A novel approach to dynamic profiling of E-customers considering clickstream data and online reviews

Houda Zaim1, Adil Haddi2, Mohammed Ramdani3

1,3Informatics Department, LIM Laboratory, Faculty of Sciences and Techniques of Mohammedia, University Hassan II, Morocco
2Informatics Department, LAMSAD Laboratory, Superior School of Technology of Berrechid, University Hassan I, Morocco

Article Info

ABSTRACT

In this paper, we present an approach for mining change in customer’s behavior for the purpose of maintaining robust profiling model over time. Most of previous studies leave important questions unanswered: In developing B2C e-commerce strategies, how do managers implicitly load customer’s profiles based on their satisfaction over the online store characteristics? And: What kind of feedback segments do they have?. Our proposed approach does not force customers to explicitly express their preference information over the online service but rather capture their preference from their online activities. The challenge does not only lay in analyzing how customer’s classifier model change and when it does so but also to adapt it to the customer’s click stream data using a new decision tree generation algorithm which takes as inputs new set of variables; categorical, continuous and fuzzy variables. Customer’s online reviews rates are considered as classes. Experiments show that this work performed well in identifying relevant customer’s stream data to judge the chinese e-commerce website “Tmall”. The extracted values of the website’s features are also useful to identifying the satisfaction level when the customer’s rate is not available.

Keywords:
Customer satisfaction
Data mining
Decision tree
E-commerce website features
Fuzzy logic
Navigation session
Web tracking

Corresponding Author:
Houda Zaim,
Informatics Department,
LIM Laboratory–Faculty of Sciences and Techniques–Mohammedia,
Hassan II University of Casablanca,
BP 146 Mohammedia, 20650, Morocco.
Email: houdazaim@gmail.com

1. INTRODUCTION

The problem of load profiling and classifying customers has been studied extensively. Customer profiling approaches were used in e-commerce where systems keep a variety of customer information such as navigational and behavioral patterns using K-means algorithm for accurate profile construction [1], [2]. A. D. Rachid et al. [3] proposed a customer prediction model in an e-commerce context, wherein a clustering phase is based on the integration of k-means method and Length-Recency-Frequency-Monetary (LRFM) model. Girish S et al. [4] integrated classifier to predict the type of purchase that a customer would make, as well as the number of visits that he/she would make during a year. User profiles can also be created through multi-resolution clustering designed for smart metering data [5]. Smart meter technologies allow retailers to supervisor individual’s consumption in real time. Efforts have been made in mining those time series data for user profiling: Y. Wang et al. [6] used data gathered by Advanced Metering Infrastructure to better understand electrical consumption behavior. Q. Wang et al. [7] attempted to automate spike detection within large volume of smart meter data for load profiling and Y. Lu et al. [8] investigated the load profile clustering of...
smart grid customers using an adaptive weighted fuzzy clustering algorithm. In [9] authors developed a classification system to identify a specific user among several users by investigating different metadata characterizing load profiles, including raw measurements, frequency characterizations and typical load shape indexes. A. E. Frosta et al. [10] tried to reduce the data to be processed during the customer’s load profile. Typical Daily Profiles (TDP) and Typical Weekly Profiles (TWP) are compared to see how the time resolution of data affects the clustering. In [11] the profile is acquired by asking users explicitly to introduce and update their profile manually and next monitoring the browsing behavior. A framework to integrate user profiling and customer journey is proposed by authors of [12] using user interviews and semantic analysis to identify the target audience. Marketing actions are then set up for customer journey and finally conversion stage to convert user into customer.

Even though these researches reflect customer consumption behavior, it ignores data from customer’s click stream and online reviews rates. Moreover, though the research reflects various customer behavior data for creating customer profile and recommendation, judging and evaluating e-commerce website characteristics based on customer’s feedbacks is ignored.

For business e-commerce website’s evaluation, P. Dong [13] established some site evaluation index: information content index, site survey index and technical index. T. Hariguna et al. [14] aimed to analyze empirically three factors antecedents of trust they are system quality, information quality, and service quality. The results of this study concluded that those had positive impact on customer intention to purchase in e-commerce transactions on social media. T. Singh [15] identified factors related to usability of e-commerce website which are: user satisfaction simplicity, attractiveness, speed, efficiency, and searching product information. A survey used these factors as a valuable input from user for assessing the usability of e-commerce website. A framework was proposed by authors of [16] for evaluating the impact of four parameters on the success of e-commerce: customer satisfaction, costs, awareness & knowledge and infrastructure. Authors of [17] believed that all the success factors for supporting on-line shopping should be considered from multiple perspectives (technological perspective such as web design usability, social perspective such as social networks, etc.), across different e-commerce life cycle stages (pre-sale, information stage and supply chain).

