Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2014, 6(4):176-181

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

A new comprehensive attribute weight algorithm with rough sets theory

Yang Su-min*, Meng Jie[1], Liu Qi-ming[2] and Wang Kai[2]

[1]*State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System, Luoyang, Henan, China*
[2]*Department of Information Engineering Ordnance Engineering College, Shijiazhuang, Hebei, China*

ABSTRACT

In view of the deficiency of the present attribute weight methods based on the rough sets theory, the author proposes one new comprehensive attribute weight method through studying deeply attribute importance on the basis of rough sets theory. The proposed method considers objective weight and subjective weight. The objective weight includes three factors, named as the importance of the attribute itself, the increment of mutual information, and its own information entropy. The subjective weight is obtained by the experts with prior knowledge in the field. Experiment results prove that the new method not only overcomes the deficiency of the existing weight methods, but also is more in line with the actual situation.

Key words: Mutual information; Attribute weight; Attribute importance; Rough sets theory; Information entropy

---

INTRODUCTION

The attribute weight is very important in the process of management decision or evaluation, which not only can directly affect the judgment of managers, but also can directly affect the final decision results. It reflects the position and function of various attributes in the process of judgment and decision. Because of the importance of the weight, many scholars are dedicated to researching on it, and a lot of effective methods of assigning weight are produced at the present. A few methods which are more popular are expert evaluation, fuzzy statistics and binary sort etc. But rough sets theory[1,2] is a valid mathematical theory developed in recent years, which can analyze and deal with imprecise, incomplete and inconsistent information effectively, and can dig out the connotative knowledge, and reveal potential rules[3-11]. Article [12] proposed the objective weight method with the rough sets theory, the proposed algorithm only considers the effect of a single attribute on the decision results, and ignores the interactive influence between attributes, so some of the attribute weights are 0, but these attributes with zero weight are necessary for the decision result. Article [13] put forward an improved weight method based on the rough sets theory, which overcomes the problem of causing weight to be equal 0. Article [15] proposed the attribute weight method based on rough sets, which considers the overall importance of condition attributes, as well as the individual importance of each attribute. Article [16] put forward one new objective attribute weight determining method by integrating information view and the algebra view of rough set. Article[17] applied rough sets theory to determine the significance of each parameter and then shifted the significance of attribution to weighting coefficients, and together with the subjective weighting method, the author brought forward a new synthetic weighting method. The weighting algorithm of article [18] took into account the role of the attribute itself and the attribute interactive influence. Above all methods utilize the rough sets theory, because the rough sets can effectively analyze and deal with imprecise, incomplete and inconsistent information, and can dig up the implicit knowledge, and can reveal the potential regularity features, the attribute weight which is obtained is very objective.

Though all above algorithms make use of the rough sets, the assigned weight isn’t still all-around, and sometimes isn’t line with the actual situation. With improving algorithm of document [16], in the paper the author puts forward one comprehensive attribute weighting algorithm, which consists of objective weight which gotten by the rough sets theory and subjective weight that the experts give by a priori knowledge, such can make sure to realize the unity of the subjective and objective weight, and improve the rationality of evaluation results. Experiment results also prove that the proposed algorithm is effective and rational.

THE BASIC ATTRIBUTE WEIGHT PRINCIPLE BASED ON THE ROUGH SETS THEORY

Each attribute has different effect on the decision results for one information system. As usual we assign one attribute importance depending on the increment after removing an attribute. If the increment is more, we think the attribute importance is bigger. The attribute weight method based on the rough sets theory is mainly through the attribute importance of decision table. The follow is the correlative definitions.

Definition 1: $S=(U,A,V,f)$ is set as an information system. Among them, $U=\{U_1,U_2,\cdots,U_{[U]}\}$ is non empty finite set which is called the domain space, $A=\{a_1,a_2,\cdots,a_{[A]}\}$ is non empty finite attribute set, which is called the attribute set. $V=\cup A_a, a\in A, V_a$ is attribute's domain range, $f:U\times A\rightarrow V_a$ is the information function. When x is a , x has unique value in V_a . On the other side, for sequence $C(c_1(x),c_2(x),\cdots,c_n(x))$ and sequence $D(d_1(x),d_2(x),\cdots,d_n(x))$, $A=C\cup D, C\cap D=\phi$,
$S=(U,A,V,f)$ is called as decision table of the information system, $c_1(x),c_2(x),\cdots,c_n(x)$ is called as the condition attribute set.

