Fuzzy set and rough set based evaluation algorithm of web customers

Yun Peng*

Computer and Information Engineering College, Jiangxi Normal University, Nanchang 330022, China

Received 1 July 2014, www.cmnt.lv

Abstract

A fuzzy algorithm of web customers evaluation based on rough set is presented. Key attributes can be gotten through rough set. The evaluation from the data objects based on key attributes can reduce the data size and algorithm complexity. After Clustering analysis of customers, then the evaluation analysis will process to the clustering data. There are a lot of uncertain data in customer clusters, so the traditional method of classification and evaluation to the incomplete data is very difficult. Superposition evaluation algorithm based on fuzzy set can improve the reliability and accuracy of web customer evaluation. Evaluation of the web customer also can improve efficiency, service quality and profitability of web businesses.

Keywords: Rough Set, Membership Function, Cluster Analysis, Fuzzy Set

1 Introduction

Following the rapid development of electronic commerce, which brought a huge number of web customer groups, and how to classify and evaluate the customers is very important. According to the customer's transactions, characteristics of the customer information, customers can be classified and evaluated and given the levels. Web services can be improve efficiency in the future based on the actual conduct of targeted customer, and increase customer attention and interest level, thereby enhancing the profitability of web. Because of the ambiguity of customer behaviour, a comprehensive evaluation is designed based on fuzzy superposition algorithm in this paper [1].

2 The proposed algorithm

Customer's attribute can be web transactions based on behaviour characteristics, also it can be static characteristics of the customer, and the attribute can be looked as the basis for cluster analysis. In the classification process, the keywords are divided into different classes for different levels of customer types. Membership function will convert all the data values less than or equal to 1, so the next step of the fuzzy clustering can be carried out [2].

Fuzzy set A in the domain U, with a membership function to describe it, namely: $U \rightarrow [0,1]$, for any $u \in U$, there is $u \rightarrow \mu_{\underset{\sim}{A}}(u)$, $\mu_{\underset{\sim}{A}}(u) \in [0,1]$, $\mu_{\underset{\sim}{A}}(u)$ is the membership of element u to set A, which represents the degree of u belonging to A. The designing of fuzzy superposition evaluation algorithm is shown as follows [3].

(1) Let U be the customer domain, U_i is the No.i customer, $i \in$ 1,2,3 ... n, A_j is the No.j attribute of U, $j \in$ 1,2,3 ... m, S_{ij} is the value of U_i and A_j, $P(K_j)$ is the weighting coefficients of A_j. The membership function of keyword property value is shown in (1).

$$\mu_{\underset{\sim}{A}}(S_{ij}) = F(S_{ij})/(S_{1j} + S_{2j} \ldots S_{nj}) * P(K_j). \quad (1)$$

After pre-treatment customers, we can design an appropriate fuzzy clustering algorithm of customer data.

(2) Set up the fuzzy similarity relation R of U. The order of R matrix is $|U|$, m is the number of attributes. Using Euclidean distance formula shown in (2), we can calculate the matrix elements of r_{ij} in R [4].

$$r_{ij} = \begin{cases} 1 & i = j \\ \sqrt{\frac{1}{m}\sum_{k=1}^{m}(sik - sjk)^2} & i \neq j \end{cases}. \quad (2)$$

(3) The graph $G = (V, E)$ can be obtained by R, and the maximum spanning tree as $T = (V, TE)$ from G can be calculated using Prim algorithm.

(4) According to the practical problems to set an appropriate $\lambda \in [0,1]$, $T(e)$ is the weight of edge e, if $T(e) < \lambda$, edge e will be removed, the connected component is the classification based on λ.

(5) The set of attributes reduction can be derived from discernibility matrix and discernibility function [5]. Discernibility matrix is shown in (3).

$$M(B) = \{m(i, j) \mid n \times n, 1 \le i, j \le n\}, \quad (3)$$

* ***Corresponding author*** e-mail 332952797@qq.com

where $m(i, j) = \{a \in A \mid a(i) \neq a(j) \text{ and } d(i) \neq d(j)\}$, $n = |U|$.

