

Unleashing the Potential of Web mining: Tailored E-commerce Proposals and Efficient Customer Relationship Management

Swati Aggarwal

Abstract— As more and more businesses are now turning to Internet to take advantage of its inherent capability of providing global exposure and extensive market availability, the downside associated with this, customers now have all information available to them just with the click of button, they are now much more capacitated to do better price/quality comparisons and know discount schemes of rival businesses. So in such a kind of cut-throat competition it is very vital to hold the fidelity of the existing customers by providing them with personalized recommendations on the basis of their past purchases pattern and frequency, to constantly monitor their rival's strategy and their current market capture so as to regulate their selling strategies accordingly. As web is huge repository of ever growing, unstructured and dynamic information, thus providing the impetus of utilizing automated tools for locating, extracting, analyzing and using that potential information for better e-business decision making process for understanding the current customer behavior patterns and grabbing prospect customers. This research paper discusses potential of web mining which can be effectively coupled with the existing technologies that can help e-businesses to understand customer's preferences, predict their actions, fine tune their strategies to reduce market risk, increase their sell ability and also availing competitive advantage by taking right decisions at the right time.

Index Terms— Customer Relationship Management, E-commerce, Information Retrieval, Knowledge Discovery, Web Mining

I. INTRODUCTION

WITH more and more corporations opting for electronic business it is imperative for them to understand the heterogeneous data coming from varied sources like web server access logs, browser logs, proxy server logs, registration data, user queries, cookies, sessions so as to decipher and predict online user behavior. The underlying problem is that as Internet provides global platform to carry out the business, shoppers with different demographics, varying needs, preferences are bound to visit, the challenge is how to convert prospect client into an actual client, where is

it is normally observed that the users tend to actually lose their tangible goal after going through the myriad of multiple options of large and complex web structure. It is quite evident that the majority of the players of e-sector have done good business in past and chances are bright that they will continue doing well, these are the positives of e-commerce but we do have some businesses which went online but unfortunately had to close, here comes the question of what should be integral to any e-business and also what factors should be considered so as to ensure the transition of potential consumer to an actual loyal consumer.

One of the emerging and promising area in this regard is web mining that works as a catalyst for discovering and extracting information from web documents so as to equip the website executives with better decision making for better ROIs (Return On Investment) and generating customized e-commerce recommendations, thus enabling better CRM (customer relationship management)[1].

II. WEB MINING

Web mining is an inter-disciplinary and functional emerging field, the way data mining works to extract hidden and relevant information from huge, random, fuzzy and incomplete data, web mining also does the same but with the difference, that the source data it mines constitutes the web domain only. This section deals with what constitutes as the source data for web mining and various parameters governing web mining process.

A. Data sources in e-business

Data for web mining comes from varied sources. Following are the normally used as data sources [2]:

1. Customer registration information. This information can be found on server which is normally provided by the customers during their first visit and this helps better understanding of their demographics, such information when coupled with session logs helps increase web mining accuracy.
2. On-line market data. All types of transactional details like type/amount/time of purchases can be found in

S. Aggarwal is with Terna Engineering College-affiliated to Mumbai University, Nerul, Navi Mumbai, India (e-mail: swati11_2006@yahoo.co.in).

traditional relational databases, this type of information when combined with individual demographics gives better insight of preferences of customers with similar demographics.

3. Query data. This information can be traced from server log through cookies or register information, that will help website executives comprehend the kind of products/services costumers are looking for.
4. Web pages. Web pages include HTML or XML pages, which comprise texts, pictures, audio, and video and so on.
5. Hyperlinks between Web pages. The navigation path followed by the customers to reach to a particular page helps understanding better relation between different web pages.
6. Server data. This is the crucial source of information, wherein information regarding individual log data can be traced to document files, generally including server logs, error logs, cookies logs and so on.

B. Knowledge Discoveries

Web mining can be efficiently utilized by e-businesses to discover the hidden patterns and knowledge that helps better designing of web site, predict customer actions, time marketing strategies, offer personalized offers, some of these are discussed below [3].

