

Optimization of Research on Case-base Structure Based on Information Entropy

Bin HUANG

Academy of Armored Force Engineering
Beijing, China
e-mail: sddy1990@yeah.net

Hao LI

Academy of Armored Force Engineering
Beijing, China
e-mail: yyyy_2010@yeah.net

Abstract—The core step of Case-based Reasoning is to retrieval and match the case with the fault feature accurately and efficiently, the scale of the case library is becoming larger because of the case studies feature, but the redundancy is becoming clearer at the same time, it influences the accuracy. The effective way to control the redundancy is to define the attributes which can distinguish the cases, the paper determines the weight of the case characteristics with the theory of entropy, making it more reasonable to divide the characteristics and more effective to organize the case, in order to control the redundancy of the case-base, thus improves the effectiveness of Case-Based Reasoning.

Keywords—case-base management; the redundancy control; entropy; weights

I. INTRODUCTION

As the core step of Case-Based Reasoning, the efficiency of case-retrieval and match is the important criterion to measure the reasonableness of case base, as the core of case base management---the redundancy control decides the case-retrieval and match effectiveness at a certain degree. There is not so many studies about the case-base management, Jian-Yang Li [1] achieves the redundancy control with the theory of similar rough set, and it deletes the case in the case of comparing the case features in order to achieve the management of the case-base; Si-Ya Pang [2] achieves the case-base management with the SOM clustering, but the object structure is relatively simple. The paper achieves the case characteristics division effectively and the reasonable management of case-base, controlling the redundancy better, laying the foundation for the management of case-base.

II. RELATIVE THEORY

A. Entropy

In Probability approximation space (U, X, P) ,
 $U = \left\{ \begin{matrix} X_1, X_2, \dots, X_n \\ p_1, p_2, \dots, p_n \end{matrix} \right\}$, which means that
 X_1, X_2, \dots, X_n is one way to divide the U ,
 $P(X_i) = p_i, \sum_{i=1}^n p_i = 1$, Uncertainty of the division can be represented by the entropy of X_i , so
 P_i is on behalf of the ratio between the cardinal number

of X_i and the cardinal number of domains.

$$H(X_i) = -\sum_{i=1}^n p_i \log_2 p_i \quad (1)$$

B. Conditional Entropy

Assume that $U = \left\{ \begin{matrix} Y_1, Y_2, \dots, Y_m \\ q_1, q_2, \dots, q_m \end{matrix} \right\}$ is another source of information, which means that Y_1, Y_2, \dots, Y_m is another division of U , $P(Y_j) = q_j, \sum_{j=1}^m q_j = 1$, so the dependence degree about X_i and Y_j can be represented by conditional entropy, so

$$H(Y_j | X_i) = -\sum_{j=1}^m P(Y_j | X_i) \log_2 P(Y_j | X_i) \quad (2)$$

C. Mutual Information

The Definition 3 mutual information between information source X and Y :

$$I(X; Y) = H(X) - H(X | Y) \quad (3)$$

It can reflect the information which can be obtained by one source of information from another one, the bigger of $I(X; Y)$, means the parameters more important which be deleted.

D. The Information about Feature Attribute

$$T = \sum_{k=1}^p \lambda_k \cdot I_{\lambda_k}(X; Y) \quad (4)$$

In order to measure the characteristic attributes of information, it should be Integrated mutual information under different confidence levels, finally the information about the characteristic properties can be obtained.

III. METHOD ABOUT WEIGHTS ALLOCATION CHARACTERISTIC PROPERTIES BASED ON ENTROPY

A. Eigenvalues Establishment

Determine the fault sample to be treated, and extract the data, assume set X is on behalf of the sample to be treated:

$$X = \{x_1, x_2, \dots, x_n\}$$

Each sample is a eigenvector composed by m eigenvalues indicators:

$$X_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$$

Then, the eigenvalues X which composes by available n \times m:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

B. Normalized Eigenvalues

To avoid the influence which causes by the different dimension, normalized eigenvalues with the following methods

(1) For the index which will be better when the value is bigger:

$$s_{ij} = \frac{x_{ij} - x_j^{\min}}{x_j^{\max} - x_j^{\min}} \quad (5)$$

(2) For the index which will be better when the value is smaller:

$$s_{ij} = \frac{x_{\min}(i, j) - x_{ij}}{x_j^{\max} - x_j^{\min}} \quad (6)$$

(3) For the index which will be better when the value approaches the fixed value Re:

$$s_{ij} = \frac{|x_{ij} - \text{Re}|}{\text{Re}} \quad (7)$$

Then, the Normalized Eigenvalues is :

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1m} \\ s_{21} & s_{22} & \cdots & s_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nm} \end{bmatrix} \quad (8)$$

