

DEVELOPMENT OF SECURED & EFFECTIVE E-COMMERCE SYSTEM WITH USER BEHAVIOURAL MINING

D. Sridhar, MCA., M.Phil., Assistant Professor, Department of Computer Science,
Dr. SNS Rajalakshmi College of Arts & Science, Coimbatore .
D.Nandhini, B.Sc (CS)., M. Sc (CS)., Department of Computer Science,
Dr. SNS Rajalakshmi College of Arts & Science, Coimbatore.

Abstract

Online shopping is becoming more and more common in our daily lives. Understanding users' interests and behavior is essential to adapt e-commerce websites to customers' requirements. The information about users' behavior is stored in the Web server logs. The analysis of such information has focused on applying data mining techniques, where a rather static characterization is used to model users' behavior, and the sequence of the actions performed by them is not usually considered. Therefore, incorporating a view of the process followed by users during a session can be of great interest to identify more complex behavioral patterns. To address this issue, this project proposes a linear-temporal logic model checking approach for the analysis of structured e-commerce Web logs. By defining a common way of mapping log records according to the e-commerce structure, Web logs can be easily converted into event logs where the behavior of users is captured. Then, different predefined queries can be performed to identify different behavioral patterns that consider the different actions performed by a user during a session. Finally, the usefulness of the proposed approach has been studied by applying it to a real case study of a Spanish e-commerce website.

Keywords— E-commerce, Web logs, User identification , Session identification,

1.0 INTRODUCTION

In the present at any point associated world, the manner in which individuals shop has changed. Individuals are purchasing increasingly finished the Internet as opposed to going conventional shopping. Web based business furnishes clients with the chance of perusing perpetual item lists, contrasting costs, being ceaselessly educated, making list of things to get and getting a charge out of a superior administration in view of their individual advantages. This expanding electronic market is very focused, highlighting the likelihood for a client to effortlessly move from one web based business when their necessities are not satis_ed.. As an outcome, online business investigators require to knowand comprehend customers' conduct when those explore through the site, and in addition attempting o distinguish the reasons that inspired them to buy, or not, an item. Getting this social information will permit internet business sites to convey a more

customized administration to clients, holding clients and expanding. However, finding client' conduct and the reasons that guide their purchasing procedure is an exceptionally complex undertaking. Web based business sites give clients a wide assortment advantages of navigational choices and activities: clients can openly travel through various item classes, take after numerous navigational ways to visit a particular item, or utilize distinctive systems to purchase items, for instance. For the most part, these client exercises are recorded in the web server logs.

Web server logs store, in an arranged way, the succession of web occasions produced by every client (regularly known as snap streams). The simple important clients' conduct is covered up in these logs, which must be found and examined. A right investigation can be along these lines used to enhance the site substance and structure, to adjust and customize substance, to prescribe items, or to comprehend the enthusiasm of clients in particular items, for example. Information mining systems have demonstrated their helpfulness for finding designs in log _les (when connected to the investigation of web server logs the term web use mining is utilized). Its primary objective is to find use designs endeavoring to clarify the clients' advantages. Diverse methods have been effectively utilized in the field of web based business, for example, arrangement procedures, bunching, affiliation rules or successive examples . In numerous application spaces these methods are utilized in conjunction with process mining strategies. Such strategies are a piece of the business knowledge area and apply specific calculations to find concealed examples and connections in vast informational collections.

2.0 RELATED WORK

Right now, there are great business apparatuses for investigating logs of online business sites, being Google Analytics one of the primary ones. Google Analytics controls the system movement, gathers data about client. Google Analytics can't import the web server logs of a site, yet it works examining the data gathered by methods for page labeling systems. These methods have a few burdens regarding the log-based investigation, for example, reliance on JavaScript and treats, the need of adding page labels to each page, the multifaceted nature of label based executions, and the way that, therefore, clients may encounter an adjustment in the download time of the site, or protection worries, for example. All things considered, Google reports are wealthy in information that, thus, require specialists in the issue area to misuse them.