Various e-commerce website attributes aimed for data collection are reviewed keeping in mind the areas where sequence of web events generated by each user is required to assess the online service quality. Therefore, we discuss general customer’s shopping steps in online stores. H. Sergio et al. [18] analyzed sections visited by users, navigational paths followed when accessing specific pages of the website, and the relation between different web sections or the sections that lead users to buy products. To improve user fulfillment and shopping experience, it has become a general practice for online sellers to allow their users to review or to communicate opinions of the products that they have sold. The major goal of the paper [19] was to solve evaluation feature extraction problem and opinion classification problem from customers using feature words and opinion words from product reviews. G. Silahtaroglu et al. [20] collected data about customer’s mouse movements, their demographic information and items they added to their shopping baskets. A global view on the sequential online-shopping events was proposed by authors of [21] where data sets consist of two parts: a collection of click events and events attached to corresponding clicks. R. Purwaningsih et al. [22] pointed out that factors affect consumer interest in online shopping are: Perceived Concentration, Perceived Enjoyment and Perceived Ease of Use.

Identifying these dimensions values implicitly based on e-customer click stream is our second task. The analysis of these data should be made to finally load the most accurate customer profile. In the next sections, we present details about the proposed customer profiling approach reflecting e-commerce website features, customer’s behavior and feedback, then we do mining on that dataset. And getting result in the form of Tmall website’s customer. Data are analyzed through a fuzzy approach to classify customers using a new decision tree algorithm by extending existing approaches [23], [24], [25], [26], [27] to profile generation, taking into account new set of evaluation variables required from customer’s feedback over the e-commerce website.

The classification task is characterized by well-defined classes and a training set of pre classified examples. Where the data to be classified is static, a simple classifier might suffice. However, our proposed approach requires classifier that is resilient to dynamic data such as time-related and click stream data taking online customer satisfaction level as predefined classes. The major innovation of the proposed approach is based on decision tree induction to obtain useful knowledge from large amounts of e-customers data to construct online customer profiles. The purpose of our study is to fill this gap, by conducting a two phases of training and inference. Training phase involves inducing a decision tree ensemble from customer reviews and click stream data. Inference phase uses the induced decision tree ensemble from training phase to classify customer implicitly based on their click stream as test data.

After providing an overview which treats together two different branches of research: online customer profiling techniques and selection based on customer’s shopping steps for load Profiles,
details regarding our proposed approach are provided in Section 2, followed by a case study in Section 3 conducted on a real user action data provided by Tmall, one of the largest B2C online retail platforms in China. The paper concludes in Section 4 with a summary of the work undertaken and direction of future works.

2. PROPOSED LOAD PROFILING METHODOLOGY

2.1. Framework for proposed model

The proposed algorithm for customer profile generation has the following steps, Figure 1:

Step 1: Collect web data (click stream and online reviews data) from e-commerce website.
Step 2: Apply preprocessing techniques to remove noise from data.
Step 3: Divide the data into 2 parts: training data (Online Customer Reviews and Click Stream Data) and testing data (Click Stream Data).
Step 4: Apply appropriate data mining techniques to create the model.
Step 5: Train the model as per given rule-based classifier.
Step 6: If the model is not trained, go to step 5.
Step 7: Test the model using test data.
Step 8: If model is accepted, use it for online customer profiling.

![Figure 1. An overview of the proposed methodology](image)

2.2. Metadata gathering and processing

User Profiling can be defined as the process of identifying data about a user interest domain. This information can be used by the system to enhance the retrieval of meeting the user’s needs. This part presents the dynamic of the proposed model using a communication diagram showing the interactions between the various components of the system. The goal is to determine graphs of data that will be subject to appropriate optimization algorithm based on customer behavior targeting.

