

# An Approach for Attribute Weights Acquisition based on Rough Sets Theory and Information Gain

Dun Liu<sup>1</sup> Pei Hu<sup>1</sup> Tianrui Li<sup>2</sup> Chaozhe Jiang<sup>3</sup>

<sup>1</sup>School of Economics and Management, Southwest Jiaotong University, Chengdu 610031, P. R. China

<sup>2</sup>School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, P. R. China

<sup>3</sup>College of Traffic and Transportation, Southwest Jiaotong University, Chengdu 610031, P. R. China

## Abstract

The irrationality of constructing weights, lending to subjectivity and without considering the redundancy of attributes, exists in traditional management decision-making. The  $\sigma$ -important rating and  $\xi$ -important rating are first proposed by the attribute reduct in rough sets theory. Then, the  $g$ -important rating is given by the information gain in information entropy. An approach for acquiring attribute weights employing the three important ratings is presented to solve the problem of subjectivity and redundancy. An empirical case study validates the rationality and validity of our method.

**Keywords:** Rough sets theory, Information gain, Attributes reduct, Weight.

## 1. Introduction

In many multiple attribute decision making (MADM) problems, the best project should be chose from the candidates. In virtue of the different importance of these attributes in candidates, each attribute should be endowed with a different weight. The classical approaches select weights based on preference information of attributes given by the experts, including eigenvector method [1], weighted least square method [2], AHP method [3] and Delphi method [4], etc. Although decision making (DM) problem can be solved through these approaches, there are two serious limitations. One is the differences of the expert's preference lead to subjectivity and deviation, the other is not considering the redundancy of attributes.

Since Professor Pawlak of Poland proposed the Rough sets Theory (RST)[5], this methodology has been developed quickly in past two decades and used widely in many domains such as industry, iatrolgy and information science, etc. Because RST is based on original data, it can solve the problems of subjectivity and redundancy in DM, which may be used to construct weights in DM. In this paper, the attribute reduct in RST is used to obtain the important ratings

of attributes. Firstly, the core and unnecessary attributes in the information system is distinguished, which can estimate the importance relations among attributes to some extent. Then, the contribute rating of a certain attribute is defined according to whether the system changed or not after deleting the attribute. The contribute rating represents the influence of the attributes to maintain the stability of the whole system. The more important the rating is, the more weight we should give.

In addition, Shannon, an American mathematician, proposed the concept of entropy as an important tool to deal with the uncertain phenomena [6], which has been developed very quickly and applied widely in information and decision science. Moreover, information gain, an important concept based on entropy for classification and prediction in data mining domain, was chose as an effective measurement to eliminate the attribute which contains less information and reserve the attribute which contains more information. Because it can reflect the important rating of attributes and the intrinsic framework of system objectively, information gain can be used to construct weights in DM.

This paper introduces the two approaches to construct the weight at first. Then, the defects of these two methods are discussed and an approach for attribute weights acquisition based on RST and information gain is presented. Finally, the advantages of proposed weight are validated by an empirical case study.

## 2. Preliminaries

The basic concepts, notations and results of rough sets as well as their extensions are briefly reviewed.

**Definition 2.1**[7] An information system is defined as a pair  $S=(U,A)$  where  $U$  is a non-empty finite set of objects,  $A=C\cup D$  is a non-empty finite set of attributes,  $C$  denotes as the set of condition attributes and  $D$  denotes as the set of decision attributes,  $C\cap D\neq\emptyset$ . Each attribute  $a\in A$  is associated with a set  $V_a$  of its value, called the domain of  $a$ .

**Definition 2.2**[7] Given an information system  $S = (U, A)$ , we denote  $\gamma_C(D) = \sum_{i=1}^n |R_{-}X_i| / |U|$  as the quality of the approximation of  $D$  by means of the attributes from  $C$ . The important rating of  $a$  is defined as:  $\sigma_{CD}(a) = \gamma_C(D) - \gamma_{C-\{a\}}(D)$ .

