Journal of Theoretical and Applied Information Technology
31st October 2012. Vol. 44 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A FUZZY EVALUATION ALGORITHM OF E-COMMERCE CUSTOMERS BASED ON ATTRIBUTES REDUCTION

[1]YUN PENG, [2]HONGXIN WAN

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

[2]Mathematics & Computer Science College, Jiangxi Science & Technology Normal University, Nanchang 330013, Jiangxi, China

ABSTRACT

The evaluation algorithm is based on the attributes of data objects. There is a certain correlation between attributes, and attributes are divided into key attributes and secondary attributes. The evaluation from the data objects in a hierarchical design based on key attributes can reduce the data size and algorithm complexity, and without prejudice on the basis of evaluation results can improve the accuracy of the algorithm. This paper proposes an algorithm of attribute reduction based on rough set and the hierarchical evaluation based on fuzzy set. The algorithm of hierarchical fuzzy evaluation based on attributes reduction is described in detail by example. According to the analysis of the current evaluation of e-commerce customers, a superposition method based on fuzzy evaluation algorithm is presented. After Clustering analysis of customers, then the evaluation analysis will be processed on the clustered data. There are a lot of uncertain data of customer clustering, so the traditional method of classification and evaluation to the incomplete data will be very complex. Superposition evaluation algorithm based on fuzzy set can improve the reliability and accuracy of e-commerce customer evaluation. Evaluation of the e-commerce customer also can improve efficiency, service quality and profitability of e-commerce businesses.

Keywords: *Rough Set, Attributes Reduction, Fuzzy Set, Fuzzy Evaluation, Similarity Relation Matrix*

1. INTRODUCTION

Following the rapid development of electronic commerce, which brought a huge number of 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. E-commerce 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 e-commerce [1].Because of the ambiguity of customer behavior, a comprehensive evaluation is designed based on fuzzy superposition algorithm in this paper.

Because of the object characteristics of things with a particularly high dimension vector property description, make it difficult to conduct effective evaluation, and comparison between the calculated vector will further increase the capacity of things. Rough set is a new mathematical tool to deal with incomplete and uncertain knowledge, and its main feature is not given in advance the number of certain characteristics or properties described, but the problem can directly derived from the classification of knowledge of a given starting indiscernible relation[2]. The data objects can export a minimum feature set of attributes, and do not affect the classification accuracy of the dimension of feature vectors. Since many uncertainties in the value of the attributes, the evaluation algorithm of key attributes is designed based on fuzzy set after attributes reduction.

2. ATTRIBUTES REDUCTION

Data object is described as S = (U, A), U is the set of objects and A is the set of attributes. Some attributes is redundant, while maintaining the same classification ability of the objects, we can delete these redundant attributes, that is, knowledge reduction, after reduction of excess property can obtain the minimum attributes set. If ind (B) = ind (B-{a}), where B belongs to A, $a \in A$, so attributes of B can be reduced. A set of objects may exist in several attributes reductions, and the reductions are defined as the intersection of nuclear effects in the classification of important attribute. Stratified and

Journal of Theoretical and Applied Information Technology
31st October 2012. Vol. 44 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

hierarchical analysis using fuzzy comprehensive evaluation of key attributes is to determine the level of a certain standard of measuring and comparing the various factors, and then give their relative importance quantitative form, the proportion of weight matrix by calculating the weight ratio of the largest eigenvalue matrix and its corresponding orthogonal eigenvectors can be obtained by the level of the weight of each factor[3]. Algorithm design steps are described as follows.

(1)According to the actual situation of data objects, we can select some attributes as evaluation attributes.

(2)The set of attributes reduction can be derived from discernibility matrix and discernibility function[4]. Discernibility matrix is shown in (1).

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

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

(3)Discernibility function is shown in (2). Σ is "$\vee$", Π is "$\wedge$".

$$\Delta = \prod_{(i,j)\in U\times U} \sum m(i,j) \qquad (2)$$

(4)Attribute reduction and nuclear can be derived from the minimal disjunctive of distinction function, which can deduce an attributes set of critical evaluation[5].

3. FUZZY EVALUTION ALGORITHM

Customer's attribute can be e-commerce transactions based on behavioral 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[6].