Discernibility function is shown in (4). Σ is "$\vee$", Π is "$\wedge$".

$$\Delta = \prod_{(i,j)\in U\times U} \sum m(i,j) \,. \tag{4}$$

(6) Attribute reduction and nuclear can be derived from the minimal disjunctive of distinction function, which can deduce an attributes set of critical evaluation [6].

(7) Set up evaluation set $V = \{v_1, v_2, ... v_m\}$, the weight distribution of U_i is: $\underset{\sim}{A_i} = \{a_{i1}, a_{i2}, \dots, a_{in_i}\}$. The evaluation factors are the fuzzy mapping from U to $F(V)$ shown in (5).

$$\begin{aligned} &\underset{\sim}{f} : U \to F(V), \forall u \in U \\ &u_i \mapsto \underset{\sim}{f}(u_i) = \frac{r_{i1}}{v_1} + \frac{r_{i2}}{v_2} + \dots + \frac{r_{im}}{v_m} \,, \end{aligned} \tag{5}$$

where $0 \le r_{ij} \le 1$, $1 \le i \le n$, $1 \le j \le m$

$\underset{\sim}{R_i}$ is the single factor evaluation matrix of U_i, so the first-class comprehensive evaluation is as follows: $\underset{\sim}{B_i} = \underset{\sim}{A_i} \circ \underset{\sim}{R_i} = (b_{i1}, b_{i2}, \dots, b_{im})$, $(i = 1, 2, \dots s)$.

(8) As an element for each U_i, using $\underset{\sim}{B_i}$ as its single factor assessment, the evaluation matrix can be derived.

(9) According to some property of a higher level, a subset S can be divided into more advanced sub-set, then return step (7) and (8). Finally, we can constitute a multi-class fuzzy evaluation.

3 Example analysis

According to the type of attribute value and the membership function, we can calculate the membership degree of each attribute, and the membership value is as the initial value for classification. Original customer data is shown in Table 1, where C is the customer, CA is the transaction attribute [7].

TABLE 1 Original customer data

C	CA_1	CA_2	CA_3	CA_4	CA_5
C_1	322	524	401	112	123
C_2	121	510	502	142	280
C_3	126	228	224	89	118
C_4	210	217	565	321	217
C_5	260	162	865	212	322
C_6	215	159	625	112	265
C_7	129	195	587	219	285
C_8	139	112	495	120	279
C_9	562	255	750	250	105
C_{10}	409	596	362	201	419

(1) Through the mapping of membership function, data can be initialized shown in Table 2, the values are changed to the values less than or equal to 1, and the values reflect the dependence of the attribute.

TABLE 2 Initialization of customer data

C	CA_1	CA_2	CA_3	CA_4	CA_5
C_1	0.134	0.140	0.401	0.500	0.390
C_2	0.343	0.144	0.300	0.257	0.144
C_3	0.257	0.451	0.500	0.457	0.310
C_4	0.544	0.457	0.400	0.357	0.229
C_5	0.457	0.467	0.067	0.487	0.300
C_6	0.450	0.057	0.500	0.257	0.320
C_7	0.257	0.437	0.520	0.457	0.400
C_8	0.542	0.400	0.184	0.400	0.457
C_9	0.434	0.300	0.540	0.460	0.387
C_{10}	0.544	0.500	0.330	0.213	0.420

(2) Fuzzy similarity relation matrix R shown in Table 3 can be calculated using Euclidean.