**Definition 2.3**[7, 8] Given an information system  $S = (U, A)$ , we denote  $POS_C(D) = \bigcup_{X \in U/D} C_X$  as the positive region of  $C$ . For  $\forall a \in B_k \subseteq C$ , we have:

- (1) If  $POS_{C-\{a\}}(D) = POS_C(D)$ , then  $a$  is an unnecessary attribute of  $C$ .
- (2) If  $POS_{C-\{a\}}(D) \neq POS_C(D)$ , then  $a$  is a necessary attribute of  $C$ .
- (3) If  $POS_{B_k}(D) = POS_C(D)$ ,  $POS_{B_k-\{a\}}(D) \neq POS_C(D)$ , then we call  $B_k$  is a reduct of  $C$ .

**Definition 2.4**[8] Given an information system  $S = (U, A)$ , let  $B_k$  ( $k=1, 2, \dots, r$ ) are the whole attribute reduct sets, we denote  $M = \bigcap_{k \leq r} B_k$  as the core attribute sets;  $N = \bigcup_{k \leq r} B_k - M$  as the relatively necessary attribute sets;  $I = C - \bigcup_{k \leq r} B_k$  as the unnecessary attribute sets respectively.

### 3. The weight acquisition method based on RST

Firstly, the result of the lower approximation can describe the credible knowledge in information system, and the weight of an attribute can be estimated by the variety rating of the lower approximation when the attribute is deleted. The important rating defined by definition 2.2 can reflect this variety very well, so it can be used to construct the weight in DM.

Considering the important rating of definition 2.2, for the core attribute, the lower approximation must be changed when we delete it. If the number of core attributes is more than one, weights constructed by this method depend on the variety degree of the system. As for unnecessary attributes, the lower approximation will not be changed if it is deleted, which is suitable for the case. But for the relatively necessary attribute, only the dimensionality of reduct sets is changed but the elements in lower approximation are not changed when it is deleted, so the important rating of these attributes will be zero. In other words, the important rating of definition 2.2 merely can distinguish the core

attribute and unnecessary attribute while ignoring the effect of the relatively necessary attribute.

For a relatively necessary attribute  $a_i$ , we can easily obtain that it is belonged to some of the attribute reduct sets of  $B_k$  according to Definition 2.3 and Definition 2.4. We can use the cardinal number of  $a_i$  divide the cardinal number of  $B_k$  as a standard to define another important rating. In order to distinguish the two different important ratings, we denote  $\sigma$ -important rating in Definition 2.2 and  $\xi$ -important rating in the following.

**Definition 3.1**[9] Given an information system  $S = (U, A)$ , let  $B_k$  ( $k=1, 2, \dots, r$ ) are all the attribute reduct sets, for  $\forall a \in C$ , we denote the

$\xi$ -important rating as  $\xi_a = \sum_{k=1}^r \chi_{a \cap B_k} / r$ , where

$$\chi_{a \cap B_k} = \begin{cases} 1 & a \cap B_k \neq \emptyset \\ 0 & a \cap B_k = \emptyset \end{cases}, (k=1, 2, \dots, r).$$

**Proposition 3.1**[9] Given an information system  $S = (U, A)$ , for any  $a \in M$ ,  $b \in N$ ,  $c \in I$ , we have  $\xi_a = 1$ ,  $0 < \xi_b < 1$  and  $\xi_c = 0$ .

Thus we can get the  $\xi$ -important ratings of these three kind attributes, which rectifies the lack of Definition 2.2. But for the core attribute, we find an interesting phenomenon that the  $\xi$ -important ratings of core attribute is equal to 1 but the  $\sigma$ -important ratings may be different if the numbers of core attributes are more than one. Therefore, for a certain attribute  $a \in C$ , the process of weights acquisition based on RST is showed as the following.

(1) The  $\sigma$ -important ratings of the attribute  $a$  is obtained and normalized as following:

$$\sigma_{CD}^*(a) = \sigma_{CD}(a) / \sum_{a \in C} \sigma_{CD}(a);$$

(2) The  $\xi$ -important ratings of the attribute  $a$  is obtained and normalized as following:

$$\xi_a^* = \xi_a / \sum_{a \in C} \xi_a;$$

(3) The weight based on RST is defined as:  $\eta_a = \nu \sigma_{CD}^*(a) + (1-\nu) \xi_a^*$ , where  $\nu$  is a constant and  $\nu \in [0, 1]$ .