Definition 2: For the given knowledge representation system $S=(U,A,V,f)$,
the in-discernable relationship of any attribute is as follows:

$$IND(B)=\{(x,y)\in U\times U:\forall a\in B(f(x,a)=f(y,a)\} \quad (1)$$

Definition 3: For the decision table $S=(U,A,V,f)$, the degree that the condition attribute depends on the decision attribute is defined as follows:

$$\gamma_B(D)=|POS_B(D)|/|U| \quad (2)$$

Definition 4: For the decision table $S=(U,A,V,f)$, $c\in C$,the importance degree of condition attribute c is defined as:

$$SGF(c)=\gamma_B(D)-\gamma_{C-\{c\}}(D) \quad (3)$$

The weight of the condition attribute is defined as:

$$W(c)=\frac{SGF(c)}{\sum_{a\in C}SGF(a)} \quad (4)$$

According to the above formula (4) we can see that the larger $SGF(a)$ is, the more important the attribute is, so the attribute weight is much greater.

ATTRIBUTE WEIGHT BASED ON ROUGH SETS THEORY

Under the algebraic representation, many concepts of rough sets are not represented expressly, people are not easy to understand their content. The researchers utilize the relationship between knowledge and information entropy for rough sets theory, so the main concepts and operations of rough sets theory can be performed by the information theory.

Definition 5: U is set a domain , P and U is two equivalent relation of domain U (knowledge), $U/ind(P)=\{x_1,x_2,\cdots,x_n\}$, $U/ind(Q)=\{y_1,y_2,\cdots,y_n\}$, then the probability distribution that P and Q effect on the U is defined as follows:

$$[X : p] = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ p(x_1) & p(x_2) & \cdots & p(x_n) \end{bmatrix}; [Y : p] = \begin{bmatrix} y_1 & y_2 & \cdots & y_m \\ p(y_1) & p(y_2) & \cdots & p(y_m) \end{bmatrix} \quad (5)$$

Among them, $p(x_i) = \frac{|x_i|}{U}, i = 1,2,\cdots,n; p(y_i) = \frac{|y_j|}{U}, j = 1,2,\cdots,m;$ the symbol $|E|$ is the base of E .

Definition 6: According to the information theory, the information entropy of knowledge P is $H(P) = -\sum_{i=1}^{n} p(x_i)\log p(x_i)$,the conditional entropy $H(Q|P)$ of the knowledge P relative to Q is :

$$H(Q|P) = -\sum_{i=1}^{n} p(x_i)\sum_{j=1}^{m} p(y_j|x_i)\log p(y_j|x_i) \quad (6)$$

The mutual information $I(P;Q)$ of the knowledge P relative to Q is:

$$I(P;Q) = H(Q) - H(Q|P) \quad (7)$$

Definition 7: The attribute importance [12] is calculated by the increment of mutual information with subtracting one attribute, the definition is as follows:

$$SGF_{old}(c) = I(C;D) - I(\{C-c\};D) = H(D|C-\{c\}) - H(D|C) \quad (8)$$

The weight $W_o(c)$ of the condition attribute c is defined as:

$$W_{old}(c) = \frac{SGF_{new}(c)}{\sum_{a\in C} SGF_{new}(a)} \quad (9)$$

THE IMPROVED ATTRIBUTE WEIGHT ALGORITHM

Attribute importance in formula (8) only considers each attribute's effect on the whole decision, without considering attribute itself effect on the result of the decision. So the following consequence appear: the increment of the mutual information is more bigger, and the corresponding attribute weight is higher, so the attribute is considered as more important, but the actual situation is on the contrary, the attribute is maybe less important. So in the paper the author improves on the algorithm of article [16],and proposes one new importance algorithm with considering important degree of the condition attribute itself, the increment of the mutual information, as well as information entropy of attribute itself on the decision results , the define is as follows:

$$\begin{aligned} SGF_{new}'(c) &= (I(C-\{c\};D) - I(C;D) + I(c;D)) / H(D|c) \\ &= (H(D|C) - H(D|C-\{c\}) + H(D) - H(D|c)) / H(D|c) \\ &= (H(D|C) - H(D|C-\{c\}) + H(D)) / H(D|c) - 1 \end{aligned} \quad (10)$$

The weight is defined as:

$$W_o(c) = \frac{SGF_{new}'(c)}{\sum_{a\in C} SGF_{new}'(a)} \quad (11)$$

From the above formula (10), we can see that the corresponding attribute importance is much greater when the mutual information increment is equal and $H(D|c)$ is smaller, attribute importance is much bigger.

