International Journal of u- and e- Service, Science and Technology
Vol.8, No.4 (2015), pp.69-76
http://dx.doi.org/10.14257/ijunesst.2015.8.4.08

Research on E-commerce Consumer Behavior Prediction based on Rough Sets

Yanrong Zhang[1], Zhijie Zhao[1], Jing Yu[2] and Kun Wang[1]

(1. College of Computer and Information Engineering, Harbin University of Commerce, Harbin 150028, China;
2. Heilongjiang science and technology museum, Harbin 150018, China)
zhangyanrong_5@163.com

Abstract

To solve the traditional problem of knowledge acquisition bottleneck in e-commerce, an improved algorithm of attribute reduction based on discernibility matrix is proposed. The algorithm is used to attribute reduction for e-commerce consumer behavior prediction. With rule extraction model of rough sets, the rules of e-commerce consumer behavior prediction are acquired. Practical example of consumer behavior prediction shows that the proposed approach can be handled found knowledge effectively and can be converted the available rules easily. It has strong ability of fault tolerance and can improve the speed and quality of knowledge acquisition. The method has good practical value.

Keywords: *Rough sets; E-commerce; Consumer behavior prediction*

1. Introduction

E-commerce consumer behavior is different from traditional consumer behavior, which has the more significant features are that the businesses and consumers can communicate through the interface of website, rather than face to face. E-commerce is a platform which contributes to the communication between businesses and consumers, they are establishing the commercial sites for supporting the consumers' online-shopping. The methods of presenting the information of website and organizing the modules of website, it would have a very strong role in stimulating for consumers. It also influences the decisions of purchasing and the behavior of buying. The key to network marketing is that how to use this information interface and attract many consumers [1].

In this paper, it bases on the e-commerce consumers' various background; review the whole process of e-commerce consumers' consuming behavior and the facts that affect the purchasing decision from the perspective of informatics. Because of the knowledge related to the e-commerce consumer behavior is large, in order to be able to dig out potentially useful information from the clutter of the huge amounts of data quickly and accurately and put it into the e-commerce consumer behavior prediction , rough set theory is applied to the prediction process of e-commerce consumer behavior in this paper. Through the collection, completion and discretization of data related to e-commerce consumer behavior, an improved algorithm of attribute reduction based on discernibility matrix has been come up, after the reduction of a set of conditional attribute about e-commerce consumer behavior based on this algorithm, the extraction and reduction to the produced rules, we obtain a new method of e-commerce consumer behavior prediction based on the rough set theory [2].

ISSN: 2005-4246 IJUNESST

2. Application of Rough Sets

2.1 Lower Approximation and the Approximation

K= (U, S) For a given knowledge base, U which means that on the field, S and U as the cluster equivalence relation, then, $\forall X \subseteq U$ and U an equivalence relation on a subset of $R \in IND(K)$. Knowledge about X and R 's Lower Approximation definition are:

$$\overline{R}(X) = \{x \mid (\forall x \in U) \Lambda ([x]_R \cap X \neq \varnothing)\} \quad (1)$$

$$\underline{R}(X) = \{x \mid (\forall x \in U) \Lambda ([x]_R \subseteq X)\} \quad (2)$$

X on the set of approximation, lower approximation and boundary region shown in Figure 1.

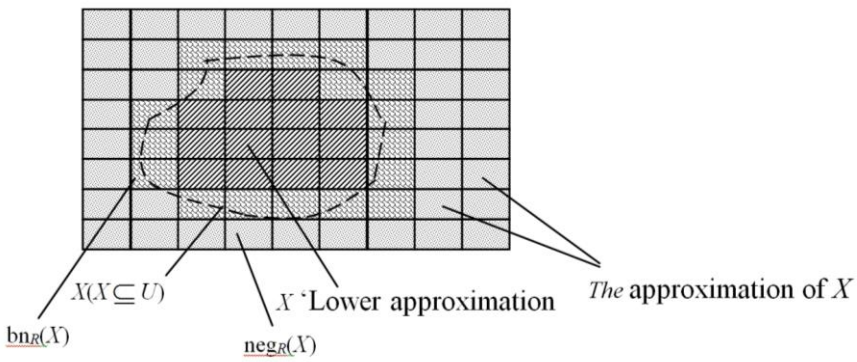


Figure1. The Schematic diagram of *X*

In the middle of them ,U is on the field, which means that the elements contained in the entire region ; universe equivalence relations R under the classification model U/R ,represents a collection of all the squares in Figure 1consists of ; R is equivalence relations, the figure is divided into horizontal and vertical lines across the region [2].