C. Construct Property FuzzyEquivalent Matrix

Assume r_{ij} as the measure to define the fuzzy similarity during different attributes of fault feature, and as the $r_{ij} = r_{ji}$, the paper defines the fuzzy similarity relation between fuzzy similarity relation with the maximum minimum method:

$$r_{ij} = \frac{\sum_{k=1}^n (s_{ik} \cap s_{jk})}{\sum_{k=1}^n (s_{ik} \cup s_{jk})} \quad (9)$$

Then, the fuzzy equivalent matrix is:

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nm} \end{bmatrix} \quad (10)$$

D. Fuzzy Equivalence Classification

(1) According to the fuzzy equivalent matrix, the appropriate level of confidence interval is easy to determine, divide it reasonably for the partitions, define the λ_k as the lower limit to divide the matrix, so the cutmatrix R_{λ_k} is got, denoted as follow: $U / R_{\lambda_k} = (U_1, U_2, \dots, U_r)$.

(2) Repeat the step above, removed the one property during the fault phenomenon characteristic property set, then the classification can be obtained according to the confidence level λ_k , denoted as follow: $U / S_{\lambda k} = (U_1, U_2, \dots, U_s)$.

At the same confidence level, if the difference between the original sample and the sample that has deleted one fault feature attributes is clearer, then it shows that the attributes more important, and the mutual information $I_{\lambda_k}(R; S)$ bigger; On the contrary, the difference is not becoming clear, then it shows that the attributes is not as important as others, and the mutual information $I_{\lambda_k}(R; S)$ smaller.

E. Divide the Parameter Weight

According to the information, get the weight after normalizing:

$$w_j = M_j / \sum_{j=1}^m M_j \quad (11)$$

IV. EXAMPLES OF APPLICATION

Assume the failure case characteristics F makes up by some property: Engine speed, Water temperature, Oil pressure, Hydraulic oil temperature, and defined them as A, B, C, D; There are 12 samples about fault F: Difficulty

starting the engine, Black smoke, Lack of working cylinders, Underpowered, Turn speed is not smooth, Water temperature is too high, Low oil pressure, Hydraulic oil temperature is too high. The fault information table shows as below.

TABLE I. TABLE FAULT INFORMATION

Number	Attribute A (r/min)	Attribute B (°C)	Attribute C (Mpa)	Attribute D (°C)	Number	Attribute A (r/min)	Attribute B (°C)	Attribute C (Mpa)	Attribute D (°C)
1	1650	108	0.12	115	7	1850	119	0.13	116
2	1800	109	0.13	119	8	1800	116	0.16	117
3	1900	112	0.15	117	9	1950	118	0.17	120
4	1850	116	0.17	118	10	1860	117	0.18	121
5	1750	109	0.14	117	11	1800	116	0.16	118
6	1750	114	0.14	119	12	1750	119	0.15	119

The process to divide the weights as follows:

(1) The fault feature matrix X obtained by the table show as follows

(2) Normalized and get the matrix S

(3) Get the Fuzzy similar matrix R with Maximum Minimum

(4) Successive square to get the fuzzy equivalent matrix

\bar{R}

$$X = \begin{bmatrix} 1650 & 1800 & 1900 & 1850 & 1750 & 1750 & 1850 & 1800 & 1950 & 1860 & 1800 & 1750 \\ 108 & 109 & 112 & 116 & 109 & 114 & 119 & 116 & 118 & 117 & 116 & 119 \\ 0.12 & 0.13 & 0.15 & 0.17 & 0.14 & 0.14 & 0.13 & 0.16 & 0.17 & 0.18 & 0.16 & 0.15 \\ 115 & 119 & 117 & 118 & 117 & 119 & 116 & 117 & 120 & 121 & 118 & 119 \end{bmatrix}$$

$$S = \begin{bmatrix} 0.167 & 0.333 & 0.667 & 0.500 & 0.167 & 0.167 & 0.500 & 0.333 & 0.833 & 0.533 & 0.333 & 0.167 \\ 1.000 & 0.918 & 0.673 & 0.345 & 0.918 & 0.509 & 0.100 & 0.345 & 0.182 & 0.264 & 0.345 & 0.100 \\ 0.100 & 0.250 & 0.550 & 0.850 & 0.400 & 0.400 & 0.250 & 0.700 & 0.850 & 1.000 & 0.700 & 0.550 \\ 0.100 & 0.700 & 0.400 & 0.550 & 0.400 & 0.700 & 0.250 & 0.400 & 0.850 & 1.000 & 0.550 & 0.700 \end{bmatrix}$$