2.2.1. Capturing dynamic customer’s behavior

In Figure 2 above communication starts with 1: \( t_s = \text{consult}_\text{website}() \) where \( t_s \) is the website’s consultation time without discovering any product. \( 2^* \): \( \text{find}_\text{products}() \) is an iterative message which could be repeated some unspecified number of times. Customer searches description of products (2.1: \( \text{search }(): t_p \)) with \( t_p \) is the consultation time of product. If customer is interested in some product(s) he consult its description (2.2 [interested]: \( \text{view}_\text{product}(): p_s \)), \( p_s \) refers to the consulted product(s). If client decides to buy, he add product to the basket -2.3 [decided to buy]: \( \text{add}_\text{to}_\text{basket}() \). 3: \( \text{Checkout}() \) includes getting list of
products from the online basket (3.1 get_products():\textit{p}'s), consulting bill (3.2: consult_bill():\textit{t}'s), updating the shopping cart(3.3: update_shopping_basket()) and purchasing (3.4[not empty (cart)]: make_order():\textit{p}''s), with \textit{p}''s refers to the purchased products.

![Communication diagram of the dynamic customer profiling for online retailing assessment](image)

### 2.2.2. Candidate entries variables

The starting and critical point for successful customer profiling based on navigational data is data preprocessing. The required high-level tasks are data cleaning, user identification, session identification, product view identification, etc. The proposed data preprocessing model for customer analysis tracks various online activities of customers to construct logical user sessions and create relevant entries variables for our proposed e-customer profiling system which take into account new set of variables; categorical, continuous and fuzzy variables for evaluating online service quality.

After the customer’s order is shipped, we detect its feedback information to assign values to Order Fulfillment as a categorical variable and Delivery Time which is numeric one. According to the Figure 2, we have three other variables: Efficiency, Content and Payment System Availability which are considered as fuzzy attributes as shown in Table 4.

### 2.2.3. Fuzzification

Restrictions placed on fuzzy variable’s values to define the possible changes are based on e-customer behavior during navigation sessions: \( S = S_{v} \cup S_{c} \cup S_{b} \cup S_{p} \) respectively, are visit sessions, consultation, basket session and purchase session [28].

\[
V_{ef} = \frac{1}{\text{card}(S)} \sum_{s \in S} (\varphi(t_{v}) \psi(s) + \Phi(s)) \text{ where:}
\]

\[
\varphi(t_{v}) = \begin{cases} 
1 & \text{if } t_{v} < t_{0} \\
0 & \text{else}
\end{cases}, \quad \psi(s) = \begin{cases} 
1 & \text{if } s \in S_{v} \\
0 & \text{if } s \in S_{c} \\
-1 & \text{if } s \in S_{b} \\
0 & \text{if } s \in S_{p}
\end{cases}
\]

The value of -1 indicates that no consultation/basket placement/Purchase was made after the visit; efficiency is therefore “Very Weak”. The value of 1 indicates that there was a consultation/basket placement/Purchase, so the website’s efficiency is considered as “Very Strong”. For the Content variable, the value of 1 indicates that there was a basket placement after the consultation of product’s information. The value of -1 indicates that there is no basket placement.
\[ V_{\text{Content}} = \frac{1}{\text{card}(S_c)} \sum_{s \in S_c} \left( \frac{1}{\text{card}(P_s)} \sum_{p \in P_s} \varphi'(t_p) \psi'(p) + \phi'(p) \right) \]

where:

\[ \varphi'(t_p) = \begin{cases} 1 & \text{if } t_p < t'_0, \\ 0 & \text{else} \end{cases} \]

\[ \psi'(p) = \begin{cases} 1 & \text{if } p \in P'_s, \\ 0 & \text{else} \end{cases} \]

\[ \Phi'(p) = \begin{cases} 0 & \text{if } p \in P'_s, \\ -1 & \text{else} \end{cases} \]

The Payment of a command is very weak if no purchase was made or product(s) added to the online basket are deleted. The value of 1 indicates that there was a successful purchase:

\[ V_{\text{Payment}} = \frac{1}{\text{card}(S_b)} \sum_{s \in S_b} \left( \frac{1}{\text{card}(P'_s)} \sum_{p \in P'_s} \varphi''(t_p) \psi''(p) + \phi''(p) \right) \]

Membership function (MF) is defined for finding membership value for each of the fuzzy inputs to five levels: very weak, weak, medium, strong and very strong. Figure 3 shows membership function that we use in our method:

![Figure 3. MFs used to present the linguistic labels of: (a) Efficiency (b) Content (c) Payment System](image)

**2.3. Proposed data mining approach**

The proposed approach consists of two phases of training and inference as outlined in Figure 4. In training phase, online customer’s reviews and click stream data are processed and a decision tree ensemble derived for change mining. The inference phase involves using the classifier derived from the training phase to dynamic profiling of online customers. The decision tree generation algorithm mentioned above present our proposed decision tree feature selection strategy adapted to the evolving data.

