

A New Method of Qualitative Attributes estimation based on Fuzzy Rough Set

Xiaohui Zeng

Chengdu University of Information Technology,
Chengdu, 610225, P. R. China
zxhui@cuit.edu.cn

Huanglin Zeng

Sichuan University of Science and Technology,
Zigong, Sichuan, 643000, PR China
zhl@suse.edu.cn

Abstract—In the paper, a new method of convenient estimation for qualitative attributes is proposed based on fuzzy sets integrated with rough set. Firstly objects with some qualitative attributes are analyzed in an information system from a view of fuzzy set, and then the concepts of a new membership function of qualitative attributes are defined to express the dependency of condition attributes with respect to decision attribute. Finally a simply example is employed to validate the reasonability and the merits of the proposed method.

Keywords—Fuzzy sets, rough set, membership function, dependency

I. INTRODUCTION

It is very important for comprehensive estimation of qualitative attributes in an information system however there are many issues that deal with weights of qualitative attributes. Although expert grading, fuzzy statistic, two-element comparing sequencing and so on [1-3] have been studied, there are still some problems in work that the ability of classification and the relationship of condition attributes with decision attribute is not taken into account for accurate researches [2-5]. The weight of qualitative attributes in an information system can be defined with membership degree of condition attributes with respect to decision attribute, but membership function is normally complex mathematical model. Therefore, we put forward a new membership function of qualitative attributes from a view of fuzzy sets integrated with rough set in this paper. The weight of qualitative attributes in an information system is defined from a view of fuzzy comprehensive estimation. And the fuzzy membership degree of condition attributes with respect to decision attribute is taken into account by dealing with the maximum-optimum attributes or minimum-optimum attributes.

II. BASIC PRINCIPLES

$S = (U, A, V, f)$ is a collected database of personal credit evaluation system, where U is a non-empty universe defined on a finite set; for any object $x \in U$, $card(U) = |U|$ is cardinal number of some sets of x in U . $A = C \cup D$ is a non-empty set where C and D respectively are set of condition attributes and decision attribute contained credit degree.

Where $U|C = (X_1, \dots, X_n)$ denotes the equivalence classes of U classified on the set of condition attributes C , where X_j denotes an equivalence class based on condition attributes, and $U|D = (Y_1, \dots, Y_m)$ denotes the equivalence classes of U classified on the set of decision D , where Y_i denotes an equivalence class based on decision attributes, and $U|R_i$ denotes the equivalence classes of U classified on a qualitative attribute $R_i \in C, i = 1, \dots, n$. V is a set of attribute values. $f : U \times A \rightarrow V$ is an information function.

Definition 1: In a decision information system $S = (U, A, V, f)$, based on rough set [8], a lower approximation set is defined as following

$$P_-(D) = \cup \{X_j \in U | ind(C) : X_j \subseteq Y\} \quad (1)$$

Moreover, a positive region of input vector with respect to output pattern classification set is defined as following

$$POS_C(D) = \cup P_-(D) \quad (2)$$

Definition 2: In a decision information system $S = (U, A, V, f)$, based on rough set [8], the performance of information dependence relationships between condition attribute $R_i \in C, i = 1, \dots, n$ and decision D is defined as following

$$k = \frac{|POS_{R_i}(D)|}{|U|} \quad (3)$$

where $||$ represents the cardinal of a set.

Definition 3: For a qualitative attribute R_i , a group of an order couple is defined as follows:

$$A = \{(x, \mu_A(x)) | x \in U\} \quad (4)$$

where $\mu_A(x) \in [0, 1]$ is referred as membership function of qualitative attribute R_i .

And several kinds of common membership function form are introduced as below:

(1) Triangle membership function:

$$\mu_A(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ \frac{c-x}{c-b}, & b \leq x \leq c \\ 0, & c \leq x \end{cases} \quad (5)$$

Another kind of expression form is as follows:

$$\mu_A(x) = \max(\min(\frac{x-a}{b-a}, \frac{c-x}{c-b}, 0)) \quad (6)$$

(2) Trapezoid membership function:

$$\mu_A(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b \leq x \leq c \\ \frac{d-x}{d-c}, & c \leq x \leq d \\ 0, & d \leq x \end{cases} \quad (7)$$