However, this method also has some shortages in practical applications shown in the following example.

**Example 3.1** Given a decision information system  $S = \{U, C, D\}$ ,  $U = \{1, 2, 3, 4, 5, 6, 7\}$ ,  $C = \{a_1, a_2, a_3, a_4\}$ ,  $D = DEC$ .

$U$	$a_1$	$a_2$	$a_3$	$a_4$	$DEC$
1	1	0	0	1	1
2	1	0	0	0	1
3	0	0	0	0	0
4	1	1	0	1	0
5	1	1	0	2	2
6	2	2	0	2	2
7	2	2	2	2	2

Table 3.1: A decision information system.

Since  $POS_{C-\{a_3\}}D = POS_C D$  and  $\forall a_i \in C \setminus \{a_3\}$ ,  $POS_{C \setminus \{a_3\} - \{a_i\}}D = POS_{C \setminus \{a_3\}}D$ , we obtain only one attribute reduct:  $\{a_1, a_2, a_4\}$ .

In addition,  $U/D = \{\{1, 2\}, \{3, 4\}, \{5, 6, 7\}\}$  and  $U/C = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{7\}\}$ ;  $U/\{a_1, a_2\} = \{\{1, 2\}, \{3\}, \{4, 5\}, \{6, 7\}\}$ ;  $U/\{a_1, a_4\} = \{\{1, 4\}, \{2\}, \{3\}, \{5\}, \{6, 7\}\}$ ;  $U/\{a_2, a_4\} = \{\{1\}, \{2, 3\}, \{4\}, \{5\}, \{6, 7\}\}$ . So we have  $POS_{C \setminus \{a_3\} - \{a_1\}}D = \{5, 6, 7\}$ ;  $POS_{C \setminus \{a_3\} - \{a_2\}}D = \{5, 6, 7\}$ ,  $POS_{C \setminus \{a_3\} - \{a_4\}}D = \{1, 2, 3\}$ .

By the definition of  $\sigma$ -important ratings and  $\xi$ -important ratings, we can get:  $\sigma_{CD}(a_1) = \sigma_{CD}(a_2) = \sigma_{CD}(a_4) = 4/7$ ,  $\sigma_{CD}(a_3) = 0$ ;  $\xi_{a_1} = \xi_{a_2} = \xi_{a_4} = 1$ ,  $\xi_{a_3} = 0$ . The weights based on RST are:  $\eta_{a_1} = \eta_{a_2} = \eta_{a_4} = 1/3$ ,  $\eta_{a_3} = 0$ .

This example shows that the weight of the element out of the attribute reduct sets will be zero; the weight of the element in the attribute reduct sets will be the same. Since the numbers of the attribute reduct sets are limited, some of the attributes may have the same weight in DM and all the attributes can't form a whole order sequence, which will affect the efficiency of the final decision making, so this approach for weight acquisition must be improved.

#### 4. The weight acquisition method based on Information gain

**Definition 4.1**[10] Supposed  $U$  is a finite set of reference actions, the universe  $U$  is divided into  $m$  classes by the decision attribute  $D$ ; For a certain

$D_i, i \in [1, 2, \dots, m]$ ,  $d_i$  is remarked as the cardinal number of  $D_i$ , Let  $\frac{d_i}{|U|}$  be the probability that an arbitrary tuple in  $U$  belongs to class  $D_i$  ( $|U|$  means the cardinal number of  $U$ ). So we can define the information entropy produced by the class  $D_i$  is:

$$I(d_1, d_2, \dots, d_m) = -\sum_{i=1}^m \left(\frac{d_i}{|U|}\right) \log_2 \left(\frac{d_i}{|U|}\right)$$

