Gao and Tan, 2012):

$$I(D\,|\,C) = -\sum_{s=1}^{q}\frac{Card(X_s)}{Card(U)}\times\sum_{k=1}^{p}\frac{Card(Y_k \cap X_s)}{Card(X_s)}\times\log_{10}\left(\frac{Card(Y_k \cap X_s)}{Card(X_s)}\right) \quad (3)$$

where, Card(*) indicates the cardinal number of the set which is the number of set elements.

Definition 6: In the decision table S, the importance of definition attribute $c \in C$ under the rough set information entropy connotation should be:

$$SGF(c) = I(D\,|\,C) - I(D\,|\,C - \{c\}) \quad (4)$$

Comprehensive intelligent evaluation model based on rough set and condition information entropy

Issue description: This study introduced the rough set theory and condition information entropy into the information system evaluation to establish a comprehensive and intelligent evaluation model based on rough set and condition information entropy. It utilized the relationship between the knowledge and information entropy in the rough set theory, to express the main concept and operation of rough set theory using the information representation method from the aspect of information (Grover and Segars, 2005).

The evaluation index system contains the level-1 index and level-2 index condition attribute set and describes the condition attribute set (index set) C of definition 1 in decision table S as follows: The condition attribute set is described as: $C = \{C_1, C_2, ..., C_m\}$; for C_x ($x = 1, 2, ..., z$), as the level-1 index condition attribute, it contains several level-2 index condition attributes which could be further described as: $C_x = \{C_{x1}, C_{x2}, ..., C_{xv}\}$. Therefore, $\forall u_i \in U$. The grade of the object under the level-2 index condition attribute C_{xv} could be described as l^i_{xn}.

Discretization of continuous data: As the rough set could only be used to process discretized data, the continuous data should be subject to discretization processing. The equidistance method is used here (Gao and Tan, 2012). The specific steps are as follows:

- The value interval length of attribute C_x during discretization should be calculated as follows:

$$l_x^* = \frac{\max(\tilde{l}_x^i) - \min(\tilde{l}_x^i)}{m} \quad (5)$$

where, l_x^* is the interval length, $\max(\tilde{l}_x^i)$ is the maximum grade in attribute C_x, $\min(\tilde{l}_x^i)$ is the minimum grade value in attribute C_x and m is the set number of discretization intervals.

- For object u_i, its discretization result under attribute C_x is calculated as follows:

$$t_x^i = \left\langle \frac{\tilde{l}_x^i - \min(\tilde{l}_x^i)}{l_x^*} \right\rangle \tag{6}$$

where, t_x^i is the discretization result of the value under attribute C_x of object u_i and <*> refers to rounding up to an integer.

Building a decision table: The data discretization results were converted into the set in decision table S = (U, A, V, f), U = $\{u_1, u_2, ..., u_n\}$ used to express various evaluation objects. The condition attribute evaluation index set C = $\{C_1, C_2, ..., C_z\}$ and decision attribute set D = $\{d_1, d_2, ..., d_k\}$.

Determination of objective weight of evaluation index: The rough set calculation method based on the information entropy concept was introduced into comprehensive and intelligent evaluation and analysis to obtain the objective weight of the attribute and to provide a new solution to the sample data issue rough set solving.

Steps of Algorithms: For the intelligent evaluation model based on the rough set condition information entropy as described in this paper, according to the decision table S = (U, A, V, f) and other input data, the condition attribute weight value w and the comprehensive evaluation value E_i (Mao *et al*., 2014) of the object to be evaluated could be calculated through the following steps:

- **Step 1:** Obtain the partition results U/IND(C) of the entire object set U, the partition result U/IND(D) on decision attribute D, as well a the partition results U/IND(C-$\{C_x\}$) after the condition attribute C_x if removed each time
- **Step 2:** Calculate the information entropy I(D | C) partitioned on the partition relative decision attribute D on the condition attribute set C; after removing the condition attributes C_x under the condition attribute set C one by one, calculate the partitioned information entropy I(D | C-$\{C_x\}$) of the partition relative decision attribute D on C-$\{C_x\}$
- **Step 3:** Calculate the importance of each condition attribute C_x under the rough set connotation

$$SGF(\{C_x\}) = I(D \mid C) - I(D \mid C - \{C_x\}) \tag{7}$$

- **Step 4:** Calculate the weight value of each condition attribute C_x

$$w(C_x) = \frac{SGF(\{C_x\})}{\sum_{x=1}^{z} SGF(\{C_x\})} \tag{8}$$

- **Step 5:** Calculate the comprehensive evaluation results of the evaluated object

$$E_i = \sum_{x=1}^{z} w(C_x) \times \tilde{l}_x^i \tag{9}$$

where, E_i is the comprehensive evaluation value of various objects to be evaluated and $\tilde{l}_x^i$ is the grade value of various attributes.