TABLE 3 Fuzzy similarity relation matrix R

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
C_1									
C_2	0.032								
C_3	0.213	0.245							
C_4	0.273	0.216	0.130						
C_5	0.139	0.130	0.270	0.230					
C_6	0.247	0.237	0.126	0.140	0.270				
C_7	0.232	0.242	0.089	0.093	0.255	0.089			
C_8	0.052	0.079	0.223	0.234	0.133	0.257	0.251		
C_9	0.186	0.191	0.073	0.099	0.260	0.126	0.145	0.211	
C_{10}	0.245	0.235	0.090	0.093	0.258	0.066	0.042	0.138	0.147

(3) The graph $G = (V, E)$ can be derived by R, and the maximum spanning tree $T = (V, TE)$ from G can be calculated using Prim algorithm, where $|V| = 10$, $|TE| = 9$, the maximum spanning tree is shown in Figure 1.

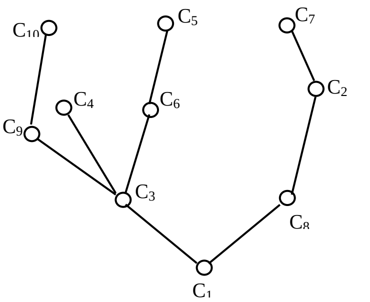

FIGURE 1 The maximum spanning tree

The three clusters above is in accordance with the feature of customers through the data analysis. Sometimes the λ is a variable value, we can draw the dynamic graph by the max-tree. The dynamic classification graph is as Figure 2.

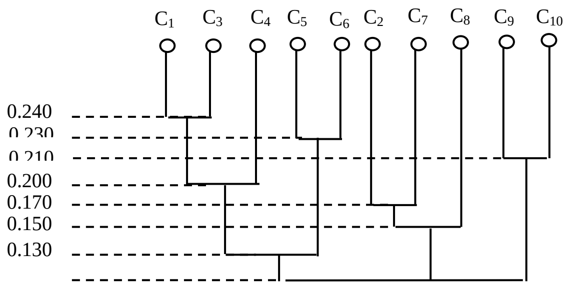

FIGURE 2 Dynamic classification graph

Set $\lambda = 0.130$, customer classification can be obtained as follows:

$C1 = \{C_1, C_3, C_4, C_5, C_6\}$, $C2 = \{C_2, C_7, C_8\}$, $C3 = \{C_9, C_{10}\}$.

(4) The distinction matrix is shown in Table 4.

TABLE 4 Distinction matrix

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
C_1									
C_2	2,5								
C_3	3,4	4,5							
C_4	3	2,4,5	4						
C_5	1,2,3	1,2,3,4	1,3,4	1,2,3,4					
C_6	2,5	1,2,4,5	1,2,4	1,2,4,5	3,4				
C_7	2,4	2,3,5	1,2	1,2	3,4,5	4,5			
C_8	1,2	2,3.4,5	2,4	2,4	1,2,3,4,5	1,2,5	1,2,4		
C_9	2,3,4	1,2,3,5	1,2	1,2	3,4,5	4,5	2	3.4,5	
C_{10}	1,3,4,5	2,3,4,5	2,3,5	1,4,5	1,3	3,4	2,4,5	2,5	1,3,5

According to Table 4, we can get distinction function as follows:

$$\begin{aligned}\Delta=&(CA_2\vee CA_5)\wedge(CA_3\vee CA_4)\wedge(CA_4\vee CA_5)\wedge CA_3\wedge(CA_2\vee CA_4\vee CA_5)\wedge CA_4\wedge(CA_1\vee CA_2\vee CA_3)\wedge\\&(CA_1\vee CA_2\vee CA_3\vee CA_4)\wedge(CA_1\vee CA_3\vee CA_4)\wedge\\&(CA_1\vee CA_2\vee CA_4\vee CA_5)\wedge(CA_1\vee CA_2\vee CA_4)\wedge\\&(CA_2\vee CA_4)\wedge(CA_2\vee CA_3\vee CA_5)\wedge(CA_1\vee CA_2)\wedge\\&(CA_3\vee CA_4\vee CA_5)\wedge(CA_2\vee CA_3\vee CA_4\vee CA_5)\wedge\\&(CA_1\vee CA_2\vee CA_3\vee CA_4\vee CA_5)\wedge(CA_1\vee CA_2\vee CA_5)\wedge(CA_2\vee CA_3\vee CA_4)\wedge(CA_1\vee CA_2\vee CA_3\vee CA_5)\wedge CA_2\wedge(CA_1\vee CA_3\vee CA_4\vee CA_5)\wedge(CA_1\vee CA_4\vee CA_5)\wedge\\&(CA_1\vee CA_3)\wedge(CA_2\vee CA_5)\wedge(CA_1\vee CA_3\vee CA_5)\\=&CA_1\wedge CA_4\wedge CA_5\end{aligned}$$