LIMITATIONS

- Tracking client profile is extremely troublesome in powerful page
- The client's perusing is put away in the nearby machines.
- The client's normal looking requires significant investment in different machines.
- Less security against DOS assaults.

3.0 OUR SYSTEM MODEL

Proposed plot finds the standard of conduct of the client by breaking down client action in single session from web logs of online business site. Diverse predefined are utilized to recognize the standard of conduct. In contrast with the grouping approaches and the remarked business instruments, the benefit of our mining method is this gives causal relations among occasions of a client follow, rather than furnishing with a worldwide perspective of the entire session. Concerning those methodologies whose primary goal is anticipating the coming conceivable occasions, the approach permits having a worldwide perspective of the sessions, making less demanding a worldwide examination of the client conduct, giving insights and encouraging the re-outline of the site for a superior adjustment to the client necessities. The reality of thinking about the causal relations of occasions inside a client session, permitting searching for intra-session designs (and not just examples rehased in various sessions) can furnish the investigators with a considerably more point by point viewpoint of a client conduct.

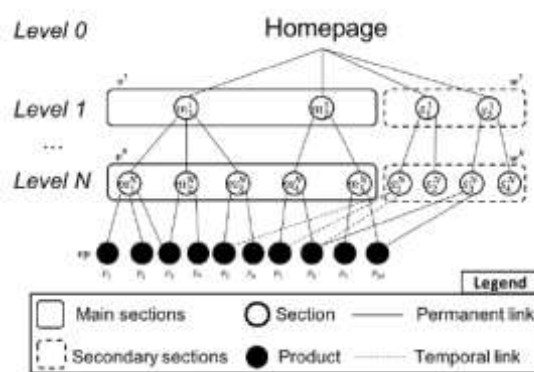


Fig1: structure for categories products in ecommerce websites

Here we execute constrained login confirmation with the goal that a programmer can figure secret word just couple of restricted occasions. On the off chance that the restricted check surpassed, the login procedure will be diverted to a phony page in which an insights about programmer data will be assembled like framework IP, framework name and so forth. At that point that record will be naturally obstructed by our framework.

Benefits

- Tracking the development of the found profiles.
- These diverse methods of utilization or the alleged mass client profiles can be found utilizing Web use procedures that can consequently remove visit get to designs from the historical backdrop of past client click streams put away in Web log records.

3.1 IMPLEMENTATION OF INCREMENTAL LEARNING

This module manages the client relationship administration (CRM). Client can login that site. In the event that login effectively, at that point approve client can enter the site. They just can see the connection points of interest. Clients require distinctive composes and alluring sorts of connections. So the connection ought to be composed in client's perspective. In this module, following client profile and improving web fame. It must be appealing and easy to use. This venture can even produce data like site page points of interest, client subtle elements, in site and outbound connections ie, client's are originating from locales. So effortlessly ascertain web prevalence. So the site proprietors can build up connection between different locales and increment their site prominence. This venture proposes a straight transient rationale demonstrate checking approach for the investigation of organized online business Web logs. By characterizing a typical method for mapping log records as indicated by the web based business structure, Web logs can be effectively changed over into occasion logs where the conduct of clients is caught. At that point, diverse predefined questions can be performed to recognize distinctive standards of conduct that consider the diverse activities performed by a client amid a session. The outcomes have recognized fascinating discoveries that have influenced conceivable to propose a few enhancements in the site to outline with the point of expanding its effectiveness.

3.2 CLASSIFICATION

In this module outlines are characterized in view of the client getting to. Toward the finish of the client session, item which are shown in this site are arranged in light of the clients click. Top scored item can acquire more finances. Furthermore, this can supportive for purchasers to purchase item.

3.3 PREDICTION

This expectation gives recommendation to purchaser to buy item. Item with high getting to are prescribed to purchase. In light of the getting to purchaser can buy and furthermore thing of that item can get rewards. Lastly this item can be taken for forecast result additionally administrator can see diagram result in view of client purchasing.