$$R = \begin{bmatrix} 1.000 & 0.539 & 0.461 & 0.287 & 0.780 & 0.463 & 0.191 & 0.349 & 0.184 & 0.207 & 0.327 & 0.242 \\ 0.539 & 1.000 & 0.537 & 0.441 & 0.701 & 0.649 & 0.323 & 0.440 & 0.372 & 0.395 & 0.501 & 0.423 \\ 0.461 & 0.537 & 1.000 & 0.618 & 0.607 & 0.524 & 0.410 & 0.631 & 0.523 & 0.482 & 0.591 & 0.421 \\ 0.287 & 0.441 & 0.618 & 1.000 & 0.409 & 0.517 & 0.412 & 0.766 & 0.704 & 0.730 & 0.838 & 0.520 \\ 0.780 & 0.701 & 0.607 & 0.409 & 1.000 & 0.637 & 0.260 & 0.501 & 0.280 & 0.304 & 0.469 & 0.393 \\ 0.463 & 0.649 & 0.524 & 0.517 & 0.637 & 1.000 & 0.275 & 0.528 & 0.425 & 0.451 & 0.602 & 0.663 \\ 0.191 & 0.323 & 0.410 & 0.412 & 0.260 & 0.275 & 1.000 & 0.391 & 0.336 & 0.323 & 0.361 & 0.316 \\ 0.349 & 0.440 & 0.631 & 0.766 & 0.501 & 0.528 & 0.391 & 1.000 & 0.525 & 0.552 & 0.913 & 0.532 \\ 0.184 & 0.372 & 0.523 & 0.704 & 0.280 & 0.425 & 0.336 & 0.525 & 1.000 & 0.762 & 0.581 & 0.513 \\ 0.207 & 0.395 & 0.482 & 0.730 & 0.304 & 0.451 & 0.323 & 0.552 & 0.762 & 1.000 & 0.607 & 0.494 \\ 0.327 & 0.501 & 0.591 & 0.838 & 0.469 & 0.602 & 0.361 & 0.913 & 0.581 & 0.607 & 1.000 & 0.612 \\ 0.242 & 0.423 & 0.421 & 0.520 & 0.393 & 0.663 & 0.316 & 0.532 & 0.513 & 0.494 & 0.612 & 1.000 \end{bmatrix}$$

$$\bar{R} = \begin{vmatrix} 1.000 & 0.701 & 0.612 & 0.612 & 0.780 & 0.649 & 0.412 & 0.612 & 0.612 & 0.612 & 0.612 & 0.649 \\ 0.701 & 1.000 & 0.612 & 0.612 & 0.701 & 0.649 & 0.412 & 0.612 & 0.612 & 0.612 & 0.612 & 0.649 \\ 0.612 & 0.612 & 1.000 & 0.631 & 0.612 & 0.612 & 0.412 & 0.631 & 0.631 & 0.631 & 0.631 & 0.612 \\ 0.612 & 0.612 & 0.631 & 1.000 & 0.612 & 0.612 & 0.412 & 0.838 & 0.730 & 0.730 & 0.838 & 0.612 \\ 0.780 & 0.701 & 0.612 & 0.612 & 1.000 & 0.649 & 0.412 & 0.612 & 0.612 & 0.612 & 0.612 & 0.649 \\ 0.649 & 0.649 & 0.612 & 0.612 & 0.649 & 1.000 & 0.412 & 0.612 & 0.612 & 0.612 & 0.612 & 0.663 \\ 0.412 & 0.412 & 0.412 & 0.412 & 0.412 & 0.412 & 1.000 & 0.412 & 0.412 & 0.412 & 0.412 & 0.412 \\ 0.612 & 0.612 & 0.631 & 0.838 & 0.612 & 0.612 & 0.412 & 1.000 & 0.730 & 0.730 & 0.913 & 0.612 \\ 0.612 & 0.612 & 0.631 & 0.730 & 0.612 & 0.612 & 0.412 & 0.730 & 1.000 & 0.762 & 0.730 & 0.612 \\ 0.612 & 0.612 & 0.631 & 0.730 & 0.612 & 0.612 & 0.412 & 0.730 & 0.762 & 1.000 & 0.730 & 0.612 \\ 0.612 & 0.612 & 0.631 & 0.838 & 0.612 & 0.612 & 0.412 & 0.913 & 0.730 & 0.730 & 1.000 & 0.612 \\ 0.649 & 0.649 & 0.612 & 0.612 & 0.649 & 0.663 & 0.412 & 0.612 & 0.612 & 0.612 & 0.612 & 1.000 \end{vmatrix}$$

Intercept with the confidence level λ_k , and provide that $\bar{R}_{ij} = 1$ if $\bar{R}_{ij} \geq \lambda_k$.