If an attribute  $a$  with values  $\{a_1, a_2, \dots, a_v\}$  is used as the test at the root of the decision tree,  $U$  is partitioned into the subsets  $\{u_1, u_2, \dots, u_v\}$ , where  $u_j$  contains those objects in  $U$  that have value  $a_j$  of  $a$ . Let  $u_j$  contain  $u_{ij}$  objects of class  $D_i$ . The expected information based on this partitioning by  $a$  is known as the entropy of  $a$ . It is the weighted average.

$$E(a) = \sum_{j=1}^v \frac{u_{1j} + u_{2j} + \dots + u_{mj}}{|U|} I(u_{1j}, \dots, u_{mj})$$

The information gain of  $a$  is defined by:

$$Gain(a) = I(d_1, d_2, \dots, d_m) - E(a)$$

In view of the size of information gain is the basis when people choose the node of the decision tree and the attribute with the highest information gain is considered the most discriminating attribute of the given set, we can use information gain as a method to construct the weight of attributes. We denote this weight as  $g$ -important ratings and the  $g$ -important ratings of attribute  $a$  is defined as:  $g_a^* = Gain(a) / \sum_{a \in C} Gain(a)$ .

Although using information gain as a measure to select attributes is extensively used in data mining, a threshold value is usually used to eliminate the independent or weak relative attribute when doing the relativity analyze by using the method of information gain [10], which may be subjectivity. In addition, we find an interesting phenomenon that if the information gain of a certain attribute has a close relation corresponding to the structure of decision attribute; it may mislead us that validated by the following example.

**Example 4.1** Given two decision information systems in the following, the condition attributes  $C$  in the left table are {headache, muscle pain, animal heat}, the condition attributes  $C'$  in the right table are {headache, snivel, animal heat}, and the decision

attribute in both of the two tables are whether get flu or not.

Patient	C			D
	H	M	A	Flu
X1	Yes	A little	Normal	No
X2	Yes	Very	High	Yes
X3	Yes	A little	Very High	Yes
X4	No	Very	Normal	No
X5	No	No	High	No
X6	No	Slightly	Very High	Yes
X7	No	No	High	Yes
X8	No	Slightly	Very High	No

Patient	C'			D
	H	S	A	Flu
X1	Yes	A little	Normal	No
X2	Yes	Very	High	Yes
X3	Yes	Slightly	Very High	Yes
X4	No	No	Normal	No
X5	No	Very	High	No
X6	No	A little	Very High	Yes
X7	No	No	High	Yes
X8	No	Slightly	Very High	No

Table 4.1: The two tables based on whether to get a flu or not.

where, H, M, S, A denote headache, muscle pain, snivel, animal heat, respectively.

Let  $s_1 = \{\text{Flu, Yes}\}$ ;  $s_2 = \{\text{Flu, No}\}$ .

$$I(s_1, s_2) = -\sum_{i=1}^2 \frac{|s_i|}{|U|} \log_2 \left( \frac{|s_i|}{|U|} \right)$$

$$= -\frac{1}{2} \log_2 \left( \frac{1}{2} \right) - \frac{1}{2} \log_2 \left( \frac{1}{2} \right) = 1$$

For the first table:

$$E(\text{muscle pain}) = 4 * \frac{1}{4} \left( -\frac{1}{2} \log_2 \left( \frac{1}{2} \right) - \frac{1}{2} \log_2 \left( \frac{1}{2} \right) \right) = 1$$

For the second table:

$$E(\text{snivel}) = 4 * \frac{1}{4} \left( -\frac{1}{2} \log_2 \left( \frac{1}{2} \right) - \frac{1}{2} \log_2 \left( \frac{1}{2} \right) \right) = 1$$

So, we can get:

$$\text{Gain}(\text{muscle pain}) = \text{Gain}(\text{snivel}) = 0$$

This example shows that the attributes of muscle pain in the first table and snivel in the second table would be eliminated according to the method of information gain. However, the attribute of snivel should be an important evidence to be used to judge a person whether gets flu or not, it is obviously unsuitable if it is deleted. Therefore, the method of

weights acquisition based on information gain also should be improved.