1. Navigation analysis. Most frequently visited page or most frequently traversed path analysis give insight into which pages are most important from customer’s perspective and thus helps redesigning of the web site by better hyper linking or placement. This reduces the burden on the customer to locate a particular page of his interest lest getting lost somewhere in between.
2. Abnormality detection. Web mining facilitates a check on deviation from the normal behavior, if found one that can then be further linked with network intrusion, credit card fraud screening, unusual customer and so on.
3. Discovering patterns and doing Classification and prediction. Web mining can utilize the server data to know the purchasing patterns of customers from similar demographics, and can thus form potential clusters, which then help marketing people to realign their selling campaigns accordingly. Also if web mining reveals some sequential patterns like if the purchase of some item ‘B’ is triggered after the purchase of ‘A’, then also e-businesses can tune their web pages alignment and their discount schemes. Web mining facilitates classification which is mapping the data items in the database to one of the given types by

constructing a classification model or classifier so that it can be used in prediction.

Figure 1 describes the process of web mining in e-commerce environment.

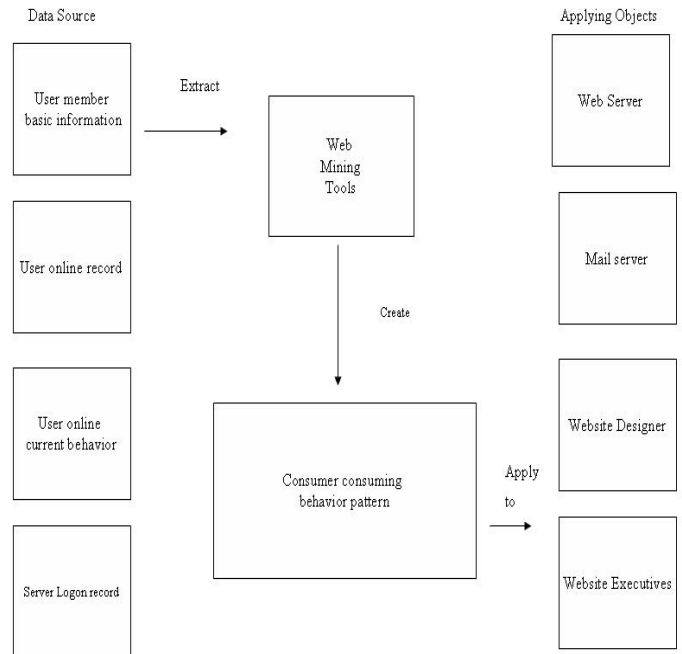


Fig 1. Process of web mining in e-commerce

III. PERSONALIZED RECOMMENDATIONS

Now more and more e-businesses are relying on personalized recommendations as this increases the chances of sell ability of product/service and also makes it convenient for the client to be glued to a specific e-shop for long, thus again increasing the likelihood of better sale and profits. Though many methods have been in use for customized recommendations like artificial intelligence, user modeling, this paper will discuss how to increase the effectiveness of recommendations significantly. What has been proposed here is that producing personalized recommendations is not the study along one dimension, many factors like analyzing and understanding behavior of customers, their preferences and history, demographics needs to be scrutinized and understood along with e-shop ontology. Advanced web mining holds the potential to carry out all the activities stated, thus increasing the convert ratio of users to buyers and retaining current customers by establishing long term bonding with them by offering one to one personalized proposals [4].

A. Methodology proposed

The framework suggested here generates customer navigational model off-line by combining usage and ontological data, thus distinguishing new users form old ones that helps sorting the recommendations by categorizing the users on the basis of the similarities drawn from navigational model that eventually helps in meeting the user preferences more efficiently.

Proposed methodology is shown in figure 2.