2.2 Knowledge Reduction

Knowledge Base knowledge is not as important, some knowledge is redundant. Knowledge reduction is irrelevant or redundant features some of the lost, does not affect the analysis and forecasting functions of its original premise, will reduce the amount of information. That is, without affecting the original knowledge classification cases, the n dimensional information space $\{x_1, x_2, ..., x_n\}$,reduced to m dimensions $\{x_1, x_2, ..., x_m\}$ $(m < n)$ .Through knowledge reduction, feature information obtained regroup generated decision rule, the use of decision rules generated by the new decision rules consistent with the results of reasoning Reduction feature information obtained before the reasoning arising drawn [2].

The following is the definition of nuclear reduction, which is a knowledge reduction in the two most fundamental and important concepts.

For the knowledge base $K = (U, S)$ and K is an equivalence relation family $P \subseteq S$, arbitrary $G \subseteq P$, if G is independent and IND (G) = IND (P), then G is P a reduction, denoted $G \in$ RED (P), where the consisting of P a collection of all the reduction indicated by RED (P).

Let P be a family of equivalence relations $P \subseteq S, \forall R \in P$, if $\mathrm{IND}(P) = \mathrm{IND}(P - \{R\})$

Establishment, R is called unnecessary in P; given a knowledge base $K = (U, S)$ and knowledge base equivalence relation family $P \subseteq S$, for any $R \in P$, if R satisfies

$\mathrm{IND}(P-\{R\}) \neq \mathrm{IND}(P)$

R is essential in P; core of P the collection consists of all the necessary relationships called P, denoted by $\mathrm{CORE}(P)$.

2.3 Knowledge Representation System

Intelligent data processing, knowledge representation and occupy an important position. Knowledge representation system is the main rough set theory knowledge representation, which is expressed as $S=(U,A,V,f)$, usually $S=(U,A)$ instead of $S=(U,A,V,f)$. Wherein U said finite non-empty set of objects, A that is, domain $V=\bigcup_{a\in A} V_a$, V_a; V_a said nonempty finite a set of attributes, f that is $U\times A\rightarrow V$, the set of attributes;, which means the property range; f as a function of information, each for each object attribute gives an information value, that is $\forall a\in A, x\in U, f(x,a)\in V_a$

2.4 Decision Tables

$A=C\bigcup D, C\bigcap D=\varnothing$, C Said condition attribute set, D said the decision attribute set. Decision table is a knowledge representation system with conditional attributes and decision attributes, as shown in Table 1.

Table 1. Decision table

U	C $c_1,c_2,\ldots c_p$	D $d_1,d_2,\ldots d_q$
u_1	$c_1(u_1),c_2(u_1),\ldots c_p(u_1)$	$d_1(u_1),d_2(u_1),\ldots d_q(u_1)$
u_2	$c_1(u_2),c_2(u_2),\ldots c_p(u_2)$	$d_1(u_2),d_2(u_2),\ldots d_q(u_2)$
u_3	$c_1(u_3),c_2(u_3),\ldots c_p(u_3)$	$d_1(u_3),d_2(u_3),\ldots d_q(u_3)$
…	…	…
u_n	$c_1(u_n),c_2(u_n),\ldots c_p(u_n)$	$d_1(u_n),d_2(u_n),\ldots d_q(u_n)$