## 5. The weight acquisition method based on RST and Information Gain

This section presents the weight acquisition method based on RST and Information Gain. We first review the two examples in section 3 and 4.

For Example 3.1, we calculate the information gain of each attribute as following:

$$\text{Gain}(a_1) = 0.6995, \text{Gain}(a_2) = 0.8774,$$

$$\text{Gain}(a_3) = 0.1981, \text{Gain}(a_4) = 0.9852$$

We can see that the weight of the four attributes form an order sequence by using the method of information gain.

For Example 4.1, the  $\eta$ -weight of muscle pain and snivel is calculated respectively as following:

$$U/D = \{\{x_1, x_4, x_5, x_8\}, \{x_2, x_3, x_6, x_7\}\}$$

$$U/C = \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_6, x_8\}, \{x_5, x_7\}\}$$

$$U/(C - \{\text{muscle pain}\}) = \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_6, x_8\}, \{x_5, x_7\}\}.$$

In virtue of  $POS_{C - \{\text{muscle pain}\}} D = POS_C D$ , that is, the attribute of muscle pain is the redundancy attribute,  $\eta\{\text{muscle pain}\} = 0$ .

$$U/D = \{\{x_1, x_4, x_5, x_8\}, \{x_2, x_3, x_6, x_7\}\}$$

$$U/C' = \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5\}, \{x_6\}, \{x_7\}, \{x_8\}\}.$$

$$U/(C' - \{\text{snivel}\}) = \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_6, x_8\}, \{x_5, x_7\}\}.$$

In virtue of  $POS_{C - \{\text{snivel}\}} D \neq POS_C D$ , that is, the attribute of snivel is not the redundancy attribute,  $\eta\{\text{snivel}\} \neq 0$ .

Therefore, we consider constructing weights by employing the two approaches together. For a certain attribute  $a \in C$ , we define the final weight of  $a$  as follows.

$$w_a = \lambda \sigma_{CD}^*(a) + \mu \xi_a^* + (1 - \lambda - \mu) g_a^* \cdot \chi_a,$$

where  $\lambda, \mu \in [0, 1]$  and  $\lambda, \mu$  is a constant,

$$\chi_a = \begin{cases} 1 & \text{when } a \text{ is not a redundancy attribute} \\ 0 & \text{when } a \text{ is a redundancy attribute} \end{cases}.$$

When  $a$  is the redundancy attribute, we have  $w_a = 0$ . At last, we normalize the final weight as following:  $w_a^* = w_a / \sum_{a \in C} w_a$ . Generally, we let  $\lambda = 1/3$ ,  $g = 1/3$ ,  $w_a = [\sigma_{CD}^*(a) + \xi_a^* + g_a^*] / 3$ .

In addition, we obtain two propositions according to Examples 3.1 and 4.1:

**Proposition 5.1:**  $a$  is a redundancy attribute, which is not equal to  $Gain(a) = 0$ .

**Proposition 5.2:** For a certain condition attribute  $a_j$ , if  $I(d_1, d_2, \dots, d_m) = I(u_{1j}, \dots, u_{mj})$ , then  $Gain(a_j) = 0$ .

## 6. Empirical study

A case about a corporation evaluating employees for a job is studied. The director of human resource department gives a test to the 10 position candidates by choosing the skill of mathematical, computer, team cooperate, space imagination, language expression, professional knowledge, foreign language as the considerable criterions, which are denoted as  $a_1, a_2, a_3, a_4, a_5, a_6$  and  $a_7$ , respectively. Then, we let  $V_C = \{A, B, C, D, E, F\}$  represent {excellent, good, commonly, bad, very bad}, respectively, corresponding to certain skills of the candidates. Let  $V_D = \{1, 2, 3, 4\}$  which means the general diathesis of the position candidate is excellent, good, commonly, bad, respectively. The detailed information is showed in Table 6.1.