4.0 ALGORITHM

We propose a robust classifier to predict buying intentions based on user behaviour within a large e-commerce website. In this work we compare traditional machine learning techniques with the most advanced deep learning approaches. We show that both Deep Belief Networks and Stacked Denoising auto-Encoders achieved a substantial improvement by extracting features from high dimensional data during the pre-train phase. They prove also to be more convenient to deal with severe class imbalance.

4.1 User identification

Each extraordinary client getting to the site is recognized in the client recognizable proof process. The point of this procedure is to recover each client's entrance attributes, at that point make client grouping and give proposal administration to the clients. Distinctive clients are distinguished by various ip addresses.

Input: refine_log_table

Output: identification of user

Begin

1. Read records in log_table
2. for each record in dataset do
3. If current IP is not in ListOfIP then add the current IP in ListOfIP mark whole record as a new user and assign

userID

4. else assign the old userID.

End else

End if

end

4.2 Session Identification

A succession of pages seen by a client amid one visit is known as the Session. The session is recorded in the log document. In pre-preparing it is important to discover session of every client. It characterizes the occasions the client has gotten to a website page. It takes all the page reference of a given client in a log and partitions them into client sessions. These sessions can be utilized as an information vector in grouping, bunching, expectation and different undertakings. In light of a uniform settled timeout a conventional session recognizable proof calculation is utilized. Another session is recognized when the interim between two successive demands surpasses the 60 minutes.

Input: user identified table

Output: identified sessions

Begin

1. Read records in log_table
2. for each record in dataset do
3. if time_required > one hour assign new sessionID for that log entry

4. increment sessionID

5. else

assign the old sessionID.

End else

End if

End

5.0 DISCUSSIONS

On account of open frameworks, where the arrangements of connections (put away as framework logs) are not obliged by a work process, process mining procedures whose goal is to extricate a procedure model will for the most part give either finished fitting spaghetti models or under fitting bloom models, from which small fascinating data can be separated. A more adaptable approach is required. In the paper we apply LTL-based model checking systems to examine internet business web logs. To empower this investigation, we have proposed a typical method for speaking to occasion composes and properties considering the internet business web structure, the item classification and the conceivable outcomes of clients to explore the site as indicated by such association. From this auxiliary perspective, the paper proposes an arrangement of question designs, converted into LTL recipes, which are of enthusiasm for the area of electronic trade. The responses to the inquiries, as far as the number (or rate) of follows fulfilling the relating equation, permits to extricate fascinating connections among groupings of occasions, which can be deciphered as far as clients' conduct.

6.0 FUTURE WORK

As a not so distant future work we need to furnish the investigation device with a graphical interface, for both the contribution of properties to be dissected and the yield of results, with the point of encouraging its utilization for non-specialized staff, giving a reflection level concealing the LTL formalism. We likewise plan to broaden the arrangement of considered examples with a specific end goal to investigate more personal conduct standards and to encourage their programmed revelation.

7.0 REFERENCES

- [1] J. B. Schafer, J. A. Konstan, and J. Riedl, "E-commerce recommendation applications," *Data Mining Knowl. Discovery*, vol. 5, no. 1, pp. 115_153, Jan. 2001.
- [2] N. Poggi, D. Carrera, R. Gavalda, J. Torres, and E. Ayguadé, "Characterization of workload and resource consumption for an online travel and booking site," in *Proc. IEEE Int. Symp. Workload Characterization (IISWC)*, Apr. 2010, pp. 1_10.
- [3] R. Kohavi, "Mining e-commerce data: The good, the bad, and the ugly," in *Proc. 7th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2001, pp. 8_13.
- [4] W.W. Moe and P. S. Fader, "Dynamic conversion behavior at e-commerce sites," *Manage. Sci.*, vol. 50, no. 3, pp. 326_335, 2004.
- [5] G. Liu *et al.*, "Repeat buyer prediction for e-commerce," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, New York, NY, USA, 2016, pp. 155_164.
- [6] J. D. Xu, "Retaining customers by utilizing technology-facilitated chat: Mitigating website anxiety and task complexity," *Inf. Manage.*, vol. 53, no. 5, pp. 554_569, 2016.
- [7] Y. S. Kim and B.-J. Yum, "Recommender system based on click stream data using association rule mining," *Expert Syst. Appl.*, vol. 38, no. 10, pp. 13320_13327, 2011.