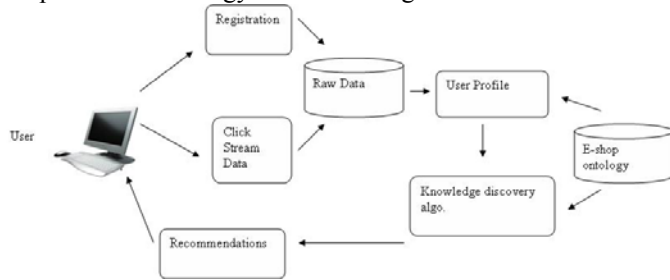


Fig. 2. Proposed recommendation approach

1) Step 1: Deriving e-shop ontological schema

In e-commerce, the website works as an interface between the customers and the product/services offered by e-businesses. So a web site can be viewed as a schema, wherein the homepage can be considered as a root, i.e. the most general class to which other pages are linked, those representing more specific classes. E-shop dedicated to selling of books online, portion of ontological schema is given for the same in figure 3.

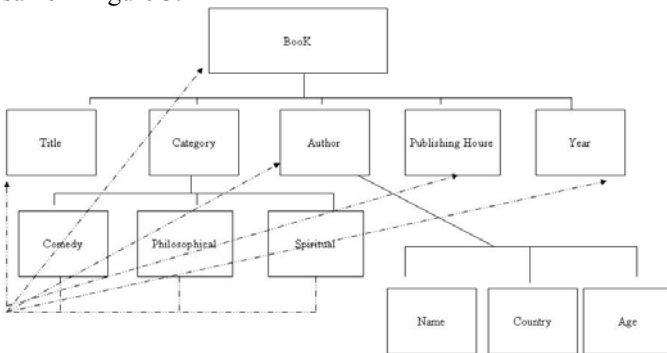


Fig. 3. Portion of ontological schema

This ontological schema is beneficial in classifying different people on the basis of different attributes that characterize them, that are helpful to find out customer preferences on the basis of customer specific history file [5]. Using specific history file of the customer his preferences can be chalked out, for instance if the current user is seen to be interested in philosophical book it is observed that he shows interests in spiritual books as well.

There are two cases to be dealt.

- 1) Old customer: In the case of the old customer, when one or more attributes match with the ontological schema that have similar attributes, other books of his interest can be generated.
- 2) New customer: For any new customer, history file would be null, so basic information can be taken from the registration form that the new user would fill initially and that then can be analyzed to generate recommendation on the basis of the ontology [6].

2) Step 2: Raw Data Extraction

As the e-businesses are interested in classifying people on the basis of their similarities in liking, so it is quite imperative to extract basic customer information regarding, user demographics, preferences, skills, usage details like ratings

given, action correlations, usage environment details like hardware, software, browser type and version, all such details can either be collected explicitly through some questionnaire or can be derived implicitly through acquisition algorithms or stereotype reasoning [7].

3) Step 3: Data Analysis

It is quite likely that the data collected might be incomplete or randomized, which might require some kind of preprocessing so as to facilitate it's manipulations like applying web mining techniques of clustering, associations, classifications for locating and extracting hidden information[8].

Old customer: In case of the old customer, the proposed recommendation approach would be utilizing web usage mining techniques and semantic metadata coupled with customer ratings and matching techniques.

New customer: For every new customer it is required to fill in registration form, so as to gather initial basic information to support him in his navigation.

Classification Process:

There would be users with similar interests, so if they can be classified together, it would be easier for the e-shops to generate recommendations for the similar liking people on the basis of the history of the people in the same class [9].

It is suggested classification be performed on the basis of naïve Bayes algorithm which states that each customer will be allocated to one of the predefined classes.

According to Bayes classifier any non-classified snapshot X with set of attributes (x₁, x₂,..... x_n) would be assigned to in the class C_i with the maximum probability. Table I describes the sample instance portion of the new current user profile.