After reduction we can get the key attributes as $\{CA_1, CA_4, CA_5\}$ and the second layer attribute as $\{CA_2, CA_3\}$. A two-tiered evaluation system for fuzzy evaluation can carried out then. To analyse the evaluation system conveniently, we simplify the system accordingly, but does not affect the algorithm analysis [5].

(5) Customer evaluation grades are divided into four grade (excellent, good, middle, bad). The weight of first-class is: $\underset{\sim}{W}=\{0.30,0.40,0.20,0.10\}$.

The weight of second-class is:

$\underset{\sim}{W}_1=\{0.35,0.45,0.20\}$;

$\underset{\sim}{W}_2=\{0.30,0.25,0.10,0.20,0.15\}$;

$\underset{\sim}{W}_3=\{0.45,0.30,0.25\}$;

$\underset{\sim}{W}_4=\{0.32,0.30,0.20,0.18\}$.

The weight of third-class is:

$\underset{\sim}{W}_{11}=\{0.50,0.30,0.20\}$;

$\underset{\sim}{W}_{12}=\{0.30,0.20,0.15,0.25,0.10\}$.

(6) The fuzzy evaluation matrix is shown in (6) and fuzzy evaluation result of second-class is shown in (7).

$$\underset{\sim}{R}_{11}=\begin{bmatrix}0.25 & 0.45 & 0.20 & 0.10\\0.30 & 0.20 & 0.35 & 0.15\\0.28 & 0.22 & 0.39 & 0.11\end{bmatrix} \tag{6}$$

$$\underset{\sim}{B}_{11}=\underset{\sim}{W}_{11}\circ\underset{\sim}{R}_{11}=$$

$$\underset{\sim}{B}_{11}\begin{bmatrix}(0.50\wedge0.25)\vee(0.30\wedge0.30)\vee(0.20\wedge0.28)\\(0.50\wedge0.45)\vee(0.30\wedge0.20)\vee(0.20\wedge0.22)\\(0.50\wedge0.20)\vee(0.30\wedge0.35)\vee(0.20\wedge0.39)\\(0.50\wedge0.10)\vee(0.30\wedge0.15)\vee(0.20\wedge0.11)\end{bmatrix}=(0.30,0.45,0.30,0.15),$$

the normalized result is obtained: $=(0.25,0.40,0.25,0.10)$.

Using the same algorithm can be obtained:

$\underset{\sim}{B}_{12}=(0.30,0.40,0.20,0.10)$;

$\underset{\sim}{B}_{13}=(0.25,0.35,0.20,0.20)$.

$$\underset{\sim}{B}_1=\underset{\sim}{W}_1\circ\underset{\sim}{R}_1=$$

$$\begin{bmatrix}(0.35\wedge0.25)\vee(0.45\wedge0.30)\vee(0.20\wedge0.25)\\(0.35\wedge0.40)\vee(0.45\wedge0.40)\vee(0.20\wedge0.35)\\(0.35\wedge0.25)\vee(0.45\wedge0.20)\vee(0.20\wedge0.20)\\(0.35\wedge0.10)\vee(0.45\wedge0.10)\vee(0.20\wedge0.20)\end{bmatrix}=, \tag{7}$$

$$(0.30,0.40,0.25,0.20)$$

The normalized result is obtained:

$\underset{\sim}{B}_1=(0.25,0.35,0.25,0.15)$.