In addition, in order to make weight achieve the subjective and objective unity, and to avoid the objective weight disaccord with the actual situation, in this paper the author assigns weight from two sides named as the subjective weight and objective weight, the subjective weight is determined directly by the expert's experience, the objective weight is obtained by formula (10), the comprehensive weight is defined as the following:

$$W_z(c) = \alpha W_s(c) + (1-\alpha)W_o(c) \qquad (12)$$

Among them: $W_z(c)$ is the subjective weight and $W_o(c)$ is the objective weight, and α is called experience factor, which reflects the extent of the attribute importance, α is affected by the expert experience, the greater it is , the expert's subjective experience affects greatly the result of the decision, when $\alpha = 1$, $W_z(c)$ is completely decided by the expert subjective weight.

ALGORITHM ANALYSIS

In order to verify the effectiveness of the above algorithm, we take it to compute the reduction set of one command and information system, as shown in table 1. From it, we can see that the system has 4 attributes, 14 experts give those evaluation results, the condition attributes are $\{c_1, c_2, c_3, c_4\}$, decision attributes are $\{d\}$, the value of set C is set as V={0,1,2}, which corresponds good, general, poor state respectively. The value of D is set as {0, 1}, which corresponds good and bad for operational effect. After pretreatment on the original data, we get the decision table shown in table 1.

Table1 Decision table of information system

Expert samples	Communication quality c_1	System exchange quality c_2	Safety measure c_3	Personnel diathesis c_4	Decision result d
x1	0	0	0	0	0
x2	0	0	0	1	0
x3	1	0	0	0	1
x4	2	1	0	0	1
x5	2	2	1	0	1
x6	2	2	1	1	0
x7	1	2	1	1	1
x8	0	1	0	0	0
x9	0	2	1	0	1
x10	2	1	1	0	1
x11	0	1	1	1	1
x12	1	1	0	1	1
x13	1	0	1	0	1
x14	2	1	0	1	0

The below is the processing in accordance with the algorithm of section 3:

We consider table 1 as an information system $U = \{x1, x2, x3, x4, x5, x6, x7, x8, x9, x10, x11, x12, x13, x14\}$ condition attribute set is $C = \{c_1, c_2, c_3, c_4\}$, decision attribute set is $D = \{d\}$,

$IND(D) = \{\{x1, x2, x6, x8, x14\}, \{x3, x4, x5, x7, x9, x10, x11, x12, x13\}$.

The important degree of each attribute is calculated according to the formula (10), the objective weight are listed in table 2:

Table 2 The attribute object weight

Attribute	$H(D)$	$H(D\|C)$	$H(D\|\{C-c\})$	$H(D\|c)$	$SGF_{old}(c)$	$W_{old}(c)$	$SGF'_{new}(c)$	$W_o(c)$
c_1	0.9403	0	0.5714	0.6935	0.5714	0.6667	1.1798	0.5537
c_2	0.9403	0	0	0.6793	0	0	0.3842	0.1803
c_3	0.9403	0	0	0.7885	0	0	0.1925	0.0904
c_4	0.9403	0	0.2857	0.8922	0.2857	0.3333	0.3741	0.1756

From table 2, we can see that the weight of attribute c_1, c_2, c_3, c_4 obtained by article [16] were 0.6667,0, 0, 0.3333 respectively, and all of the weight of attribute c_2 and c_3 is 0, which is inconsistent with the actual situation. But the weight of attribute c_1, c_2, c_3, c_4 is obtained according to the algorithm proposed in this paper is 0.5537, 0.1803, 0.0904, 0.1756 respectively. The objective weight of attribute c2 isn't 0, and it is slightly higher than the objective weight of attribute c_4, this is consistent with the actual system. In addition, considering of subjective weights, we invite the experts in the field of command and information system to assign the subjective weight for four attribute c_1, c_2, c_3, c_4 as (0.35, 0.3, 0.15, 0.35) respectively. The experts set experience factor $\alpha = 0.4$ in considering of objective data, the comprehensive weight depending on the formula (11) is obtained as follows:

$$W_z(c1) = 0.4722 \quad W_z(c2) = 0.2281 \quad W_z(c3) = 0.1142 \quad W_z(c4) = 0.1855$$

Form the table 2, we can see that the attribute weight of c_2 and c_3 listed in $W_{old}(c)$ column is 0 based on the formula(8),which is not line with the actual situation.