TABLE I
SAMPLE INSTANCE PORTION OF THE NEW CURRENT USER PROFILE

Login key	A088
age	34
country	India
state	Mumbai
education	High school
occupation	Business

For the given unknown snapshot, the attribute values will be matched with the attributes of the varied predefined classes, and will be finally be associated with a class with maximum probability. The classification described here calls for a supervised machine learning where training data with pre assigned classes will be used to train the classifier to further gradually learn to assign new data in one of the classes and such classification would be evolving with time subjected to deletions and entry of new classes in the classification model [10].

Generation of association rules:

Web server would have all the details pertaining to the log files, so all the click stream data would be easily available from it. Main purpose of applying association rules is to derive correlations between heterogeneous information which have no obvious dependence[11]. The set of pages visited and the books purchased details can be traced from the user profile stored on to the web server that helps extracting interesting association rules like

$$\{U_{\text{spiritual books}}\} \rightarrow \{U_{\text{philosophical books}}\}$$

Support=0.01 and confidence=0.74

suggests that customers who purchased spiritual books tends to show interest in philosophical books as well with a confidence of 74% and 1% of the support data is available for the same.

4) Step 4: Customized Recommendations

The discovered association rules and classes would be used in conjunction with each other to generate an automated recommendation for end users [12].

Following cases are to be considered for discussion:

Case 1: User is old

If the user is old, then the server is having his past details and a specific user profile must be present for that user.

Let current user be represented by U_c and the book he is interested is, lets say B.

- a. For current user U_c , extract $U_{cbhistory}$, books that user has purchased in past.
- b. Using this $U_{cbhistory}$, system would generate set of users, $U_{similar}$, consisting of users who have purchased similar books as of the current user U_c .
- c. Every user U_i that is an element of $U_{similar}$ is associated with number S_i , depicting similarity of U_i with current user U_c .
- d. Extract user U_{same} from $U_{similar}$, who have read book B (the current user U_c is asking for) and having similarities close to the current user U_c (threshold values need to be specified by system administrator).
- e. For all users U_i , that are part of U_{same} , extract their $U_{ihistory}$ and then the model proposed would use ontology to discover associations between items that the user has rated.
- f. Model would generate $R_{ontology}$, the list of recommendations that can be concluded on the basis of user's profile having details of his past transactions.
- g. If the current user U_c , in the current session has accessed set of books B1, then on the basis of the current session, $R_{current-recommendations}$ can be generated using recent accesses and association rules mining.
- h. Similarly on the basis of the present session $R_{current-ontology}$ can be derived.
- i. Finally the recommendations generated by the system would be

$$j. R_{\text{final-recommendation}} = R_{\text{current-recommendations}} \cup R_{\text{current-ontology}} \cup R_{\text{ontology}}$$

Case 2: User is new

- a. For current user U_c , extract information from his registration form and classify user to class C_i using Bayes classifier. (C_i would comprise of the users $U_{similar}$ having similar attribute values as specified by eth current user in his registration form)
- b. Every user U_i that is an element of $U_{similar}$ is associated with number S_i , depicting similarity of U_i with current user U_c .
- c. Extract user U_{same} from $U_{similar}$, who have read book B (the current user U_c is asking for) and having similarities close to the current user U_c (threshold values need to be specified by system administrator).
- d. Using clustering algorithm, every class is associated with set of recommended books $R_{cluster}$.
- e. If the current user U_c , in the current session has accessed set of books B1, then on the basis of the current session, $R_{current-recommendations}$ can be generated using recent accesses and association rules mining.
- f. Similarly on the basis of the present session $R_{current-ontology}$ can be derived.
- g. Finally the recommendations generated by the system would be

$$R_{\text{final-recommendation}} = R_{\text{current-recommendations}} \cup R_{\text{current-ontology}} \cup R_{\text{cluster}}$$

IV. CONCLUSION

With Internet flooded with numerous websites dedicated solely for e-commerce, to sustain in such an environment is a challenge. As we have many examples of top-notch web sites doing extremely well in e-business, the dark side associated does reveal that many could not survive. The bottom line is that customers are the real asset of any business, but with e-business, it is a daunting task to convert casual one time visit to a sellable visit and to ensure multiple future visits as well because to catch hold of the customers in today's scenario is exigent as just with a click of button, they tend to deviate towards better options, so it is imperative to have better customer relationship management. Better CRM is possible only if e-business stores the individual profile along with past transactions carried out, churn the historical data, frequently making analysis, generating personalized offers and thus ensuring better profitability. With current techniques and superior methods available, web mining holds the potential of making efficient use of information at hand to calculate and devise tactics about prospect e-business track lest headless stratagem would lead to a dead-end.