![Figure 4. Change mining of customer profile based on navigational data](image)
2.3.1. Learning phase

Given training set T containing n samples belonging to k classes \( \{ c_1, c_2, \ldots, c_k \} \) \((1 \leq k \leq p)\). Let \( A_q \) be the attribute that may influence consumer’s satisfaction, its value set \( \{ x_1, x_2, \ldots, x_v \} \), \( v \) is the number of values. The dissimilarity \( \text{dis}(x_u, x_v) \) between any two \( A_q \) values for the class \( C_k \) can be computed by:

\[
\text{dis}(x_u, x_v) = \begin{cases} 
\frac{|x_u - x_v|}{\text{Max}(A_q) - \text{Min}(A_q)} & \text{if } A_q \text{ is numeric} \\
0 \text{ or } 1 & \text{if } A_q \text{ is categorical} \\
\text{d}(x_u, x_v) & \text{if } A_q \text{ is fuzzy}
\end{cases}
\]

\( \text{d}(x_u, x_v) = \left( \sum_{i=1}^{v} (x_{ui} - x_{vi})^2 \right)^{1/2} \) is a euclidean distance where \( x_{ui} \) and \( x_{vi} \) represent the vector of belonging to fuzzy objects \( \{ a_1, a_2, \ldots, a_5 \} \) which are represented by functions mapping the fuzzy attribute scale (Figure 3), \( x_u = x_{u1}, x_{u2}, \ldots, x_{u5} \)

The average similarity of \( A_q \) for the class \( C_k \) can be computed by equation expressed by \( \text{sim} \left(A_q, C_k\right)\):

\[
\text{sim} \left(A_q, C_k\right) = \frac{\sum_{x_u \in C_k} \text{dis}(x_u, x_v)}{|C_k| \times (|C_k| - 1)} \in [0, 1]
\]

\( |C_k| \) is the size of \( C_k \). If \( |C_k| = 1 \), then \( \text{sim} \left(A_q, C_k\right) = 1 \).

The global similarity of the attribute \( A_q \) is the sum of the similarity of each subset expressed by \( \text{SIM}(A_q) \), which can be defined as:

\[
\text{SIM}(A_q) = \sum_{k=1}^{p} \left( P_k \times \text{sim}(A_q, C_k) \right) \in [0, 1]
\]

The Set of class values is \( \{ C_1, C_2, \ldots, C_k \} \), their probability is \( \{ P_1, P_2, \ldots, P_k \} \).

Node Selection (T, \( A_q \)):

For each \( A_q \) in attribute_list do

\( \text{SIM}(A_q) = 0 \)

For each class \( C_k \) do

\( \text{SIM}(A_q) = \text{SIM}(A_q) + \text{sim}(A_q, C_k) \times P_k \)

End for

End for

Return the attribute \( A_q \) with the maximum similarity

Calculate Splitting Threshold (T, \( A_q \)):

Sort the dataset T by numeric attribute’s values

\( \text{max} = 0 \)

\( A_q \) is the numeric attribute

For each value \( x_u \)

Calculate the average similarity of \( A_q \) for two subsets \( T_u \) that divided by \( x_u \)

If \( \text{max} < \text{SIM}(A_q) \)

\( \text{max} = \text{SIM}(A_q) \)

Splitting Value = \( x_u \) where \( \text{SIM}(A_q) \) is maximum

End if

End for

Return \( x_u \)

Figure 5 shows provides a better understanding of decision tree induction process in data stream scenario, particularly focusing on the mathematical foundations of choosing the root and splitting continuous, categorical and fuzzy criteria in decision tree nodes.