In Table 1, Conditions subset of attributes $C'\subseteq C$, the importance about D is that:

$$\sigma_{CD}(C')=\kappa_C(D)-\kappa_{C-C'}(D) \tag{3}$$

In them, $\kappa_C(D)=|\mathrm{pos}_C(D)|/|U|$, D is dependent on the C .In particular, the time $C'=\{a\}$, $a\in C$ the importance of knowledge about D the properties as follows:

$$\sigma_{CD}(a)=\kappa_C(D)-\kappa_{C-\{a\}}(D) \tag{4}$$

Order X_i Representative U/C in the equivalence class; Y_j Representative U/D the equivalence classes of X_i description s with said $\mathrm{des}(X_i)$; Y_j pair described by representation $\mathrm{des}(Y_j)$.Decision rule is defined as follows:

r_{ij}:$\mathrm{des}(X_i) \rightarrow \mathrm{des}(Y_j)$, $Y_j \cap X_i \neq \varnothing$

When $\mu(X_i,Y_j)=|Y_j\cap X_i|/|X_i|, 0<\mu(X_i,Y_j)\leq 1$, r_{ij} is certain. When $\mu(X_i,Y_j)$, r_{ij} is not certain.

2.5 Distinguish and Differentiate Function Matrix

Let knowledge representation system $S=(U,A,V,f)$, $|U|=n\cdot S$ distinction matrix is a matrix $n\times n$, whose elements are:

$$\alpha(x,y)=\{a\in A \mid f(x,a)\neq f(y,a)\} \tag{5}$$

Therefore, the difference $\alpha(x, y)$ is a collection of x and y, all the objects and attributes. Δ Denoted by distinguishing function, for each attribute $a \in A$, given a Boolean variable a, if using Boolean $\alpha(x, y) = \{a_1, a_2, ..., a_k\} \neq \varnothing$, $\sum \alpha(x, y)$ function representation $a_1 \vee a \vee ... \vee a_k$; If the distinction $\alpha(x, y) = \varnothing$ between the function Δ can be defined as:

$$\Delta = \prod_{(x,y) \in U \times U} \sum \alpha(x, y) \tag{6}$$

3. Predict Consumer Behavior based on Rough Set of E-commerce

This paper, rough set theory to predict consumer behavior in e-commerce, the use of improved attribute reduction algorithm to delete redundant condition attributes of e-commerce consumer behavior condition attribute set reduction, the necessary condition attribute set; the improved forecast consumer behavior rules reduction algorithm to predict consumer behavior in e-commerce rules extraction and reduction, draw e-commerce consumer behavior decision rules. As a rough set analysis methods of data processing can be maintained in the case of knowledge classification ability, through the knowledge reduction, classification or decision rules derived problems [1].

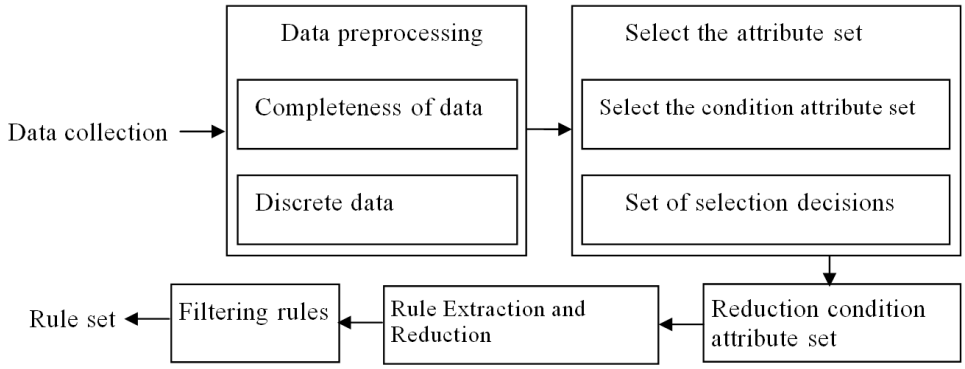


Figure 2. E-commerce Consumer Behavior Prediction based on Rough Set Theory

Predict the specific operation of e-commerce consumer behavior based on rough set theory shown in Figure 2, includes the following steps:

(1) By collecting consumption data for E-commerce, consumer completeness of the data and discrete, the completion of thee-commerce consumer behavior prediction basic data preprocessing;

(2) To remove redundant data e-commerce consumer behavior, the completion of the reduction of consumer behavior to predict the condition attributes set;

(3) Through thee-commerce consumer behavior prediction rules extraction and reduction, draw the necessary set of rules to predict consumer behavior;

(4) By calculating the rule of confidence and coverage, the rules filtered to give E-commerce decision rule to predict consumer behavior.