RESULTS AND DISCUSSION

Case analysis: The purchase of e-government system integration service of an organization was taken as an example. The buyer was to select 1-2 optimal enterprises from 12 information system integration service suppliers (enterprise u_i, i = 1, 2, ..., 12) in the central region of China to provide the software and hardware construction and follow-up services for its e-government system. According to the business requirement and requirement on performance of the e-government system of the buyer, there are 5 main evaluation indexes (attributes): System quotation (C_1); time promised to complete the integration service (C_2); usability, safety and expansibility of the software system (C_3); research and development capacity and after-sales service capacity of the enterprise (C_4) and the reputation of the enterprise (C_5) (Huang and Yang, 2009). The evaluating conducted comprehensive evaluation on various attributes of the 12 enterprises u_i. Decision attribute D is the experts' satisfaction. The data obtained are listed in Table 1.

The evaluation indexes are divided into 3 levels. According to the discretization interval length of each index as determined via Eq. 5 and the discretization results of each index as determined via Eq. 6, the specific results are shown in Table 2.

- **Step 1:** Calculate the partition results U/IND(C) of object set U in the condition attribute set block C and its partition result U/IND(D) in decision attribute D

Table 1: Initial information in the information system

Information system evaluation	Condition attribute set C					Decision attribute D
	C_1	C_2	C_3	C_4	C_5	
u_1	95	96	62	85	65	74.38
u_2	80	80	80	96	65	90.13
u_3	85	82	72	90	96	88.33
u_4	62	65	62	85	78	75.21
u_5	78	70	65	78	80	83.13
u_6	68	90	82	78	65	81.49
u_7	70	68	95	72	82	92.53
u_8	72	72	81	80	65	72.37
u_9	62	65	92	65	94	70.13
u_{10}	95	92	95	78	62	86.67
u_{11}	82	85	80	72	68	80.38
u_{12}	90	68	65	85	80	96.35

Table 2: Decision table after data discretization

Information system evaluation	C					D
	C_1	C_2	C_3	C_4	C_5	
u_1	3	3	1	2	1	1
u_2	2	2	2	3	1	3
u_3	3	2	1	3	3	3
u_4	1	1	1	2	2	1
u_5	2	1	1	2	2	2
u_6	1	3	2	2	1	2
u_7	1	1	3	1	2	3
u_8	1	1	2	2	1	1
u_9	1	1	3	1	3	1
u_{10}	3	3	3	2	1	2
u_{11}	2	2	2	1	1	2
u_{12}	3	1	1	2	2	3

U/IND(C) = {{1}, {2}, {3}, {4}, {5}, {6}, {7}, {8}, {9}, {10}, {11}, {12}}
U/IND(D) = {{1, 4, 8, 9}, {5, 6, 10, 11}, {2, 3, 7, 12}}

- **Step 2:** Remove the level-1 index condition attributes C_n (n = 1, 2, ..., 5) one by one in the condition attribute set C and obtain the partition results U/IND(C-{C_n})

U/IND(C-{C_1}) = {{1}, {2}, {3}, {4, 5, 12}, {6}, {7}, {8}, {9}, {10}, {11}}
U/IND(C-{C_2}) = {{1}, {2}, {3}, {4}, {5}, {6, 8}, {7}, {9}, {10}, {11}, {12}}
U/IND(C-{C_3}) = {{1, 10}, {2}, {3}, {4}, {5}, {6}, {7}, {8}, {9}, {11}, {12}}
U/IND(C-{C_4}) = {{1}, {2,11}, {3}, {4}, {5}, {6}, {7}, {8}, {9}, {10}, {12}}
U/IND(C-{C_5}) = {{1}, {2}, {3}, {4}, {5}, {6}, {7, 9}, {8}, {10}, {11}, {12}}