<table>
<thead>
<tr>
<th>Generate_Decision_Tree (T, A_q, Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inputs:</strong> Training data: T, set of continuous, categorical and fuzzy valued attributes: A_q, Target attribute: Y</td>
</tr>
<tr>
<td><strong>Output:</strong> A decision tree</td>
</tr>
<tr>
<td><strong>Start</strong></td>
</tr>
<tr>
<td>If T is empty Then</td>
</tr>
<tr>
<td>Finish the function without building a node</td>
</tr>
<tr>
<td>End if</td>
</tr>
<tr>
<td>If T is made only for the values of the same target Then</td>
</tr>
<tr>
<td>Return a single node of this value</td>
</tr>
<tr>
<td>End if</td>
</tr>
<tr>
<td>If A_q is empty Then</td>
</tr>
<tr>
<td>Return a single node with value as the most common value of the target attribute values in T</td>
</tr>
<tr>
<td>End if</td>
</tr>
<tr>
<td>D ← Node_Selection (T, A_q)</td>
</tr>
<tr>
<td>//D is the root of the tree</td>
</tr>
<tr>
<td>If D is continuous type Then</td>
</tr>
<tr>
<td>$x_u$ ← Calculate splitting threshold (T, A_q)</td>
</tr>
<tr>
<td>End if</td>
</tr>
<tr>
<td>//Grow the tree and partition the data using the continuous split</td>
</tr>
<tr>
<td>{x_1, x_2, \ldots, x_v} ← Values of the attribute D</td>
</tr>
<tr>
<td>Return the constructed tree whose root is D and the arcs are labeled by $x_1, x_2, \ldots, x_v$</td>
</tr>
<tr>
<td>Generate_Decision_Tree for sub-trees (T_1, A_q-{D}, Y), (T_2, A_q-{D}, Y), \ldots, (T_v, A_q-{D}, Y)</td>
</tr>
<tr>
<td>End</td>
</tr>
</tbody>
</table>

**Figure 5.** Pseudo code of decision tree induction algorithm

### 2.3.2. Classification of a new point

To classify a new input point, we simply traverse down the tree. At each node in the decision tree, we ask a question about our data point. For categorical and numeric attributes, if the condition at the node is met, we go left; if it is not true, then we go right. For fuzzy criteria we calculate the value of $d(x_u, x_v)$ translating the membership possibility of the new input point $x_u$ to each fuzzy subspace $x_v$ in the tree. The best mark $d(x_{u'}, x_{v'})$ in a subspace is the mark with the minimum value; it identifies the fuzzy split value which responds the condition at the node in question.

### 3. VALIDATION METHODOLOGY AND RESULTS

#### 3.1. Dataset

We choose reviews and click stream data from TMALL as the data source. TMALL is an important business unit of Alibaba Group which is known as the top one B2C platform in China. The user’s behavior of browsing TMALL reflects their preference of items. This data set contains 25432915908 records of user-item interactions. Features of each row are listed in Table 1. Table 2 presents review data for partial "user-item" pairs, which contains the review and rating on the item/merchant/logistic. This data set contains 241919749 rows, corresponding to 241919749 reviews.

Table 4 describes used training data where class is deduced from online reviews which are filtered by 5 starts referring to customer satisfaction rate as mentioned in Table 3. For each class of each customer, restrictions can be placed on website’s features values to define their possible changes based on e-customer behavior during navigation sessions as mentioned in Fuzzy Rules Design section. These features will be classified into 5 linguistic terms using click stream data into membership function values mentioned in Figure 3 e.g. (0.0/VW Very Weak+0.0/W Weak+0.5M/Medium+0.5/S Strong+0/VS Very Strong). Customers who don’t describe their feedback were chosen as testing set as shown in Table 5.
A novel approach to dynamic profiling of e-customers considering click stream...

### Table 1. User-Item Interaction

<table>
<thead>
<tr>
<th>Item</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Used_id</td>
<td>A string as &quot;u0774184&quot;, denoting an unique user</td>
</tr>
<tr>
<td>Item_id</td>
<td>An integer in [1, 8133507], denoting an unique item</td>
</tr>
<tr>
<td>Action</td>
<td>Type of behavior, a string like &quot;click&quot;, &quot;collect&quot;, &quot;cart&quot;, &quot;alipay&quot;, represents for 'click', 'add to favorite', 'add to cart' and 'purchase', respectively</td>
</tr>
<tr>
<td>Vtime</td>
<td>Timestamp of user’s behavior</td>
</tr>
</tbody>
</table>