In the actual command and information system, the quality of communication and interoperability play a decisive role on the operational effectiveness of the whole system. From the experiment results we can see that the weight value of c_1 and c_2 is the two largest in the all attributes, the weight of c_3 is the least in the all attributes, which is completely in line with the actual situation, so the experiment results verify that the proposed algorithm in this paper is rational and effective.

CONCLUSION

In view of the deficiency of the attribute weight method based on the rough sets theory, the author proposes one new comprehensive attribute weight algorithm，which contains two parts, one is the objective weight gotten by the mutual information based on the rough sets theory, the other is the subjective weight gotten by expert experience. The subjective weight can correct the deviation caused by the object weight. The objective weigh includes three factors, named as the importance of the attribute itself, the increment of mutual information, and its own information entropy change. Experiment results prove that the new proposed method completely avoids the weight 0, and it is more effective and reasonable than the existing algorithm. By the new algorithm the decision makers can make more effective decision and judgment.

Acknowledgements
This research is supported by State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System under Grants No CEMEE2014K0368B.

REFERENCES

[1] Pawlak Z. *International Journal of Computer and Information Sciences*, vol.11, no.3, pp. 341-356, **1982.**
[2] Pawlak Z. *Rough sets theoretical aspects of reasoning about data. Dordrecht: Kluwer Academic Publisher,* **1991.**
[3] Kryszkiewicz M. *Information Sciences*, vol.112, no.1, pp. 39-49, **1988.**
[4] Kim Y. M. , Kim, C. K. , Lee, J. C. . *Advances in Engineering Soft war e*, vol.40, no.3, pp. 202- 211, **2009.**
[5]Liu qing. *Rough Set and Rough reasoning[M],Beijing：Science Press*，**2001**.
[6]Liang jiye, Li deyu. *The uncertainty and knowledge acquisition in information system [M]. Beijing：Science Press*,**2005**.
[7] Zhang wenxiu, Wu weizhi. *The theory and method of Rough Sets theory[M]. Beijing：Science Press*,**2001**.
[8] Wu W Z, Mi J S, Zhang W X. *IEEE Computer society Press*, vol.3, no.3, pp. 1713 - 1718, November **2003.**
[9] Zhao Xu, Huang Yong-zhong, An Liu-yang. *Journal of Computer applications*, vol.32, no.10, pp. 167-169, **2012.**
[10] Zhang Wen-yu, Ma Yue, Chen Xing, Zhang Yu-Fei. *Measurement & Control technology* , vol. 32, no. 10, pp. 125-128, **2013.**
[11] Li, M . , Yu, B. , Rana, O, F. , et al. *IEEE Transactions on Knowledge and Data Engineering*, vol.20, no.6, pp.851-862, **2008.**
[12] Teoh H . J. , Cheng C. H . , Chu, H . H . , *et al. Data & Know ledge Engineering*, vol. 67, no.1, pp. 103-117, **2008.**
[13] Wang Hongkai, Yao Bingxue,Hu haiqing. *Computer engineering and application*, vol.67, no.3, pp. 20- 21, **2003.**
[14] Tan zongfeng, Xu zhangyan, Wangshuai. *Computer engineering and application*, vol. 48, no.18, pp. 115-118, **2012.**
[15] Miao Duoqian,Fan Shidong. *system engineering theory & practice*, vol. 22, no.1, pp. 48- 56, **2002.**
[16] Wang guoKuang. *Rough theory and knowledge acquisition[M] . Xi'an jiao tong university press*, **2001.**
[17] Zhong Jia-ming1,LI Ding- fang. *Computer Engineering and Applications*, vol.44, no.20, pp. 51-54, **2008.**
[18] Sun Limin, Jin Xiangju. *Computer Engineering and Applications*, vol. 49, no.2, pp. 51-54, **2013.**

[19]Yanyan,Yang huizhong. *tsinghua science and technology*, vol.47, no.S2, pp. 1903-1906, **2007.**
[20] Bao xinzhong,Liu cheng. *Journal of management*, vol. 6, no.6, pp. 729-732, **2009.**