3.1 Data Preprocessing

Rough set theory to predict consumer behavior before the e-commerce data analysis, data must first be pretreated to collect valid data, the establishment of e-commerce consumer behavior prediction information table. Predict consumer behavior all the data

into a summary of information systems, information systems, also known as knowledge representation systems or attribute a value table, which can be represented by two-dimensional table. Value information table each row corresponds to an object and its attributes, the attribute values of each column corresponds to each object's attributes. The data needs to be processed and then aggregated information on the table.

Based on e-commerce and consumer behavior to predict the characteristics of the prediction process, the establishment of e-commerce knowledge of consumer behavior prediction tables, and then build e-commerce consumer behavior prediction system. Prior to this, the consumer behavior predicted parameter values are summarized in a table into knowledge. Summary of the e-commerce consumer behavior prediction information table format as shown in Table 2.

Table 2. E-commerce Consumer Behavior Prediction Information Summary

	r_1	r_2	r_3	……	r_m
t_1					
t_2					
t_3					
……					
t_{n-1}					
t_n					

Table 1 can use $K=\{U,R,V,f\}$ to express, Domain $U=\{t_1,t_2,...,t_n\}$, Set of attributes $R=\{r_1\ ,r_2,...,r_m\}$; in them r_i represents the i argument , i; Attributes and parameters and equivalence relations are corresponded each other. In the specific model of this paper, the properties and parameters Is regarded as one and the same concept; $V=\bigcup_{r\in R}V_r$ the range of attributes r is V_r . f can show as $U\times R\rightarrow V$, Which is a function of information, each attribute information are assigned a value, that is $\forall r\in R, x\in U, f(x,r)\in V_r$. In this paper, $K=\{U,R\}$ use as a simplified form to represent $K=\{U,R,V,f\}$, as the Table 2 shows ,on the field $U=\{t_1,t_2,t_3,t_4,t_5,t_6\}$, the condition attribute set $R=\{r_1,r_2,r_3,r_4\}$, which represent different attributes of different commodities . Let condition attribute set to $C=\{r_1,r_2,r_3\}$ = $\{a,b,c\}$; decision attribute set to D $=\{r_4\}$ $=\{d\}$

Table 3. Record Customer Spending Decision Table

	condition attributes			decision attribute
	r_1	r_2	r_3	r_4
t_1	Y	Y	N	Y
t_2	N	N	Y	N
t_3	Y	N	Y	Y
t_4	Y	Y	Y	N
t_5	N	Y	N	N
t_6	Y	Y	N	Y

3.2 Reduction Condition Attributes Set

Using the difference matrix method simplify attribute set , if there is a difference matrix a single attribute in the matrix element , you cannot find the reduction of the decision table, and therefore the article use an improved attribute reduction algorithm which is based on discernibility matrix to analysis e-commerce consumer behavior condition attributes. Firstly, by calculating the degree of dependency of decision attribute to condition attribute decision to make preliminary processing of the data in the table, and then take full advantage of the difference matrix, calculate the nuclear of decision table rapidly, and get a reasonable rule [2] by the attribute importance and credibility Reduction the value of the degree . Algorithm is described as follows:

(1) Enter $K = \{U, C \bigcup D, V, f\}$;

(2) Calculating a conditional attribute dependency $\gamma_a(D)$ $(a \in C)$, if $\gamma_a(D) = 0$, $C = C - \{a\}$;

