

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

Houda Zaim¹, Adil Haddi², Mohammed Ramdani³

^{1,3}Informatics 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

Article history:

Received Mar 1, 2018

Revised Aug 4, 2018

Accepted Dec 18, 2018

Keywords:

Customer satisfaction

Data mining

Decision tree

E-commerce website features

Fuzzy logic

Navigation session

Web tracking

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.

Copyright © 2019 Institute of Advanced Engineering and Science.
All rights reserved.

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. Silaharoglu 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 features 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:

Step1: Collect web data (click stream and online reviews data) from e-commerce website.

Step2: Apply preprocessing techniques to remove noise from data.

Step3: Divide the data into 2 parts: training data (Online Customer Reviews and Click Stream Data) and testing data (Click Stream Data).

Step4: Apply appropriate data mining techniques to create the model.

Step5: Train the model as per given rule-based classifier.

Step6: If the model is not trained, go to step5.

Step7: Test the model using test data.

Step8: If model is accepted, use it for online customer profiling.

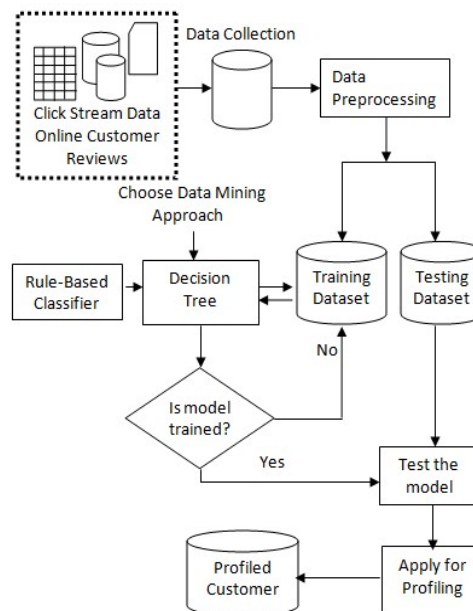


Figure1. An overview of the proposed methodology

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: $t_s = \text{consult_website}()$ where t_s is the website's consultation time without discovering any product. 2*: $\text{find_products}()$ 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]: $\text{view_product}():p_s$), p_s refers to the consulted product(s). If client decides to buy, he add product to the basket -2.3 [decided to buy]: $\text{add_to_basket}()$. 3: $\text{Checkout}()$ includes getting list of

products from the online basket (3.1 get_products():p"s), consulting bill (3.2: consult_bill():t"p), updating the shopping cart(3.3: update_shopping_basket()) and purchasing (3.4[not empty (cart)]: make_order():p"s),with p"s refers to the purchased products.

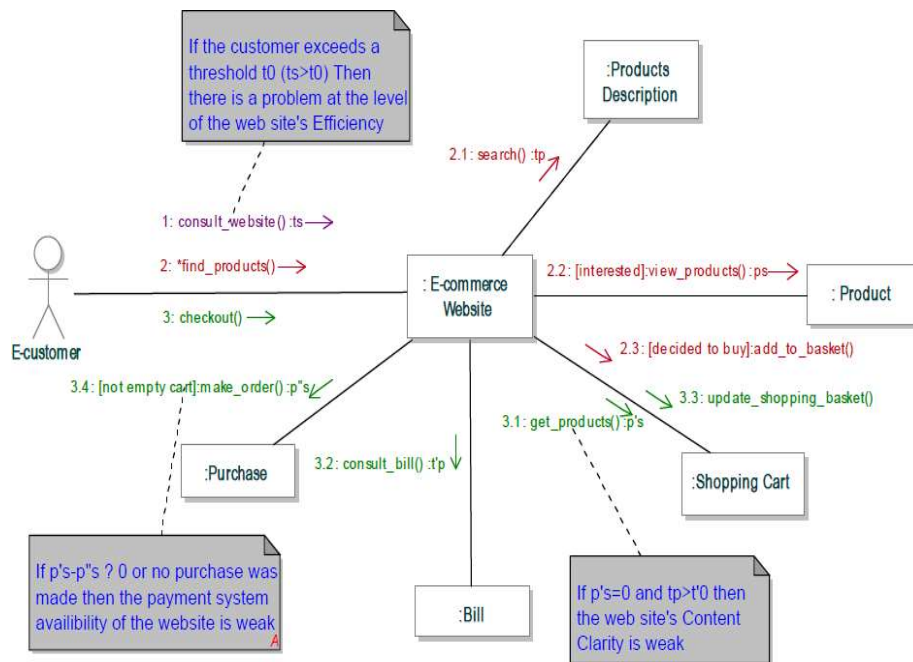


Figure2. Communication diagram of the dynamic customer profiling for online retailing assessment