When $\lambda_1 = 0.65$, the set matrix M shows as follow:

$$M = \begin{vmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix}$$

As it infers, when $\lambda_1 = 0.65$, the set can be divided into five parts: $\{1,2,5\}$, $\{3\}$, $\{4,8,9,10,11\}$, $\{7,12\}$, $\{8\}$;

So, When $\lambda_1 = 0.75$, the set can be divided into seven parts: $\{1\}$, $\{2,5,6\}$, $\{3\}$, $\{4,8,11\}$, $\{7\}$, $\{9,10\}$, $\{12\}$;

When $\lambda_1 = 0.85$, the set can be divided into eleven parts: $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, $\{5\}$, $\{6\}$, $\{7\}$, $\{8,11\}$, $\{9\}$, $\{10\}$, $\{12\}$.

Delete fault feature property A, repeat the above step to get the sorts:

When $\lambda_1 = 0.65$, the set can be divided into two parts: $\{1,2,3,4,5,6,8,9,10,11,12\}$, $\{7\}$;

When $\lambda_1 = 0.75$, the set can be divided into six parts: $\{1,3,5\}$, $\{2\}$, $\{4,8,9,10,11\}$, $\{6\}$, $\{7\}$, $\{12\}$;

When $\lambda_1 = 0.85$, the set can be divided into ten parts: $\{1\}$, $\{2\}$, $\{3\}$, $\{4,8,11\}$, $\{5\}$, $\{6\}$, $\{7\}$, $\{9\}$, $\{10\}$, $\{12\}$;

(5) Confirm information entropy and weight

When the confidence level $\lambda = 0.65$, the system information entropy is:

$$\begin{aligned} H(X) &= -\sum_{i=1}^n p_i \log_2 p_i \\ &= -\left(\frac{3}{12} \log_2 \frac{3}{12} + \frac{1}{12} \log_2 \frac{1}{12} + \frac{5}{12} \log_2 \frac{5}{12} + \frac{2}{12} \log_2 \frac{2}{12} + \frac{1}{12} \log_2 \frac{1}{12}\right) \\ &= 2.056 \end{aligned}$$

The information entropy becomes as follows if the attribute A is deleted:

$$\begin{aligned} H(Y|X) &= -\sum_{j=1}^m P(Y_j|X_i) \log_2 P(Y_j|X_i) \\ &= -\left(\frac{11}{12} \log_2 \frac{11}{12} + \frac{1}{12} \log_2 \frac{1}{12}\right) \\ &= 0.414 \end{aligned}$$

So, the information capacity that the case attains from attribute A shows as follow when the confidence level $\lambda = 0.65$

$$\begin{aligned} I_{0.65}(X;Y) &= H(X) - H(X|Y) \\ &= 2.056 - 0.414 \\ &= 1.642 \end{aligned}$$

In a similar way, $I_{0.75}(X;Y) = 0.4046$,

$$I_{0.85}(X;Y) = 0.2296;$$

So, the information capacity of attribute A is :

$$T_A = \sum_{k=1}^p \lambda_k \cdot I_{\lambda_k}(X;Y) = 1.566;$$

Repeat the steps as above, then the capacity of the other attributes can be obtained: $T_a=0.436$, $T_c=1.659$, $T_d=1.155$.

So, each characteristic attribute's weight $\{W_A, W_B, W_C, W_D\} = \{0.338, 0.094, 0.317, 0.251\}$, so the attribute B can be deleted when the fault set F is set up.

V. CONCLUSIONS

Make certain the weights with the case characteristic attribute information entropy, not only differentiate the cases rationed to organize the case-base more efficiency, but also set off the attributes more reasonably, establish the base for case retrieve, but it should base on the plurality date, and rely on the characteristic attribute date at a certain extant, so it does not differentiate the cases with qualitative descriptive perfectly, it needs to study more to differentiate the fault cases with ration and qualitative descriptive.

REFERENCES

- [1] Jian-yang Li, Zhi-wei Ni, and Hui-ting Liu. A Similarity Rough Set Technique Approach to Case- base Maintenance. Hefei University of Technology, Computer Engineering and Applications, 2005, pp.23-25.
- [2] Ya-si Pang, Xue-ning CHU, Hui CHENG. Research on Fuzzy SOM Clustering Approach to Case Base Maintenance, Machine Design & Research, 2009, pp.86-89.
- [3] Qing Liu. Rough Set and Rough Inference. Science Press, 2001, pp.40.
- [4] Wen-xiu Zhang, Wei-da Wu, Ji-ye Liang. Science Press, 2001, pp.15.
- [5] Pu-yan Liu, Meng-da Wu. Fuzzy theory and its application. National University of Defence Technology Press, 1998, p.184.