(3) On the condition attribute set of $\gamma_a(D) \neq 0$, write the lower triangular matrix $M_{n\times n}(K) = (c_{ij})_{n\times n}$, where $i, j = 1, 2, ..., n$.

$$c_{ij} = \begin{cases} \{\alpha \mid (\alpha \in C) \wedge (f_\alpha(x_i) \neq f_\alpha(x_j))\}, f_D(x_i) \neq f_D(x_j), \\ \varnothing, f_D(x_i) \neq f_D(x_j) \wedge f_C(x_i) = f_C(x_j), \\ -, f_D(x_i) = f_D(x_j)。 \end{cases}$$

(4) Search for differences in the matrix, if the value of all the elements of the matrix are not equal $\varnothing$, then go to (4); If there is a value of the matrix element $\varnothing$, then exit;

(5) Search for differences in the matrix, and assign all of its single-property element $\mathrm{CORE}_C(D)$, output

$$\mathrm{CORE}_C(D) = \{\alpha \mid (\alpha \in C) \wedge (\exists c_{ij}, ((c_{ij} \in M_{n\times n}(K)) \wedge (c_{ij} = \{\alpha\})))\};$$

(6) Draw all possible combinations of attributes D contain relatively nucleus, if satisfied $\forall c_{ij} \in M_{n\times n}(K)$, when $c_{ij} \neq \varnothing$, $B \bigcap c_{ij} \neq \varnothing$; B independence. Then it is assigned to $\mathrm{RED}_C(D)$, and through all of the property portfolio contains D relatively nucleus;

(7)Output $\mathrm{RED}_C(D)$ calculate $\mathrm{RED}_C(D)$ the importance attributes $\sigma_{CD}(a) = \gamma_C(D) - \gamma_{C-\{a\}}(D)$, wherein, $a \in C$ if $\sigma_{CD}(a)$ >0.9, then $\mathrm{RED}'_C(D) \Leftarrow \mathrm{CORE}_C(D) \bigcup a$ traverse all the combinations of properties $\mathrm{RED}_C(D)$,calculate the reliability of $\mathrm{RED}'_C(D)$;

(8) Output $\mathrm{RED}'_C(D)$, the algorithm ends .

As shown in Table 4 attribute reduction based on discernibility matrix obtained by the improved algorithm.

Table 4. Difference Matrix

	1	2	3	4	5	6
1						
2	c					
3	c					
4		ac	ac			

5		ab	ab		
6	ac			c	b

3.3 Extraction and Reduction Rules

Consumer records above example, the rules generated by extraction and rules reduction obtained are as follows:

（buyr2）and（buyr3）$\Rightarrow$（buyr4）；

（buyr2）and（do not buyr3）$\Rightarrow$（do not buyr4）。

Among them, the algorithm described in consumer behavior prediction rules reduction are as follows:

(1) Output $K' = \{U, C \cup D, V, f\}$

(2) $B_0 = \mathrm{CORE}_C(D)$, $A = C - B_0 = \{\beta_1, \beta_2, ..., \beta_m\}$ $(\beta_i \in A, m \le \mathrm{card}(C), i = 1, 2, ..., m)$ Sorting according to the attribute of importance, namely strike OA, $\mathrm{T}_{l+1}(\mathrm{OA})$ and $\mathrm{OT}_{l+1}(\mathrm{OA})$ $(0 \le l \le m)$, $\mathrm{pos}_{B_0}(D)$ and $\mathrm{pos}_C(D)$;

(3) Determine equality. If equal, output $B_0 = \min\{\mathrm{RED}_C(D)\}$, go to (11); otherwise go to (4);

(4) If $i = 1, \mathrm{flag} = 0, Z, B, B_0$;

(5) If $Y = \mathrm{OT}_i(\mathrm{OA})$

(6) Take $y \in Y, B \Leftarrow B_0 \cup \{y\}$, calculate $\mathrm{pos}_B(D)$,and then determine whether $\mathrm{pos}_B(D)$ and $\mathrm{pos}_C(D)$ are equal, if $\mathrm{pos}_B(D) = \mathrm{pos}_C(D)$,and $\mathrm{flag} = 0$, then $Z = B$, $\mathrm{flag} = 1$;If $\mathrm{card}(U / Z) > \mathrm{card}(U / B)$, then $Z = B$, $\mathrm{flag} = 0$;