- **Step 3:** Calculate the information entropy I(D | C) and I(D | C-{C_n})

Calculate the information entropy I(D | C) according to Eq. 3, where, S = 1, 2, ..., 12; k = 1, 2, 3
When S = 1, X_1 = {1}, k = 1, 2, 3:

$$\sum_{k=1}^{p}\frac{\mathrm{Card}(Y_k \cap X_1)}{\mathrm{Card}(X_1)}\times\log_{10}\left(\frac{\mathrm{Card}(Y_k \cap X_1)}{\mathrm{Card}(X_1)}\right)=\log_{10}1=0$$

Then, the values are all 0 when the S = 2, 3, ..., 12 is calculated.
Through calculation, we have I(D | C) = 0.
In a similar way, information entropy I(D | C-{C_n}) could be calculated based on Eq. 3. where, S = 1, 2, ..., 10; k = 1, 2, 3.
Then we have:

$$\mathrm{I(D\,|\,C\text{-}\{C_1\})}=-\frac{3}{12}\times\log_{10}\frac{1}{3}=\frac{1}{4}\log_{10}3$$

In a similar way, the following equations could be drawn:

$$\mathrm{I(D\,|\,C\text{-}\{C_2\})}=-\frac{2}{12}\times\log_{10}\frac{1}{2}=\frac{1}{6}\log_{10}2$$

$$\mathrm{I(D\,|\,C\text{-}\{C_3\})}=\mathrm{I(D\,|\,C\text{-}\{C_4\})}=\mathrm{I(D\,|\,C\text{-}\{C_5\})}=\frac{1}{6}\log_{10}2$$

- **Step 4:** Based on Eq. 7, the relative importance SGF({C_n}) of various level-1 index condition attributes C_n (n = 1, 2, ..., 5) in the condition attribute set block C under the rough set connotation is calculated

$$\mathrm{SGF(\{C_1\})}=\mathrm{I(D\,|\,C)}-\mathrm{I(D\,|\,C-\{C_1\})}=0-\frac{1}{4}\log_{10}3=-\frac{1}{4}\log_{10}3$$

Table 3: Comprehensive evaluation results and ranking

Information system	Comprehensive evaluation (E_i)	Ranking
u_1	83.710	3
u_2	80.157	5
u_3	85.000	2
u_4	68.586	12
u_5	75.021	8
u_6	74.742	9
u_7	75.802	7
u_8	73.568	10
u_9	72.662	11
u_{10}	86.690	1
u_{11}	78.394	6
u_{12}	80.278	4

$$\mathrm{SGF}(\{C_2\}) = \mathrm{I}(D \mid C) - \mathrm{I}(D \mid C - \{C_2\}) = 0 - \frac{1}{6}\log_{10} 2 = -\frac{1}{6}\log_{10} 2$$

$$\mathrm{SGF}(\{C_3\}) = \mathrm{SGF}(\{C_4\}) = \mathrm{SGF}(\{C_5\}) = -\frac{1}{6}\log_{10} 2$$

- **Step 5:** Calculate the weight values of the level-1 indexes using the rough set methods in the information entropy concept

$$w(C_1) = 0.3728,\ w(C_2) = 0.1568,\ w(C_3) = 0.1568,\ w(C_4) = 0.1568,\ w(C_5) = 0.1568$$

- **Step 6:** Calculate the comprehensive evaluation results of the 12 enterprises in this region to be evaluation using Eq. 9. The specific results are shown in Table 3

As indicated by the evaluation results drawn, the optimal enterprise is u_{10}, followed immediately by u_3. This result coincides with the actual operation condition of the enterprises. This real case demonstrated that the model built in this study can satisfactorily solve the difficulty in quantitative evaluation of qualitative indexes in the information system evaluation and overcome the uncertainty and fuzziness of subjective judgment. It can promote the scientific property of decision making. Compared with the evaluation method used in references, the method proposed in this study better meets the actual application demand and boasts higher feasibility.