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].

$$V_{Ef} = \frac{1}{card(S)} \sum_{s \in S} (\varphi(t_s) \psi(s) + \Phi(s)) \text{ where:}$$

$$\varphi(t_s) = \begin{cases} 1 & \text{if } t_s < t_0 \\ 0 & \text{else} \end{cases}, \quad \psi(s) = \begin{cases} 1 & \text{if } s \in S - S_v \\ 0 & \text{if } s \in S_v \end{cases}, \quad \Phi(s) = \begin{cases} -1 & \text{if } s \in S_v \\ 0 & \text{if } s \in S - S_v \end{cases}$$

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:

$$V_{Content} = \frac{1}{card(S_c)} \sum_{s \in S_c} \left(\frac{1}{card(P_s)} \sum_{p \in P_s} \varphi'(t_p) \psi'(p) + \Phi'(p) \right) \text{ where:}$$

$$\varphi'(t_p) = \begin{cases} 1 & \text{if } t_p < t'_0 \\ 0 & \text{else} \end{cases}, \quad \psi'(p) = \begin{cases} 1 & \text{if } p \in P'_s \\ 0 & \text{if } p \in P_s - P'_s \end{cases}, \quad \Phi'(p) = \begin{cases} 0 & \text{if } p \in P'_s \\ -1 & \text{if } p \in P_s - P'_s \end{cases}$$

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:

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

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:

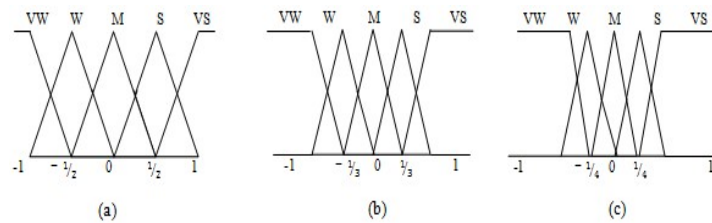


Figure 3. MFs used to present the linguistic labels of: (a) Efficiency (b) Content (c) Payment System Availability

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.

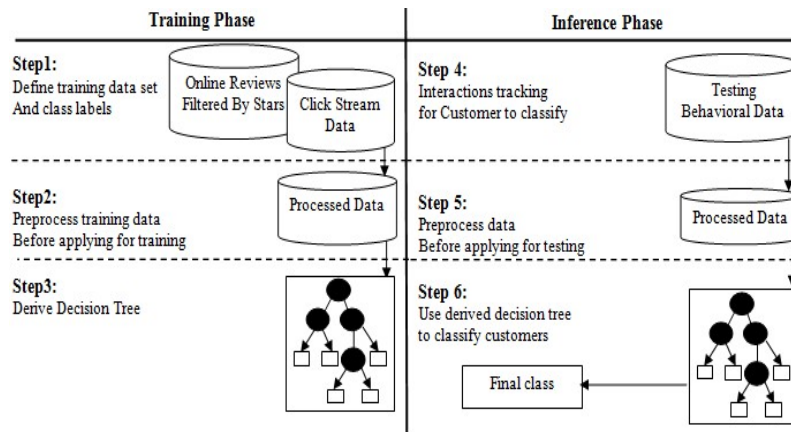


Figure 4. Change mining of customer profile based on navigational data

2.3.1. Learning phase

Given training set T containing n samples belonging to k classes $\{c_1, c_2, \dots, c_k\}$ ($1 \leq k \leq p$). Let A_q be the attribute that may influence consumer's satisfaction, its value set $\{x_1, \dots, x_u, \dots, x_v\}$, 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:

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

$d(x_u, x_v) = \left(\sum_{i=1}^5 (x_{ui} - x_{vi})^2 \right)^{1/2}$ is a euclidean distance where x_{ui} and x_{vi} represent the vector of belonging to fuzzy objects $\{a_1, \dots, a_i, \dots, a_5\}$ which are represented by functions mapping the fuzzy attribute scale (Figure 3), $x_u = x_{u1}, x_{u2}, \dots, x_{u5}$

The average similarity of A_q for the class C_k can be computed by equation expressed by $\text{sim}(A_{qC_k})$:

$$\text{sim}(A_{qC_k}) = \frac{2 * \sum_{x_u \in C_k, x_v \in C_k} \frac{1}{|C_k| * (|C_k| - 1) + \text{dis}(x_u, x_v)}}{|C_k| * (|C_k| - 1)} \in [0, 1]$$

$|C_k|$ is the size of C_k . If $|C_k| = 1$, then $\text{sim}(A_{qC_k}) = 1$.

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:

$$\text{SIM}(A_q) = \sum_{k=1}^p (P_k * \text{sim}(A_{qC_k})) \in [0, 1]$$

The Set of class values is $\{C_1, C_2, \dots, C_k\}$, their probability is $\{P_1, P_2, \dots, P_k\}$.

Node_Selection (T, A_q):

For each A_q in attribute_list do

SIM(A_q) = 0

For each class C_k do

*SIM(A_q) = SIM(A_q) + sim(A_{qC_k) * 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

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 max < SIM(A_q)

max = SIM(A_q)

Splitting Value = x_u where SIM(A_q) is maximum

End if

End for

Return x_u

Figure 5 shows provides a better understandin 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.

```

Generate_Decision_Tree (T,Aq, Y)
Inputs: Training data: T, set of continuous, categorical and fuzzy valued attributes: Aq, Target attribute: Y
Output: A decision tree
Start
  If T is empty Then
    Finish the function without building a node
  End if
  If T is made only for the values of the same target Then
    Return a single node of this value
  End if
  If Aq is empty Then
    Return a single node with value as the most
    common value of the target attribute values in T
  End if
  D ← Node_Selection (T,Aq)
  //D is the root of the tree
  If D is continuous type Then
    xu ← Calculate splitting threshold (T, Aq)
  End if
  //Grow the tree and partition the data using the continuous split
  {x1,x2,... xv} ← Values of the attribute D
  Return the constructed tree whose root is D and the arcs are labeled by x1,x2,... xv
  Generate_Decision_Tree for sub-trees (T1, Aq-{D}, Y), (T2, Aq-{D}, Y), ..., (Tv, Aq-{D}, Y)
End

```

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 stars 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.

Table 1. User-Item Interaction

Item	Definition
Used_id	A string as "u9774184", denoting an unique user
Item_id	An integer in[1, 8133507], denoting an unique item
Action	Type of behavior, a string like "click", "collect", "cart", "alipay", represents for 'click', 'add to favorite', 'add to cart' and 'purchase', respectively
Vtime	Timestamp of user's behavior

Table 2. User-Reviews

Item	Definition
Used_id	A string as "u9774184", denoting an unique user
Item_id	An integer in[1, 8133507], denoting an unique item
Feedback	A string containing multiple key words, separated by ' '. These words are extracted from the raw title by an NLP system
Rate_pic_url	An URL linked to corresponding image online
Gmt_create	Timestamp of the review, A string as "yyyy-mm-dd hh:mm:ss"

Table 3. Tmall Data Sampling

User Id	Item Id	User Action	V-Time	Feedback
u41	i161	Click	30/09/2014 15:19:00	★★★★★
	i534	Alipay	03/09/2014 15:46:01	
u57	i135	Cart	26/09/2014 20:32:01	★★★
u1209	i110	Alipay	01/09/2014 22:28:29	★★★
u1641	i109	Alipay	02/09/2014 11:50:52	★★