(7) $Y = Y - \{y\}$;

(8) If $Y \ne \varnothing$, turn to (6);

(9) If $\mathrm{flag} = 1$,then $\min\{\mathrm{RED}_C(D)\} = Z$,turn to (11)；

(10) $i = i + 1$, if $i \le m$, turn to (5)；

(11) Output $\min\{\mathrm{RED}_C(D)\}$, the algorithm ends.

4. Conclusion

The paper adopt an improved attribute reduction algorithm prediction condition attribute reduction algorithm which is based on discernibility matrix to analysis the set of e-commerce consumer behavior , and by the extraction and reduction rules generated draw a new theory which is based on rough set e-commerce consumer behavior prediction ,and this method achieves better practical results.

Acknowledgements

This work was supported by Educational Science and Technique Research Program of Heilongjiang Province (12541214), Heilongjiang postdoctoral scientific research projects (LBH-213127) and Doctoral Scientific research Project of Harbin University of Commerce (Analysis of e-commerce consumer behavior and its influencing factors).

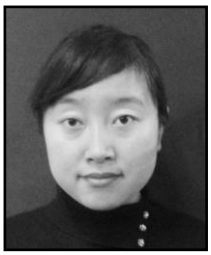
References

[1] Y. Zhang and C.-Y. Han, “Management of marketing channel in the environment of electronic commerce”, Journal of HIT (social sciences edition), vol. 1, no. 9, **(2007).**

[2] Y. Zhang, “The Research on the Model and Algorithm of Forest Disease and Pest Forecast based on the Rough Set Theory”, Northeast Forestry University, **(2012).**

[3] Z. H. Liu, Q. L. Zeng and Y. S. Li, "Research of equipment selection and matching expert system in fully mechanized caving face based on ontology [J]", Key Engineering Materials, **(2010),** pp. 419-420,117-120.
[4] L. F. Zhang and J. Ai, "Method for Bridge Bearing Capacity Assessment Based on Analytic Hierarchy Process [J]", Transactions of Nanjing University of Aeronautics & Astronautics, vol. 26, no. 3, **(2009**), pp. 236-241.
[5] Z. H. Liu, Q. L. Zeng and Y. S. Li, "Research of equipment selection and matching expert system in fully mechanized caving face based on ontology", Key Engineering Materials, **(2010).**
[6] X. Hao, D. Zhang, X. Liu and H. Zhao, "Algorithm and Application Research of Data Mining based on Rough Set Theory", Computer engineering and Applications, vol. 1, no. 9, **(2007).**
[7] F. Abbattista, M. Degemmis, P. Licehelli, G. Semeraro and F. Zambetta, "Improving the Usability of an E-commerce website through Personalization", In proceedings of the workshop on Recommendation and Personalization in Ecommerce, **(2002).**
[8] R. J. Kuo, J. L. Liao and C. Tu, "Integration of ART2 neural network and genetic k- means algorithm for analyzing web-browsing Paths in electronic commerce, Decision Surport Systems, vol. 40, **(2005),** pp. 355-374.
[9] D. Zhang and W. S. Lee, "A Web2based Question Answering System [A]", In Proceedings of the SMA Annual Symposium 2003, NUS, Singapore, **(2003)** January.
[10] Y. Li and D. Cui, "Improvement and application of MVC design patterns [J]", Jisuanji Gongcheng/Computer Engineering, **(2006),** pp. 62-64.

Authors



Zhang Yan-rong, she is a Doctor, Lecturer and teacher of School of the Computer and Information Engineering, Harbin University of Commerce. Her main research fields are E-commerce and Artificial Intelligence.

Zhao Zhi-jie, he Doctor, Professor, IEEE Member, His research fields are image processing and intelligence information processing.