Position candidates	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	Dec
1	B	A	B	B	A	B	A	1
2	E	D	E	D	D	C	C	4
3	E	C	E	D	C	D	C	4
4	D	C	D	C	B	A	B	3
5	B	A	B	B	A	B	A	1
6	D	C	D	C	B	A	B	3
7	E	D	E	D	C	D	C	4
8	E	C	C	D	C	C	C	3
9	C	B	C	B	A	B	A	2
10	D	B	C	C	B	A	A	2

Table 6.1: The decision table of the 10 position candidates.

Let  $U = \{x_1, x_2, \dots, x_{10}\}$ ,  $C = \{a_1, a_2, a_3, a_4, a_5\}$ , then we have:

$$U/C = \{\{x_1, x_5\}, \{x_4, x_6\}, \{x_2\}, \{x_3\}, \{x_7\}, \{x_8\}, \{x_9\}, \{x_{10}\}\}$$

$$U/D = \{\{x_1, x_5\}, \{x_9, x_{10}\}, \{x_4, x_6, x_8\}, \{x_2, x_3, x_7\}\}$$

$$\text{Since } \forall a_i \in C, POS_{C-\{a_i\}}D = POS_C D,$$

$$\sigma_{CD}(a_i) = \gamma_C(D) - \gamma_{C-\{a_i\}}(D) = 0, i = 1, 2, \dots, 7.$$

The system doesn't have the core attribute.

We use the software RSES produced by Warsaw University in Poland to calculate the whole attribute reduct sets as following:

$$B_1 = \{a_1, a_3\}, B_2 = \{a_2, a_3\}, B_3 = \{a_2, a_6\},$$

$$B_4 = \{a_3, a_4\}, B_5 = \{a_3, a_5\}, B_6 = \{a_3, a_6\},$$

$$B_7 = \{a_3, a_7\}, B_8 = \{a_1, a_5, a_6, a_7\}$$

From Definition 2.4, we can get 8 attribute reduct sets. The core attribute sets  $M = \bigcap_{k \leq 8} B_k = \emptyset$ .

The relatively necessary attribute sets  $N = \bigcup_{k \leq 8} B_k - M = U$ .

The unnecessary attribute sets  $I = C - \bigcup_{k \leq 8} B_k = \emptyset$ .

From Definition 3.1,  $\forall a_i \in C$ ,

$$\xi_{a_i} = \sum_{k=1}^r \chi_{a_i \cap B_k} / r.$$

$$\text{So } \xi_{a_1} = 2/8 = 0.25; \xi_{a_2} = 2/8 = 0.25;$$

$$\xi_{a_3} = 7/8 = 0.875; \xi_{a_4} = 1/8 = 0.125;$$

$$\xi_{a_5} = 2/8 = 0.25; \xi_{a_6} = 3/8 = 0.375;$$

$$\xi_{a_7} = 2/8 = 0.25.$$

Then, we have

$$\xi_{a_1}^* = \xi_{a_2}^* = \xi_{a_5}^* = \xi_{a_7}^* = 0.1053;$$

$$\xi_{a_3}^* = 0.3684; \xi_{a_4}^* = 0.0526; \xi_{a_6}^* = 0.1579.$$

From Definition 4.1, we denote  $d_1 = 1, d_2 = 2,$

$d_3 = 3, d_4 = 4$  for the domain of decision attribute.

Then we can calculate the information entropy and information gain of the 7 condition attributes in Table 6.1. The results are showed in Table 6.2.

Therefore, we have  $g_{a_1}^* = 0.1329, g_{a_2}^* = 0.1868,$   
 $g_{a_3}^* = 0.1924, g_{a_4}^* = 0.0789, g_{a_5}^* = 0.1299,$   
 $g_{a_6}^* = 0.1377, g_{a_7}^* = 0.1414.$

Since the system doesn't have core attribute, we do not consider the effect of  $\sigma$  - important ratings. So,

we choose  $\lambda = 0$  and  $\mu = 1/2$ , the final weight acquisition and sequence are showed in Table 6.3.