Table 4. Descriptive Example of Used Training Data

User-Id	Class	Efficiency				Content					Payment System Availability					Order Fulfillment	Delivery Time	
		VS	S	M	W	V	W	S	M	W	V	W	VS	S	M			W
u48	Satisfied	0.5	0.5	0	0	0	0	0	1	0	0	0.6	0.4	0	0	0	Available-To-Promise	3
u41	Very Satisfied	1	0	0	0	0	0	0	1	0	0	1	0	0	0	Available-To-Promise	3	
u1725	Moderately Satisfied	1	0	0	0	0	0	1	0	0	0	0.7	0.3	0	0	Available-To-Promise	10	
u2502	Moderately Satisfied	1	0	0	0	0	1	0	0	0	0	0	0	1	0	Available-To-Promise	20	
u394	Not Satisfied	0.8	0.2	0	0	0	0	1	0	0	0	0	0	1	0	Not Available-To-Promise	20	
u50	Very Satisfied	1	0	0	0	0	0	0	0.5	0.5	1	0	0	0	0	Available-To-Promise	6	
u490	Moderately Satisfied	0.6	0.4	0	0	0	0	1	0	0	0	1	0	0	0	Available-To-Promise	8	
u1675	Somewhat Satisfied	1	0	0	0	0	0	1	0	0	0	0	0	1	0	Not Available-To-Promise	10	

Table 5. Tmall Testing Data

User Id	Item Id	User Action	V-Time
u39	i146	Click	03/09/2014 21:01
	i53	Cart	29/09/2014 11:17
	i366	Alipay	04/09/2014 19:23
	i34	Click	06/09/2014 20:24
u3286	i152	Collect	27/09/2014 0:28
u2984	i41	Cart	24/09/2014 21:36
	i381	Click	18/09/2014 11:22

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

A novel approach to dynamic profiling of e-customers considering click stream... (Houda Zaim)

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.

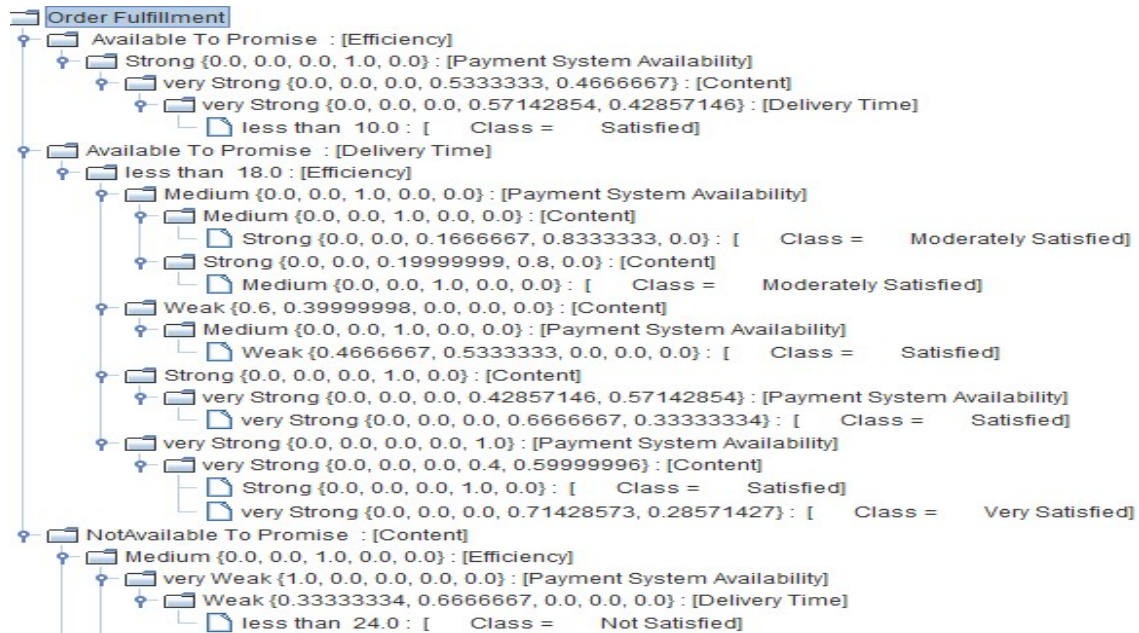


Figure 6. A part of induced decision rules

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 as [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 FUTUR 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