Another kind of expression form is as follows:

$$\mu_A(x) = \max(\min(\frac{x-a}{b-a}, 1, \frac{d-x}{d-c}, 0)) \quad (8)$$

(3) Gaussian membership function:

$$\mu_A(x) = \exp(-\frac{1}{2}(\frac{x-c}{\sigma})^2) \quad (9)$$

(4) Cauchy membership function:

$$\mu_A(x) = \frac{1}{1 + |\frac{x-c}{a}|^{2b}} \quad (10)$$

(5) Sigmoid membership function:

$$\mu_A(x) = \frac{1}{1 + \exp(-a(x-c))} \quad (11)$$

III. 3 A NEW MEMBERSHIP FUNCTION OF QUALITATIVE ATTRIBUTES BASED ON FUZZY ROUGH SET

In order to quantitatively describe something and translate qualitative attributes into quantitative attributes, the definition of fuzzy optimal membership degree is introduced based on fuzzy rough sets:

$$\mu_A(x) = \frac{|[x]_{R_i} \cap A|}{|[x]_{R_i}|} \quad (11)$$

Where $[x]_{R_i}$ is denoted by $U|R_i$ the equivalence classes of U classified on a qualitative attribute $R_i \in C, i = 1, \dots, n$.

According to definition above, properties of a membership function $\mu_A(x)$ of a qualitative attribute is introduced as below:

- 1) $\mu_U(x) = 1$
- 2) $\mu_\Phi(x) = 0$
- 3) $y \in [x]_R, \mu_A(x) = \mu_A(y)$
- 4) $x \in A, \mu_A(x) \neq 0$
- 5) $\mu_A(x) = 1 \Rightarrow x \in A$
- 6) $A \subseteq B, \mu_A(x) \leq \mu_B(x)$

Here $\mu_A(x)$ meets with the following operation relations:

- 1) $\mu_{\bar{A}}(x) = 1 - \mu_A(x)$
- 2) $\max(0, \mu_A(x) + \mu_B(x) - 1) \leq \mu_{A \cap B}(x) \leq \min(\mu_A(x), \mu_B(x))$
- 3) $\mu_{A \cup B}(x) = (\mu_A(x) + \mu_B(x)) - \mu_{A \cap B}(x)$
- 4) $\max(\mu_A(x), \mu_B(x)) \leq \mu_{A \cup B}(x) \leq \min(1, \mu_A(x) + \mu_B(x))$

The concrete steps performing comprehensive estimation for qualitative attributes are as follows:

Step 1: The problem is expressed as a knowledge system that is a list of attribution value of knowledge representation in which data will be carried out as discrete-time data.

Step 2: calculate $U|R_i$ which denotes the equivalence classes of U classified on a qualitative attribute $R_i \in C, i = 1, \dots, n$.

Step 3: calculate membership function of qualitative attribute based on formula (11).

The effectiveness of the method proposed here is demonstrated by illustration with an example of an automobile knowledge expression system.

Where a, b, c represent size, fuel, color of cars as condition attributes set $C = \{a, b, c\}$; d, e represent velocity, acceleration as decision attributes set $D = \{d, e\}$ respectively.

Because of $U = \{1, 2, 3, 4, 5, 6, 7, 8\}$ and different condition attribute values in each object, we can divide them into unrelated subsets or equivalence classes, that are

$$\begin{aligned} U|a &= \{x_1, x_2, x_3\} = \{(1, 4, 5), (2, 8), (3, 6, 7)\} \\ U|b &= \{(1, 3, 5), (2, 4, 7, 8), (6)\} \\ U|c &= \{(1, 5), (2, 7, 8), (3, 4, 6)\}. \end{aligned}$$

In the same way, on account of their different decision attribute values $D = \{d, e\}$, we can divide them into 4 equivalence classes

$$U|D = \{W_1, W_2, W_3, W_4\} = \{(1, 4, 5), (2, 7), (3, 6), (8)\}.$$

For condition attribute a :

$$\mu_{w_1}(x_1) = 1, \mu_{w_2}(x_1) = 0, \mu_{w_3}(x_1) = 0, \mu_{w_4}(x_1) = 0;$$

$$\mu_{w_1}(x_2) = 0, \mu_{w_2}(x_2) = \frac{1}{2}, \mu_{w_3}(x_2) = 0, \mu_{w_4}(x_2) = \frac{1}{2};$$

$$\mu_{w_1}(x_3) = 0, \mu_{w_2}(x_3) = \frac{1}{3}, \mu_{w_3}(x_3) = \frac{2}{3}, \mu_{w_4}(x_3) = 0;$$

For condition attribute b :

$$\mu_{w_1}(x_1) = \frac{2}{3}, \mu_{w_2}(x_1) = 0, \mu_{w_3}(x_1) = \frac{1}{3}, \mu_{w_4}(x_1) = 0;$$

$$\mu_{w_1}(x_2) = \frac{1}{4}, \mu_{w_2}(x_2) = \frac{1}{2}, \mu_{w_3}(x_2) = 0, \mu_{w_4}(x_2) = \frac{1}{4};$$

$$\mu_{w_1}(x_3) = 0, \mu_{w_2}(x_3) = 0, \mu_{w_3}(x_3) = 1, \mu_{w_4}(x_3) = 0;$$