Fuzzy set A in the domain U, with a membership function μ_{A} to describe it, namely: U → [0,1], for any u ∈ U, there is u → (u), (u) ∈ [0,1], $\mu \underset{\sim}{A}(u)$ is the membership of element u to set A, which represents the degree of u belongs to A[7]. The designing of fuzzy superposition evaluation algorithm is shown as follows.

(1)Let U be the customer domain, U_i is the No.i customer, i ∈ 1,2,3 ... n, A_j is the No.j attribute of U, j ∈ 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: $(S_{ij}) = F (S_{ij}) / (S_{1j} + S_{2j} \ldots S_{nj}) * P (K_j)$.After pretreatment customers, we can design an appropriate fuzzy clustering algorithm of customer data[8].

(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 (3), we can calculate the matrix elements of r_{ij} in R.

$$rij = \begin{cases} 1 & i = j \\ \sqrt{\frac{1}{m}\sum_{k=1}^{m}(sik - sjk)^2} & i \ne j \end{cases} \qquad (3)$$

(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 λ ∈ [0,1], T (e) is the weight of edge e, if T (e) <λ, edge e will be removed, the connected component is the classification based on λ.

(5)The evaluation factors set U according to the type of customer attributes is divided into sub-set denoted by U_1, U_2, ... U_S and shown in (4).

$$\bigcup_{i=1}^{S} U_i = U, U_i \cap U_j = \varnothing \quad (i \ne j) \quad (4)$$

Where $U_i = \{u_{i1}, u_{i2}, \ldots, u_{in_i}\}$ (i=1,2,…,s) and $\sum_{i=1}^{n} n_i = n, n = |U|$.

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

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

Where 0 ≤ rij ≤ 1, 1 ≤ i ≤ n, 1 ≤ j ≤ m. Fuzzy relation can be derived from , and get the fuzzy matrix shown in (6).

$$\underset{\sim}{R} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nm} \end{bmatrix} \qquad (6)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$\underset{\sim}{R}_i$ is the single factor evaluation matrix of U_i, so the first-class comprehensive evaluation is $\underset{\sim}{R}_i = \underset{\sim}{A}_i \circ \underset{\sim}{R}_i = (b_{i1}, b_{i2}, ..., b_{im})$ (i = 1,2, ..., s) [9].

(7)As an element for each U_i, using $\underset{\sim}{B}_i$ as its single factor assessment, the evaluation matrix is shown in (7).

$$\underset{\sim}{R} = \begin{bmatrix} \underset{\sim}{B}_1 \\ \underset{\sim}{B}_2 \\ \vdots \\ \underset{\sim}{B}s \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ b_{s1} & b_{s2} & \cdots & b_{sm} \end{bmatrix} \quad (7)$$

(8)It is the single-factor evaluation matrix of $\{U_1, U_2, ..., U_i\}$, each U_i reflecting the certain property of U, the importance can be given according to their weight distribution as $\underset{\sim}{A} = (a_1^*, a_2^*, \cdots, a_s^*)$,the second-class evaluation as

$\underset{\sim}{B} = \underset{\sim}{A} \circ \underset{\sim}{R}$, $\underset{\sim}{B} = (b_1, b_2, \cdots, b_m)$,

it is the V on a fuzzy set. If the evaluation result is not 1, it should be normalized.

(9)According to some property of a higher level, a subset S can be divided into more advanced sub-set, then return above steps. At last, we can constitute a multi-class fuzzy evaluation.

4. EXAMPLE ANALYSIS

Data objects is shown in Table 1. It includes objects U_1 ...U_{10} and attributes A_1 ...A_n. We take Top 5 attributes for analysis, the distinction matrix is shown in Table 2.

(1)According to Table 2. we can get distinction function as follows:

$\Delta = (A_4 \vee A_5) \wedge (A_2 \vee A_4) \wedge (A_3 \vee A_5) \wedge A_4 \wedge (A3 \vee A_4 \vee A_5) \wedge A_3 \wedge (A_1 \vee A_2 \vee A_3 \vee A_4) \wedge (A_1 \vee A_2 \vee A_5) \wedge (A_1 \vee A_2 \vee A_4 \vee A_5) \wedge (A_3 \vee A_4) \wedge (A_2 \vee A_4) \wedge (A_1 \vee A_2 \vee A_3 \vee A_5) \wedge (A_1 \vee A_2) \wedge (A_4 \vee A_5) \wedge A_2 \wedge (A_2 \vee A_3 \vee A_4 \vee A_5) \wedge (A_2 \vee A_4 \vee A_5) \wedge (A_1 \vee A_3) \wedge (A_1 \vee A_4) \wedge (A_1 \vee A_3 \vee A_4 \vee A_5) = A_2 \wedge A_3 \wedge A_4$