- [1] C. Mihai, C. Popa and P. M. Mircea, "Load Profiling For Gas Stations Using Cluster Techniques", *IEEE International Power Electronics and Motion Control Conference (PEMC)*, Varna, pp. 1041-1048, 2016.
- [2] J. Yang, C. Liu, M. Teng, M. Liao and H. Xiong, "Buyer Targeting Optimization: A Unified Customer Segmentation Perspective", *IEEE International Conference on Big Data (Big Data)*, Washington, DC, pp. 1262-1271, 2016.
- [3] A. D. Rachid, A. Abdellah, B. Belaid, L. Rachid, "Clustering Prediction Techniques in Defining and Predicting Customers Defection: The Case of E-Commerce Context", *International Journal of Electrical and Computer Engineering (IJECE)*, ISSN: 2088-8708. Vol 8, No 4: August, <http://doi.org/10.11591/ijece.v8i4.pp%25p>, 2018.
- [4] Girish S, Ramamurthy B, Senthilnathan T, "Mining the Web Data for Classifying and Predicting Users' Requests", *International Journal of Electrical and Computer Engineering (IJECE)*, ISSN: 2088-8708. Vol 8, No 4, 2018.
- [5] Ran Li, Furong Li, Nathan D. Smith, "Multi-Resolution Load Profile Clustering for Smart Metering Data", in *IEEE Transactions on Power Systems*, vol. 31, no. 6, pp. 4473-4482, Nov. 2016. doi: 10.1109/TPWRS.2016.2536781
- [6] Y. Wang, Q. Chen, C. Kang, M. Zhang, K. Wang, Y. Zhao, "Load Profiling and Its Application to Demand Response: A Review", in *Tsinghua Science and Technology*, vol. 20, no. 2, pp. 117-129, April 2015.
- [7] Q. Wang, X. Yu, P. Chou, D. Savage, X. Zhang, "Power Usage Spike Detection using Smart Meter Data for Load Profiling", *IEEE 25th International Symposium on Industrial Electronics (ISIE)*, pp. 732-737, 2016.
- [8] Y. Lu, T. Zhang, Z. Zeng, "Adaptive Weighted Fuzzy Clustering Algorithm for Load Profiling of Smart Grid Customers", *IEEE/CIC International Conference on Communications in China*, Chengdu, pp. 1-6, 2016.
- [9] M. Bicego, A. Farinelli , E. Grosso , D. Paolini , S.D. Ramchurn, "On The Distinctiveness Of The Electricity Load Profile". Volume 74, Pages 317-325, February 2018.
- [10] A. E. Frosta , M. Azazaa , H. Lia, F. Wallin,"On The Distinctiveness Of The Electricity Load Profile", 2017.
- [11] K. Gatzliolis, A. C. Boucouvalas, "User Profile Extraction Engine", *Pan-Hellenic Conference on Informatics*, 2016.
- [12] L. Boss, F. Müller, A. Selimi, A. Rossmann, "Converting Users Into Customers: The Role Of User Profile Information And Customer Journey Analysis", *Digital Enterprise Computing*, 2016.
- [13] P. Dong, "Research On The Evaluation Of E-Commerce Website Under The Environment Of Big Data", 2015.
- [14] T. Hariguna, Berlilana, "Understanding of Antecedents to Achieve Customer Trust and Customer Intention to Purchase E-Commerce in Social Media, an Empirical Assessment", *International Journal of Electrical and Computer Engineering (IJECE)*, ISSN: 2088-8708. Vol. 7, No. 3, pp. 1240~1245, June 2017.
- [15] T. Singh, S. Malik, D. Sarkar, "E-Commerce Website Quality Assessment based on Usability", ICCCA 2016.
- [16] M. Choshin, A. Ghaffari, "An Investigation Of The Impact Of Effective Factors On The Success Of E-Commerce In Small- And Medium-Sized Companies", <https://doi.org/10.1016/j.chb.2016.09.026>, 2016.
- [17] M. L. R. Varela, M. F. Ferreira, G. G. Vieira , V. K. Manupati, K. Manoj, "A Multi-perspective Integrated Framework of Critical Success Factors for Supporting On-line Shopping", 2016.
- [18] H. Sergio, P. Alvarez, J. Fabra, J. Ezpeleta, "Analysis Of Users' Behaviour In Structured E-Commerce Websites", 2016.
- [19] P. Solainayagi, R. Ponnusamy, "To Improve Feature Extraction and Opinion Classification Issues in Customer Product Reviews Utilizing an Efficient Feature Extraction and Classification (EFEC) Algorithm", *International Journal of Electrical and Computer Engineering (IJECE)*, ISSN: 2088-8708. Vol. 10, No. 2, May 2018, pp. 587~595.
- [20] G. Silahtaroglu, H. Dönertasli, "Analysis and Prediction of E-Customers' Behavior by Mining Clickstream Data", *IEEE International Conference on Big Data (Big Data)*, Santa Clara, CA, 2015, pp. 1466-1472, 2015.
- [21] G. Chen, S. Ma, W. Wu, X. Wang, "Large-scale Online Sequential Behavior Analysis with Latent Graphical Model", *International Conference on Wireless Communications & Signal Processing (WCSP)*, Nanjing, 2015, pp. 1-6, 2015.
- [22] R. Purwaningsih, B. Adison, "Path Analysis Method to Identify Factors Affecting Consumer Interest on Online Shopping", *2nd International Conference on Science in Information Technology*, Balikpapan, pp. 20-25, 2016.
- [23] Y. Song, S. Yao, D. Yu, Y. Shen, Y. Hu, "A New K-Ary Crisp Decision Tree Induction with Continuous Valued Attributes", in *Chinese Journal of Electronics*, vol. 26, no. 5, pp. 999-1007, 9 2017.