### Table 2. User-Reviews

<table>
<thead>
<tr>
<th>Item</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Used_id</td>
<td>A string as &quot;u0774184&quot;, denoting an unique user</td>
</tr>
<tr>
<td>Item_id</td>
<td>An integer in [1, 8133507], denoting an unique item</td>
</tr>
<tr>
<td>Feedback</td>
<td>A string containing multiple key words, separated by ',' . There words are extracted from the raw title by an NLP system</td>
</tr>
<tr>
<td>Rate_pic_url</td>
<td>An URL linked to corresponding image online</td>
</tr>
<tr>
<td>Gmt_create</td>
<td>Timestamp of the review, A string as &quot;yyyy-mm-dd hh:mm:ss&quot;</td>
</tr>
</tbody>
</table>

### Table 3. Tmall Data Sampling

<table>
<thead>
<tr>
<th>User Id</th>
<th>Item Id</th>
<th>User Action</th>
<th>V-Time</th>
<th>Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>u41</td>
<td>i61</td>
<td>Click</td>
<td>30/09/2014 15:19:00</td>
<td>★★★★★</td>
</tr>
<tr>
<td>i534</td>
<td></td>
<td>Alipay</td>
<td>03/09/2014 15:46:01</td>
<td></td>
</tr>
<tr>
<td>u57</td>
<td>i135</td>
<td>Cart</td>
<td>26/09/2014 20:32:01</td>
<td>★★★</td>
</tr>
<tr>
<td>u1209</td>
<td>i110</td>
<td>Alipay</td>
<td>01/09/2014 22:28:29</td>
<td>★★★</td>
</tr>
<tr>
<td>u1641</td>
<td>i109</td>
<td>Alipay</td>
<td>02/09/2014 11:50:52</td>
<td>★★★</td>
</tr>
</tbody>
</table>

### Table 4. Descriptive Example of Used Training Data

<table>
<thead>
<tr>
<th>User Id</th>
<th>Class</th>
<th>Efficiency</th>
<th>Content</th>
<th>Payment System</th>
<th>Order</th>
<th>Deli-V</th>
<th>Delivery Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>VS S M W V</td>
<td>V S</td>
<td>VS S M W V W</td>
<td></td>
<td>V W</td>
<td></td>
</tr>
<tr>
<td>u48</td>
<td>Satisfied</td>
<td>0.5 0.5 0 0</td>
<td>0 0 0 1 0</td>
<td>0 0 0 0 0 0 0 0</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>u41</td>
<td>Very Satisfied</td>
<td>1 0 0 0 0 0</td>
<td>1 0 0 0</td>
<td>1 0 0 0 0 0 0 0</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>u1725</td>
<td>Moderately Satisfied</td>
<td>1 0 0 0 0 0</td>
<td>1 0 0 0</td>
<td>0 0 0 0 0 0 1</td>
<td>0 0</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>u2502</td>
<td>Moderately Satisfied</td>
<td>1 0 0 0 0</td>
<td>1 0 0 0</td>
<td>0 0 0 0 0 0 1</td>
<td>0 0</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>u394</td>
<td>Not Satisfied</td>
<td>0.8 0.2 0 0</td>
<td>0 0 0 1 0</td>
<td>0 0 0 0 0 0 1</td>
<td>0 0</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>u50</td>
<td>Very Satisfied</td>
<td>1 0 0 0 0 0</td>
<td>0 0 0 0</td>
<td>0 0 0 0 0 0 0</td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>u490</td>
<td>Moderately Satisfied</td>
<td>0.6 0.4 0 0</td>
<td>0 0 0 1</td>
<td>0 0 0 0 1 0 0</td>
<td>0 0</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>u1675</td>
<td>Somewhat Satisfied</td>
<td>1 0 0 0 0</td>
<td>0 0 0 1</td>
<td>0 0 0 0 0 0 1</td>
<td>0 0</td>
<td></td>
<td>10</td>
</tr>
</tbody>
</table>