Table 1: Initial Data Objects

U	A_1	A_2	A_3	A_5	A_5	…	A_n
U_1	0.574	0.257	0.700	0.477	0.400	…	…
U_2	0.577	0.277	0.087	0.577	0.200	…	…
U_3	0.576	0.067	0.700	0.577	0.500	…	…
U_5	0.572	0.277	0.730	0.577	0.500	…	…
U_5	0.533	0.200	0.135	0.200	0.287	…	…
U_7	0.533	0.210	0.700	0.277	0.277	…	…
U_7	0.533	0.200	0.605	0.577	0.530	…	…
U_8	0.578	0.143	0.700	0.277	0.540	…	…
U_9	0.533	0.207	0.403	0.577	0.550	…	…
U_{10}	0.579	0.206	0.804	0.200	0.271	…	…

(2)After reduction we can get the key attributes as {A2, A3, A4} and the second layer attribute as {A1, A5}. A two-tiered evaluation system for fuzzy evaluation can carried out then. To analyze the evaluation system conveniently, we simplify the system accordingly, but does not affect the algorithm analysis[9]. 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 3., where C is the customer, CA is the transaction attribute.

(3)Through the mapping of membership function, data can be initialized shown in Table 4., the values are changed to the values less than or equal to 1, and the values reflect the dependence of the attribute.

(4)Fuzzy similarity relation matrix R shown in Table 5. can be calculated using Euclidean.

(5)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. Set λ = 0.310, remove the edge of value less than λ, the connected component is shown in Figure 1.

(6)Thus, customer classification can be obtained as follows: C1 = {c1, c2, c5, c6, c7, c8, c10} ,C2 = {c3, c4}, C3 = {c9} ,Set z0 = 0.250, the ratio of the number of classes less than z0 should be deleted. P (C1) = 7 / 10 = 0.700, P (C2) = 2 / 10 = 0.200,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

P(C3)= 1 / 10 = 0.100, So, C1 can be used as sample class for fuzzy evaluation.

Table 2: Distinction Matrix

	U_1	U_2	U_3	U_4	U_5	U_6	U_7	U_8	U_9
U_1									
U_2	4,5								
U_3	2,4	3,5							
U_4	4	3,4,5	3						
U_5	1,2,3,4,5	1,2,3,4	1,2,3,4	1,2,3,4					
U_6	1,2,5	1,2,4,5	1,2,4,5	1,2,4,5	3,4				
U_7	2,4	1,2,3,5	1,2	1,2	3,4,5	4,5			
U_8	2	2,3.4,5	2,4	2,4	1,2,3,4,5	1,25	1,2,4		
U_9	1,2,4	1,2,3,5	1,2	1,2	3,4,5	4,5	2	2,3.4,5	
U_{10}	2,4,5	2,3,4,5	2,4,5	2,4,5	1,3	1,4	1,4,5	2,4,5	1,3,4,5

Table 3: Original Customer Data

C	CA_1	CA_2	CA_3	CA_4	CA_5
C_1	233	515	401	113	123
C_2	121	410	502	143	280
C_3	126	228	335	89	118
C_4	110	317	465	221	217
C_5	360	162	865	213	322
C_6	215	159	634	112	264
C_7	129	194	487	319	285
C_8	139	113	695	120	279
C_9	462	245	750	250	104
C_{10}	309	596	362	201	119

Table 4: Initialization Customer Data

C	CA1	CA3	CA3	CA4	CA5
C_1	0.133	0.133	0.301	0.300	0.400
C_2	0.333	0.133	0.300	0.357	0.133
C_3	0.257	0.357	0.500	0.357	0.400
C_4	0.533	0.357	0.400	0.357	0.300
C_5	0.357	0.357	0.057	0.457	0.300
C_6	0.457	0.057	0.500	0.257	0.400
C_7	0.257	0.357	0.500	0.457	0.400
C_8	0.533	0.300	0.133	0.300	0.357
C_9	0.433	0.300	0.500	0.357	0.357
C_{10}	0.533	0.300	0.500	0.457	0.300