- [24] Swanto, O. Wahyunggoro, A. I. Cahyadi, "Path Planning Based on Fuzzy Decision Trees and Potential Field", *International Journal of Electrical and Computer Engineering (IJECE)*, Vol. 6, No. 1, pp. 212-222, February 2016.
- [25] A. Khamis, Yan Xu, A. Mohamed, "Comparative Study in Determining Features Extraction for Islanding Detection using Data Mining Technique: Correlation and Coefficient Analysis", *International Journal of Electrical and Computer Engineering (IJECE)*, ISSN: 2088-8708. Vol. 7, No. 3, pp. 1112 – 1124, June 2017.
- [26] Y. Wang, C. Song, S. Xia, "Improving Decision Trees by Tsallis Entropy Information Metric Method", 2016.
- [27] Suzan Kantarci-Savaş, Efendi Nasibov, "Fuzzy ID3 Algorithm On Linguistic Dataset By Using WABL Defuzzification Method", *IEEE International Conference on Fuzzy Systems*, Naples, 2017, pp. 1-5, 2017.
- [28] ZAIM, Houda and Ramdani, Mohammed and Haddi, Adil, "A Model of E-Commerce Self-Assessment System Based on E-Customer Behavior," *Smart Application and Data Analysis for Smart Cities (SADASC'18)*. Available at SSRN: <https://ssrn.com/abstract=3179244>, May 15, 2018.
- [29] Neville, Kevin, "Channel Attribution Modelling Using Clickstream Data From An Online Store", 2017.
- [30] R. S. Algharabat, "The Role of Telepresence and User Engagement in Co-Creation Value and Purchase Intention: Online Retail Context", *Journal of Internet Commerce*, 17:1, 1-25, DOI: 10.1080/15332861.2017.14226, 2018.
- [31] R. C. Prati, F. Chartre, F. Herrera, "A First Approach towards A Fuzzy Decision Tree For Multi-Label Classification", *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Naples, 2017, pp. 1-6, 2017.
- [32] Zaim H, Ramdani M, Haddi A, "Multi-Criteria Analysis Approach Based on Consumer Satisfaction to Rank B2C E-Commerce Websites," *In: SITA 2016—11th international conference on intelligent systems: theories and applications*. <https://doi.org/10.1109/sita.2016.7772260>, 2016.