V. REFERENCES

- [1] M. Mehta, R. Agrawal, and J. Rissanen, "SLIQ: A fast scalable classifier for data mining", Proc. of the Fifth Int'l Conference on Extending Database Technology, Avignon, France, 1996.
- [2] Zhong Xie, "Recommendation System of Commercial Site Research Based on Web Data Mining", *Master's Degree Thesis of Southwest Normal University*, 2002(05), pp. 8-10.
- [3] P. R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," *Proceedings of the 20th International Conference on Very Large Databases*, Santiago, Chile, September 1994, pp. 487-499.
- [4] P. Markellou, P., Rigou, M., and Sirmakessis, S., "Mining for Web Personalization", *Web Mining: Applications and Techniques*, A. Scime (Ed.), Idea Group, 2005, pp. 27-48.
- [5] Ding, Y., Fensel, D., Klein, M., Omelayenko, B., and Schulten, E., "The Role of Ontologies in eCommerce", *Handbook on Ontologies*, S. Staab and R. Studer (Eds.), Springer-Verlag, 2004, pp. 593-615.
- [6] Choa, Y.H., and Kimb, J.K., "Application of Web Usage Mining and Product Taxonomy to Collaborative Recommendations in E-commerce", *Expert Systems with Applications*, 2004, 26, pp. 233-246.
- [7] Chakrabarti, S., "Mining Web: Discovering Knowledge from Hypertext Data", Morgan Kaufmann, 2003.
- [8] Eirinaki, M., Vazirgiannis, M., and Varlamis, I., "SEWeP: Using Site Semantics and a Taxonomy to Enhance the Web Personalization Process", *Proc. Of 9th ACM Conference on Knowledge Discovery and Data Mining*, Washington, USA, 2003, pp. 99-108.
- [9] Eirinaki, M., and Vazirgiannis, M., "Web Mining for Web Personalization", *ACM Transactions on Internet Technology*, 2003, 3(1), pp. 1-27
- [10] Zaiane OR Xin M, Han J. Discovering web access pattern and trends by applying OLAP and data mining technology on web logs. In: Proc of Advances in Digital Libraries Conf (Adl' 98). Santa Barbara, CA, 1998, 19~29
- [11] Eirinaki, M., Lamos, C., Paulakis, S., and Vazirgiannis, M., "Web Personalization Integrating Content Semantics and Navigational Patterns", *Proc. Of 6th ACM International Workshop on Web Information and Data Management*, Washington, USA, 12-13 November 2004, pp. 72-79.
- [12] Markellou, P., Rigou, M., and Sirmakessis, S., "Mining for Web Personalization", *Web Mining: Applications and Techniques*, A. Scime (Ed.), Idea Group, 2005, pp. 27-48.
- [13] P. Kosala, R., and Blockeel, H., "Web Mining Research: Survey", *SIGKDD Explorations*, 2000, 2(1), pp. 1-15.

VI. BIOGRAPHY



Ms. Swati Aggarwal has done B.Tech (Computer Science), university rank holder and M.Tech (I.T). She has 7 years of teaching experience. During her tenure of teaching graduate and post graduate engineering students, she has taught subjects like Artificial Intelligence, Data Warehouse and Mining, Project Management, Software Engineering. Her areas of interests are Artificial Intelligence, Fuzzy Logic, Data Warehousing and Mining. She has published research papers on topics related to neural networks, data mining and web mining. Continuing her passion of learning and teaching, she is currently employed with an engineering college and has plans to pursue her PhD shortly.