CONCLUSION

To get a more subjective evaluation decision conclusion, the rough set theory and condition information entropy are introduced into the information system evaluation analysis to build a comprehensive intelligent model based on rough set condition information entropy and obtain the weight values of various evaluation indexes using the hierarchical calculation method in a subjective fashion, so as to finally complete the decision evaluation analysis of the evaluation system. In all, the method to determine weight based on rough set theory and condition information entropy could both reflect the importance of the attribute in the attribute set, can reflect the importance of the attribute indexes for attribute decision set, thus guaranteeing the subjectivity and consistency of evaluation and ensuring comprehensive and reasonable overall evaluation results. Through real case analysis, it was demonstrated that the intelligent evaluation model built in this paper based on rough set theory is feasible and reasonable. This study provided new thoughts and paths for information system evaluation.

ACKNOWLEDGMENTS

This study was supported by the Scientific Research Fund of Hunan Provincial Education Department (No. 14C0184), by the Hunan Province Philosophy and Social Science Fund Project (No. 14YBA065) and also supported by Hunan Province Science and Technology Program Project (No. 2014GK3042). Supported by the construct program of the key discipline in Hunan province.

REFERENCES

Chen, Y.M., K.S. Wu and X.J. Li, 2013. A kind of outlier mining algorithm based on information entropy. Contr. Decision, 28: 867-872.

Danielson, M., 2005. Generalized evaluation in decision analysis. Eur. J. Oper. Res., 162: 442-449.

Ding, W.P., J.D. Wang, Z.J. Guan and H. Zhu, 2010. Algorithm of attribute reduction based on extension entropy of variable precision thresholding in incomplete information system. J. Chin. Comput. Syst., 31: 2372-2376.

Dzazali, S. and A.H. Zolait, 2012. Assessment of information security maturity: An exploration study of Malaysian public service organizations. J. Syst. Inform. Technol., 14: 23-57.

Gao, W.C. and X. Tan, 2012. Analysis of students' evaluation of teaching based on rough set theory with unknown decision attribute. Comput. Eng. Applic., 48: 238-241.

Grimsley, M. and A. Meehan, 2007. e-Government information systems: Evaluation-led design for public value and client trust. Eur. J. Inform. Syst., 16: 134-148.

Grover, V. and A.H. Segars, 2005. An empirical evaluation of stages of strategic information systems planning: Patterns of process design and effectiveness. Inform. Manage., 42: 761-779.

Huang, J.W. and L. Yang, 2009. Modified TOPSIS decision making method for information system selection. J. Guangxi Acad. Sci., 25: 92-94.

Irani, Z., 2002. Information systems evaluation: Navigating through the problem domain. Inform. Manage., 40: 11-24.

Irani, Z., A.M. Sharif and P.E. Love, 2005. Linking knowledge transformation to information systems evaluation. Eur. J. Inform. Syst., 14: 213-228.

Leung, Y., M.M. Fischer, W.Z. Wu and J.S. Mi, 2008. A rough set approach for the discovery of classification rules in interval-valued information systems. Int. J. Approximate Reasoning, 47: 233-246.

Liang, J.Y. and Y.H. Qian, 2008. Information granule and entropy theory in information system. Sci. China (Ser. E: Inform. Sci.), 38: 2048-2065.

Mao, T.T., K. Xiao and X. Tan, 2013. Local government information resources allocation evaluation based on rough set theory. J. Intell., 32: 116-122.

Mao, T.T., K. Xiao and K. Zou, 2014. A research on multiple-indicators comprehensive evaluation method based on rough set and conditional information entropy. Stat. Res., 31: 92-96.

Nonaka, I. and H. Takeuchi, 1995. The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation. Oxford University Press, New York, ISBN-13: 9780195092691, Pages: 284.

Pawlak, Z., 1982. Rough sets. Int. J. Comput. Inform. Sci., 11: 341-356.

Xu, K.Z., J.P. He and W.Q. Xia, 2014. A fusion evaluation model of dam safety based on rough set and conditional information entropy. Eng. J. Wuhan Univ., 47: 46-49.

Yao, J.T., 2010. Novel Developments in Granular Computing: Applications for Advanced Human Reasoning and Soft Computation: Applications for Advanced Human Reasoning and Soft Computation. IGI Global, Hershey, ISBN: 13- 9781605663258, pp: 401-416.

Zhen, Y. and H. Shi, 2011. Attribute reduction algorithm based on relation coefficient and conditional information entropy. Comput. Eng. Applic., 47: 26-28.