For condition attribute c :

$$\mu_{w_1}(x_1) = 1, \mu_{w_2}(x_1) = 0, \mu_{w_3}(x_1) = 0, \mu_{w_4}(x_1) = 0;$$

$$\mu_{w_1}(x_2) = 0, \mu_{w_2}(x_2) = \frac{2}{3}, \mu_{w_3}(x_2) = 0, \mu_{w_4}(x_2) = \frac{1}{3};$$

$$\mu_{w_1}(x_3) = \frac{1}{3}, \mu_{w_2}(x_3) = 0, \mu_{w_3}(x_3) = \frac{2}{3}, \mu_{w_4}(x_3) = 0.$$

From the calculated result above, we know that $\mu_A(x)$ is represented with the performance of information dependence which is interrelated by $POS_{R_i}(D)$, moreover, the rough member function actually is similar to conditional probability[8]. For example, for an automobile with medium velocity and excellent acceleration, it can be 100% decided by condition attributes as medium size, but diesel fuel and gray color has none business with the decision result. For a same automobile with medium velocity and excellent acceleration, it can be 2/3 decided by condition attribute fuel as diesel and 1/4 by condition attribute fuel as gas, and never by color. In this example, of course, we use a very simply data table to discuss attribute dependency between condition attribute with decision attribute, but it's necessary to know that from

a view of fuzzy membership degree, comprehensive estimation of any examples can be gained.

IV. SUMMERIES

In this paper, the significance of equivalence class with qualitative attributes in an information system is studied based on the quality of approximation classification. The quality of approximation classification of qualitative attributes is taken into account by a method of simply estimation in which is computed by way of the indiscernibility of rough set and the degree of fuzzy optimal membership. The whole process provides a new view to discuss the relationship of condition attributes with respect to decision attribute in fuzzy sets integrated the approximation classification in rough set.

ACKNOWLEDGEMENTS

This work was supported by a grant from Foundation of major special project of key laboratory of liquor and biotechnology of Sichuan province under No. NJ2010-01, and a Foundation of major applied science project of Science and Technology Department of Sichuan Province under No. 2011JY0051.

REFERENCES

- [1] Pawlak, Z.. Some issues on rough sets. Swiniarski, R.W., Szczuka, M.S., eds.: Transactions on Rough Sets I. LNCS 3100. Springer-Verlag, Berlin. PP.1-58, 2004
- [2] Yang Lunbiao, Gao Yingyi. Theory and Application of Fuzzy Maths (in Chinese). The Press of South China University of technology. PP.95-100, 2001
- [3] Huanglin Zeng, Hengyou Lan, Zeng Xiaohui. Redundant Data Processing Based on Rough-Fuzzy Approach. Rough Sets and Knowledge Technology [M], Springer-Verlag Berlin Heidelberg, Germany. PP.156-161, 2006
- [4] Qing Shen. Attribute Reduction of Multi-valued Information System Based on Conditional Information Entropy. Proceedings of IEEE International Conference on Granular Computing, China. PP.562-566, 2008
- [5] Sankar K Pal. Rough- Fuzzy Granular Computing with Application to Bioinformatics and Web Intelligence. Proceedings of IEEE International Conference on Granular Computing, China, PP.14-16, 2008
- [6] Huanglin Zeng, Xiaohui Zeng. Studies on consistence of the knowledge simplification of an uncertain system. Proceedings of 8th International Conference on Cognitive Informatics, HongKong. PP.162-166, 2009
- [7] Huanglin Zeng, Xiaohui Zeng. Reasoning Decision Rules of an Uncertain System. Proceedings of Proceedings of the Fourth International Conference on Rough Sets and Knowledge Technology, Australia, Springer-Verlag Berlin Heidelberg. PP.634-642, 2009
- [8] Huanglin Zeng, Intelligent Computing, Theoretical Aspects and Applications of Rough Sets, Fuzzy logic and Neutral Networks [M]. (in Chinese), ChongQing: University Press of ChongQing, 2004

TABLE I. DATA OF AN AUTOMOBILE KNOWLEDGE EXPRESSION SYSTEM

Cars	a	b	c	d	e
1	medium	diesel	gray	medium	excellent
2	small	gas	white	high	excellent
3	big	diesel	black	high	good
4	medium	gas	black	medium	excellent
5	medium	diesel	gray	medium	excellent
6	big	propane	black	high	good
7	big	gas	white	high	excellent
8	small	gas	white	medium	good