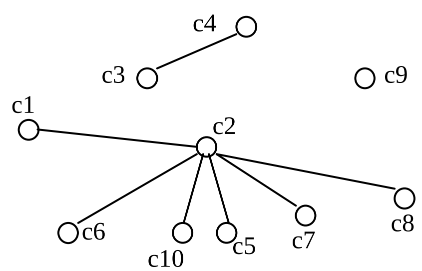

Figure 1: Connected Components

$\underset{\sim}{W}$(7)Customer evaluation grades are divided into four grades (excellent, good, middle, bad). The weight of first-class is : ={0.35,0.30,0.18,0.17}.

The weight of second-class is: ={0.32，0.40，0.28}； ={0.28，0.20，0.18，0.19，0.15}； ={0.40，0.31，0.29}； ={0.32，0.30，0.20，0.18}.

The weight of third-class is: ={0.45，0.35，0.20}； ={0.25，0.22，0.18，0.19,0.17}.

(8)The fuzzy evaluation matrix is shown in (8) and fuzzy evaluation result of second-class is shown in (9).

$$\underset{\sim}{R}_{11} = \begin{bmatrix} 0.27 & 0.45 & 0.18 & 0.10 \\ 0.20 & 0.21 & 0.37 & 0.22 \\ 0.23 & 0.24 & 0.38 & 0.15 \end{bmatrix} \quad (8)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$\underset{\sim}{B}_{11} = \underset{\sim}{W}_{11} \circ \underset{\sim}{R}_{11} = \begin{bmatrix} (0.45\wedge 0.27)\vee(0.35\wedge 0.20)\vee(0.20\wedge 0.23) \\ (0.45\wedge 0.45)\vee(0.35\wedge 0.21)\vee(0.20\wedge 0.24) \\ (0.45\wedge 0.18)\vee(0.35\wedge 0.37)\vee(0.20\wedge 0.38) \\ (0.45\wedge 0.10)\vee(0.35\wedge 0.22)\vee(0.20\wedge 0.15) \end{bmatrix} = (0.27, \quad 0.45, \quad 0.35, \quad 0.22). \tag{9}$$

Table 5: Fuzzy Similarity Relation Matrix R

	U1	**U2**	**U3**	**U4**	**U5**	**U6**	**U7**	**U8**	**U9**
U1									
U2	0.042								
U3	0.213	0.225							
U4	0.223	0.211	0.130						
U5	0.149	0.130	0.270	0.240					
U6	0.237	0.237	0.126	0.130	0.270				
U7	0.242	0.242	0.089	0.094	0.255	0.089			
U8	0.052	0.079	0.223	0.244	0.133	0.257	0.251		
U9	0.186	0.191	0.073	0.099	0.260	0.126	0.115	0.211	
U10	0.235	0.235	0.099	0.094	0.258	0.066	0.042	0.248	0.107

The normalized result is obtained:

$\underset{\sim}{B}_{11} = (0.20,0.35,0.25,0.20)$.

Using the same algorithm can be obtained:

$\underset{\sim}{B}_{12} = (0.21,0.40,0.25,0.14)$,

$\underset{\sim}{B}_{13} = (0.22,0.38,0.21,0.19)$.

$$\underset{\sim}{B}_{1} = \underset{\sim}{W}_{1} \circ \underset{\sim}{R}_{1} = \begin{bmatrix} (0.32\wedge 0.20)\vee(0.40\wedge 0.21)\vee(0.28\wedge 0.22) \\ (0.32\wedge 0.35)\vee(0.40\wedge 0.40)\vee(0.28\wedge 0.38) \\ (0.32\wedge 0.25)\vee(0.40\wedge 0.25)\vee(0.28\wedge 0.21) \\ (0.32\wedge 0.20)\vee(0.40\wedge 0.14)\vee(0.28\wedge 0.19) \end{bmatrix} = (0.22, 0.40, 0.25, 0.20) \tag{10}$$

(9)The normalized result is obtained: $\underset{\sim}{B}_1 =$ (0.20,0.40,0.20,0.20). Using the same algorithm can be obtained:

$\underset{\sim}{B}_2 = (0.15,0.35,0.26,0.24)$,

$\underset{\sim}{B}_3 = (0.25,0.40,0.10,0.25)$,

$\underset{\sim}{B}_4 = (0.24,0.30,0.17,0.29)$.