- [8] R. Kosala and H. Blockeel, "Web mining research: A survey," *SIGKDD Explor. Newsl.*, vol. 2, no. 1, pp. 1_15, Jun. 2000.
- [9] F. M. Facca and P. L. Lanzi, "Mining interesting knowledge from Weblogs: A survey," *Data Knowl. Eng.*, vol. 53, no. 3, pp. 225_241, 2005. [10] C. J. Carmona and S. Ramírez-Gallego, F. Torres, E. Bernal, M. J. del Jesus, and S. García, "Web usage mining to improve the design of an e-commerce website: OrOliveSur.com," *Expert Syst. Appl.*, vol. 39, no. 12, pp. 11243_11249, 2012.
- [11] Q. Song and M. Shepperd, "Mining Web browsing patterns for e-commerce," *Comput. Ind.*, vol. 57, no. 7, pp. 622_630, 2006. [12] O. Arbelaitz, I. Gurrutxaga, A. Lojo, J. Muguerza, J. M. Pérez, and I. Perona, "Web usage and content mining to extract knowledge for modelling the users of the bidasoatourismo website and to adapt it," *Expert Syst. Appl.*, vol. 40, no. 18, pp. 7478_7491, 2013.
- [13] J. K. Gerrickagoitia, I. Castander, and F. Rebón, and A. Alzua-Sorzabal, "New trends of intelligent e-marketing based on Web mining for e-shops," *Procedia-Social Behavioral Sci.*, vol. 175, pp. 75_83, Sep. 2015.
- [14] Y. H. Cho and J. K. Kim, "Application of Web usage mining and product taxonomy to collaborative recommendations in e-commerce," *Expert Syst. Appl.*, vol. 26, no. 2, pp. 233_246, 2004.
- [15] K.-J. Kim and H. Ahn, "A recommender system using GA K-means clustering in an online shopping market," *Expert Syst. Appl.*, vol. 34, no. 2, pp. 1200_1209, 2008.
- [16] Q. Su and L. Chen, "A method for discovering clusters of e-commerce interest patterns using click-stream data," *Electron. Commerce Res. Appl.*, vol. 14, no. 1, pp. 1_13, 2015.
- [17] J. Srivastava, R. Cooley, M. Deshpande, and P.-N. Tan, "Web usage mining: Discovery and applications of usage patterns from Web data," *SIGKDD Explor. Newsl.*, vol. 1, no. 2, pp. 12_23, Jan. 2000.
- [18] Q. Zhang and R. S. Segall, "Web 2mining: A survey of current research, techniques, and software," *Int. J. Inf. Technol. Decision Making*, vol. 7, no. 4, pp. 683_720, 2008.
- [19] D. Sridhar and Dr.V.Kathiresan "The Psychology behind users mental health status in social media" , International journal of Computer Sciences and Engineering, vol. 6, Issue 5, 2018
- [20]. D. Sridhar and Dr.V.Kathiresan, "Predicting mental behaviour of Social Media users using AOMM(Auto Observation Mining Method)", Journal of Advanced Research in Dynamical and Control Systems, vol. 10, Issue 01. 2018
- [21] K.Vengatesan, S.Sadhana, Published a paper on "Performance Enhancement of Robust Rough Fuzzy Clustering using Silhouette Index" International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 5, May 2016, pp.9885-9890.
- [22] K.Vengatesan, Dr.S.Selvarajan, Published a paper on " The performance Analysis of Microarray Data using Occurrence Clustering" International Journal of Mathematical Science and Engineering, Issue-2, Volume-3, December 2014, pp 69-75.