	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$
$I(d_1, d_2, d_3, d_4)$	1.971	1.971	1.971	1.971	1.971	1.971	1.971
$E(\square)$	0.8	0.325	0.275	1.275	0.826	0.75	0.725
$G(\square)$	1.171	1.646	1.695	0.695	1.114	1.221	1.246

Table 6.2: The information entropy and information gain of the 7 condition attributes.

	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$
$\xi_a^*$	0.1053	0.1053	0.3684	0.0526	0.1053	0.1579	0.1053
Rank	3	3	1	7	3	2	3
$g_a^*$	0.1329	0.1868	0.1924	0.0789	0.1299	0.1377	0.1414
Rank	5	2	1	7	6	3	4
$w_a^*$	0.1191	0.1461	0.2804	0.0658	0.1176	0.1478	0.1234
Rank	5	3	1	7	6	2	4

Table 6.3: The weight acquisition and sequence of the 7 condition attributes.

Therefore, we obtain the order sequence of these 7 condition attributes decreasingly: the skill of team cooperate, professional knowledge, computer, foreign language, mathematical, language expression and space imagination, which is consistent with the human's experience. In addition, the process of the weight acquisition is based on the data from the information system and don't add any expert's preference, which makes this approach more reasonable in real applications.

## 7. Conclusions

The methods of RST and information gain are employed for weight acquisition in DM in this paper. The advantage of this method is that it eliminates the personal subjectivity as and deals with the redundancy of attributes properly. A case study validates the rationality and validity of our method. Future research work of our research is to study the attribute weight acquisition in dynamic DM environment. It also seems worthwhile to explore if the proposed approach can be extended to other generalized rough set models such as fuzzy rough set theory.

## Acknowledgement

This work is partially supported by National Nature Science Foundation of China (60474022) and the MEPRC Doctoral Science Research Fund Foundation of China (20060613019).

## References

- [1] T.L. Saaty, A scaling method for priorities in hierarchical structures, *Journal of Mathematical Psychology*, 15: 234-281,1977.
- [2] A.T.W. Chu, R.E. Kalaba, K. Spingarn, A comparison of two methods for determining the weights of belonging to fuzzy sets, *Journal of Optimisation Theory and Application*, 27: 531-538, 1979.
- [3] T. Chen, Decision-making analysis, *The Press of Science*, 1997.
- [4] C.L. Hwang, M.J. Lin, *Group Decision Making under Multiple Criteria: Methods and Applications*, Springer-verlag, 1987.
- [5] Z. Pawlak. Rough Sets . *International Journal of Computer and Information Sciences*, 11: 341-356, 1982.
- [6] Shannon. A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27: 379-423, 623-656,1948.
- [7] W.X. Zhang, W.Z. Wu. Rough sets Theory and methodology. *The Press of Science*, 2001
- [8] W. X. Zhang. Uncertain decision making based on Rough Sets. *The Press of tsinghua University*, 2005.
- [9] D. Liu, P. Hu. A new Hierarchical Cluster Analysis methodology based on Rough Sets Theory and its application. *Statistic and Decision*, 236: 26-28,2007.

- [10] J. Han, M. Kamber. Data Mining: Concepts and techniques (2<sup>nd</sup> Edition), *China Machine Press*, 2006.
- [11] J. Wu, C. Y. Liang, W.N. Li. Method to determine attribute weights based on subjective and object integrated, *System Engineering and Electronics*, 29: 383-387, 2007.
- [12] L.E. Shirland, R.R. Jesse. Determining attribute weights using mathematical programming, *The International Journal of Management Science*. 31: 423-437, 2003.
- [13] X.Z. Xu. Note on the subjective and objective integrated approach to determine attribute weights, *European Journal of Operational Research*, 156: 530-532, 2004.
- [14] B.S. Ahn, Comparing methods for multiattribute decision making with ordinal weights, *Computer & Operations Research*, In Press.
- [15] Y.C. Hu. Assessing weights of product attributes from fuzzy knowledge in a dynamic environment. *European Journal of Operational Research*, 154: 125-143, 2004.
- [16] Z. Pawlak, Rough set theory and its application to data analysis. *Cybernetics and Systems*. 29: 661 - 688, 1998.