(10)According to the second-class evaluation results matrix, first-class level evaluation shown in (10) can be derived.

(11)The normalized result:

$\underset{\sim}{B}$ =(0.18,0.34,0.25,0.23). Based on the maximization of fuzzy set membership, the evaluation results of customer class C1 can be rated as "good".

$$\underset{\sim}{B} = \underset{\sim}{W} \circ \underset{\sim}{R} = \begin{bmatrix} (0.35\wedge 0.20)\vee(0.30\wedge 0.15)\vee(0.18\wedge 0.25)\vee(0.17\wedge 0.24) \\ (0.35\wedge 0.40)\vee(0.30\wedge 0.35)\vee(0.18\wedge 0.40)\vee(0.17\wedge 0.30) \\ (0.35\wedge 0.20)\vee(0.30\wedge 0.26)\vee(0.18\wedge 0.10)\vee(0.17\wedge 0.17) \\ (0.35\wedge 0.20)\vee(0.30\wedge 0.24)\vee(0.18\wedge 0.25)\vee(0.17\wedge 0.29) \end{bmatrix} = (0.20, 0.35, 0.26, 0.24) \tag{11}$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5. CONCLUSION

The division of attributes at different levels of data objects can provide a more detailed evaluation by key attributes. Rough set method can divide attributes into key attributes and secondary attributes. The evaluation of the key attributes reflects the importance of key attributes, and increases the accuracy and efficiency. The remaining key attributes and secondary attributes can also be layered to form a multi-evaluation system. We adopt a comprehensive fuzzy evaluation strategy, it can eliminate the interference of uncertain data, and get a more objective evaluation results.

The evaluation of e-commerce customers adopts multi-class structure and takes the static and dynamic attributes into account. The evaluation result is an important reference for commercial distribution of e-commerce 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 e-commerce customers, compliance with requirements of enterprise, true reflection of the needs and behavior 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 e-commerce customers.

ACKNOWLEDGEMENTS

This work was supported by 2012 National Natural Science Foundation of China (Granted ID:61262015).

REFERENCES:

[1] Toly Chen, Yi-Chi Wang, "A Fuzzy Set Approach For Evaluating And Enhancing The Mid-term Competitiveness Of A Semiconductor Factory", Fuzzy Sets and Systems, Vol.160, No.5, 2009, pp. 569-585.

[2] D. Alisantoso, LP Khoo, "Early Analysis Of Design Concepts Using Rough Set Theory", International Journal of Systems Science, Vol.40, No.2, 2009, pp.121-130.

[3] S. Nefti, M. Oussalah, U. Kaymak, "A New Fuzzy Set Merging Technique Using Inclusion-Based Fuzzy Clustering", IEEE Transactions on Fuzzy Systems, Vol.16, No.1, 2008, pp.145-161.

[4] Xiaohong Zhang, "Topological Residuated Lattice: A Unifying Algebra Representation of Some Rough Set Models", LNCS: Rough sets and knowledge technology, Vol. 5589, 2009, pp.102-110.

[5] Hexiang Bai, Yong Ge, "Handling Spatial-Correlated Attribute Values in a Rough Set", LNCS: Computational science and its applications, Vol. 5592, 2009, pp.467-478.

[6] Deng-Neng Chen, B. Jeng, Wei-Po Lee, Cheng-Hung Chuang, "An agent-based model for consumer-to-business electronic commerce", Expert Systems with Applications, Vol.34, No.1, 2008, pp. 469-481.

[7] Yan Zou, Zhi Xiao, "Data Analysis Approaches Of Soft Sets Under Incomplete Information", Knowledge-Based Systems, Vol. 21, No. 8, 2008, pp. 941-945.

[8] C. W. Gao, H. Z. Cheng, H. B. Zhai, "Fuzzy Evaluation Model for Uncertainties in Electrical Network Planning", International Journal of Power & Energy Systems, Vol. 28, No. 3, 2008, pp. 298-304.

[9] K. Saffu, J. H. Walker, R. Hinson, "Strategic Value and Electronic Commerce Adoption Among Small and Medium-Sized Enterprises in a Transitional Economy", The Journal of Business & Industrial Marketing, Vol. 23, No. 6, 2008, pp. 395-404.