### Table 5. Tmall Testing Data

<table>
<thead>
<tr>
<th>User Id</th>
<th>Item Id</th>
<th>User Action</th>
<th>V-Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>u39</td>
<td>i46</td>
<td>Click</td>
<td>03/09/2014 21:01</td>
</tr>
<tr>
<td>i35</td>
<td></td>
<td>Cart</td>
<td>29/09/2014 11:17</td>
</tr>
<tr>
<td>i366</td>
<td></td>
<td>Alipay</td>
<td>04/09/2014 19:23</td>
</tr>
<tr>
<td>i34</td>
<td></td>
<td>Click</td>
<td>06/09/2014 20:24</td>
</tr>
<tr>
<td>u3286</td>
<td>i152</td>
<td>Collect</td>
<td>27/09/2014 0:28</td>
</tr>
<tr>
<td>u2984</td>
<td>i41</td>
<td>Cart</td>
<td>24/09/2014 21:36</td>
</tr>
<tr>
<td>i381</td>
<td></td>
<td>Click</td>
<td>18/09/2014 11:22</td>
</tr>
</tbody>
</table>

### 3.2. Programming setup

The deployment procedure is performed on a system among Intel Core i7, 8GB memory, along with Windows 7 system. Here, the method implemented in JAVA using Eclipse neon.3 with CSV Files. The Proposed algorithm is calculated with numerous kinds of customer’s dataset to estimate the effectiveness of
proposed approach. We ran online TMALL customer’s reviews and click stream data through step 1 and step 2 of the process outlined in Figure 4. Data shown in Table 4 were randomly chosen as training set. This data set contains rows corresponding to records of user-item interactions and reviews, considering customer satisfaction degree as target attribute and TMALL website’s characteristics as non-target attributes.

The simulation results are summarized in Figure 6. The algorithm takes as output a tree that resembles to an orientation diagram where each end node (leaf) is a decision (a class) and each non-final node (internal) represents a test. At each node in the decision tree, we ask a question about our in store navigational data point considered as input data as shown Table 5.

The simulation results are summarized in Figure 6. The algorithm takes as output a tree that resembles to an orientation diagram where each end node (leaf) is a decision (a class) and each non-final node (internal) represents a test. At each node in the decision tree, we ask a question about our in store navigational data point considered as input data as shown Table 5.

Figure 6. A part of induced decision rules

3.3. Performance and discussion

Our methodology enhances e-commerce website’s feature extraction and customer’s opinion classification. As the results show in Figure 6, the tests carried on numeric, categorical and fuzzy features show that the proposed algorithm select appropriate threshold for stopping growth respecting the three types of features and gives good classification rates for the shortest computing times. Feature selection reduces the dimensions of problem, but also improves customer’s classification performance by discarding noise, redundant, and unimportant features.

First, references such us [29] were analyzed for click stream data usage, they didn't involve fuzzification of those data based on online customer navigation sessions. Moreover in the area of value creation in an e-business model, such an approach also didn't attempt to create value for customers based on their navigational pathways. Second, comparing reference [30] selection features of success in e-business, our research fuzzy mine click stream and use it as a key to select and judge features properly characterizing the global online store performance. Third, several references like [19] conducted research for online products which are relevant in terms of having high sales, while this paper research for ensure online customer’s satisfaction. Using the proposed tree induction technique, marketing rules can be generated to match customer to satisfaction categories. For [31], classification decision tree algorithm has an input training dataset which consists of a number of attributes which are either categorical or continuous. Dataset used in the current paper is from the public information in Tmall. More types of data are available in our approach for learning decision tree classifier which can handle categorical, discrete, continuous and fuzzy attributes.

4. CONCLUSIONS AND FUTURE WORK

Modeling the user’s interests is a challenging task. Our work draw a line between the actual user interest and the acquired user profile by inducing a decision tree ensemble from customer’s behavior and
classifying customers based on their navigation sessions and reviews over the online service that may be used. The case study shows that by fitting the model variables and variable’s restrictions, and by using our classification model, an accurate customer profile can be achieved with regards to assessment of e-service quality. As future works, our approach can be could be enhanced and extended in several directions:

a. Whereas explicit customer’s click data is collected for statistical analysis, data for textual customer’s reviews need a pre-processing process. We can attempt to explore how text mining could be applied to mine and summarize customer’s reviews. This deserves more study.

b. Compare the proposed customer profiling approach to our proposed multi criteria classification model [32] to show promising results for classification techniques adoption.

c. People tend to take experienced customer’s opinion before making their own purchase. Thus, an implicit user profiling through social information discovery can be proposed to adequately capture the user’s interests.

REFERENCES