Using the same algorithm can be obtained:

$\underset{\sim}{B}_2=(0.25,0.35,0.20,0.20)$;

$\underset{\sim}{B}_3=(0.30,0.45,0.10,0.15)$;

$\underset{\sim}{B}_4=(0.25,0.30,0.20,0.25)$.

(7) According to the second-class evaluation results matrix, first-class level evaluation shown in (8) can be derived.

$$\underset{\sim}{B}=\underset{\sim}{W}\circ\underset{\sim}{R}=$$

$$\begin{bmatrix}(0.30\wedge0.25)\vee(0.40\wedge0.25)\vee(0.20\wedge0.30)\vee(0.10\wedge0.25)\\(0.30\wedge0.35)\vee(0.40\wedge0.35)\vee(0.20\wedge0.45)\vee(0.10\wedge0.30)\\(0.30\wedge0.25)\vee(0.40\wedge0.20)\vee(0.20\wedge0.10)\vee(0.10\wedge0.20)\\(0.30\wedge0.15)\vee(0.40\wedge0.20)\vee(0.20\wedge0.15)\vee(0.10\wedge0.25)\end{bmatrix}=. \tag{8}$$

$$(0.25,0.35,0.25,0.20)$$

The normalized result: $\underset{\sim}{B}=(0.24,0.34,0.24,0.18)$.

Based on the maximization of fuzzy set membership, the evaluation results of customer class C1 can be rated as "good".

4 Conclusions

The evaluation of web customers adopts multi-class structure and takes the static and dynamic attributes into account. To delete the redundant attributes and improve the computation efficiency, attributes reduction is necessary. The evaluation result is an important reference for commercial distribution of web businesses. The weight factors and the fuzzy evaluation algorithm can be adjusted according to the actual customer data. In designation of superposition algorithm to evaluate web customers, compliance with requirements of enterprise, true reflection of the needs and behaviour of customers and adaptive algorithm for data processing are all to be considered. Tested by actual data analysis, cluster analysis can reduce the size of customer data and data noise, and the key class evaluation analysis can improve the evaluation efficiency of web customers.

Acknowledgments

This work is supported by Social Science Planning Project of Jiangxi Province under grant No. 13TQ08, and Science Research Project of Jiangxi Normal University under grant No.4546.

References

[1] Kopylov A V, Moskovtsev A F 2013 The model of assessment of strategic resources of industrial enterprise on the basis of fuzzy set theory *World Applied Sciences Journal* **23**(1) 37-45

[2] Deng-Neng Chen, Jeng B, Wei-Po Lee, Cheng-Hung Chuang 2008 An agent-based model for consumer-to-business electronic commerce *Expert Systems with Applications* **34**(1) 469-81

[3] Yan Zou, Zhi Xiao 2008 Data analysis approaches of soft sets under incomplete information *Knowledge-Based Systems* **21**(8) 941-5

[4] Cho J, Lee J 2013 Development of a new technology product evaluation model for assessing commercialization opportunities using Delphi method and fuzzy AHP approach *Expert Systems with Applications* **40**(13) 5314-30

[5] Jia X, Tang Z, Liao W, et al. 2014 On an optimization representation of decision-theoretic rough set model *International Journal of Approximate Reasoning* **55(**1) 156-66

[6] Li B, Tang P, Chow T W S 2013 Quantization of rough set based attribute reduction *Journal of Software Engineering and Applications* **5**(12) 117-23

[7] Eid M I 2011 Determinants of web customer satisfaction, trust, and loyalty in Saudi Arabia *Journal of Electronic Commerce Research* **12**(1) 78-93

Authors

Yun Peng, born on October 27, 1972, China

Current position, grades: a Lecturer from 2002 to 2007 and currently an Associate Professor in the Computer and Information Engineering College at Jiangxi Normal University
University studies: B.E. degree in electronic engineering from Nanchang University, Nanchang, P.R.China, in 1994 and the M.E. degree in computer software and theory from Jiangxi Normal University
Experience: He is a member of CCF and ACM. His research interests include artificial intelligence, data mining and software engineering.