A Novel System based on GMM to understand the User's Web Navigation Pattern

M.Bangaru Lakshmi[1], A.Nageswara Rao[2,] G.Venugopal[3]

[1,2] Department of CSE,
Pragati Engineering College, Kakinada, (A.P.), India
[3] Department of CSE,
Narayana Engineering College, Nellore, (A.P.), India

Abstract -The latest development in the area of communication technology paved way towards the usage of E-commerce technologies. The B2B transaction has helped the customers to get the best resource by comparing various resources of same kind, leaving the business community in guessing about the user choice and the reasons beyond the user's attraction about a product. To understand this changing patterns, intelligent techniques are needed which can help to understand the user's navigation pattern. This paper proposes a model, using Gaussian Mixture Model to understand the web navigation pattern of the user, and help the business community to understand user guess and supply the customers according to their choice.

Keywords: E-commerce, Web navigation, Gaussian Mixture Model.

1. INTRODUCTION

Web personalization is a process of gathering the information about the visitors profile [1][2][3][4],the navigation path taken by the user. The information about the same are stored for further analysis using various techniques specified by the user. Web personalization plays a dominating role in the every organization and helps to derive useful results that help in the increase of number of visitors for a particular web page / buying a specific product and retaining the visitors. Web personalization uses the technologies, such as modeling, social networks, web mining etc…to facilitate the user to navigate the best path for effective navigation path analysis. The visitor's information is gathered and the analysis is done using the cache content. The visitors information gathered will facilitate to design a user profile containing visitors choices, purchases made during the visit. The user's profiles can be deemed to be either static or dynamic, and the profiling can be implicit where the details of the visitors visited sites are considered and in explicit profiling, help to understand the user's behavior. The   visitors behavioral pattern and purchasing ability can be assessed from the browsing pattern and purchase database .To analyze the visitors patterns several methods including rule based [5] and filtering technique [6] are discussed in literature, and most of the research is based on identifying the behavior pattern of the user, however the main disadvantage, is that if number of user exhibits the same common interest, then analyzing the navigation patterns of the user from the overlapping interest

is a challenging problem. In order to overcome this disadvantage, we propose a model based technique, based on GMM, which analyzes the pattern of the user navigation and using the maximum likelihood estimate of the model parameters μ, σ,we identify the users navigation pattern. To explain the model, a web advertisement scenario is considered, where the interest lies in the identification of user behavior based on the advertisement, and thereby identifying the actual time of the display of advertisement, time duration and correlating the advertisement with the users buying sequences and testing the impact. The rest of the paper is organized as follows; Review of the work in this area is presented in Section-2, Section-3 describes the GMM model, Section-4 of the paper deals with the experimental results and the final Section-5 concludes the paper.

2. LATEST REVIEW OF THE WORK IN THIS AREA

2.1 K.S.Kuppusamy and G.Aghila et al (2012) has proposed a contemporary content filtering technique based system, where the bulk data is given ranking according to the number of times it is viewed by the users. If a person searches for a subject (word) then we get the results according to the ranking, the most viewed document is given first rank (or) first in a screen and so on.

2.2 Tasawar Hussain, Dr.Sohail Asghar, Dr.Nayyer Masood(2010).In their paper expressed that the web page contains huge amount of data, it may be difficult for the user to understand because of huge and irrelevant information are represented in a single web page. It makes confusion and also make hurdle to search correct data (or) information but all these information is available online, which is hidden for the user.

2.3 Tasaware proposed about the web usage mining technique in which the web users are categorized based on their click history and similarity in search. So this technique converts the information of data preprocessing to a web log data and this web log data is converted to numerical data based on which the similar web data are clubbed in to one cluster. So with the help of this technique, we can obtain structured information.

The recent reviews have been mostly projected to understand the web user behavior pattern based on number of hits and click history. But to have more realistic investigation, generative models like Gaussian Mixture Models will be well suited, since mixture models understand the realistic behavioral patterns better. Hence in this paper, GMM model is utilized to understand the web navigation pattern.

3. GAUSSIAN MIXTURE MODEL

Mixture models can be used for many different purposes, for instance, density estimation, clustering, and classification. Here, we focus on clustering. You will see why it's useful for clustering as we proceed.

The PDF of the Gaussian Mixture Model is given by f(x) =

$$\frac{1}{\sqrt{2\,\Pi\,\sigma}} \; e^{-\left(\frac{x-\mu}{\sigma}\right)^2}$$

where x denotes the number of customers, μ=mean and σ = standard deviation.

The user navigation pattern is modelled using the GMM. A list of users navigated a particular website basing on the advertisement and without advertisement are obtained. The user's navigation pattern is identified by calculating the PDF in each case,

Products statistical based on With Advertisement and Without Advertisement:

TABLE 1

With Advertisement		Without Advertisement	
No. of users	No. of web pages	No. of user's	No. of web pages
10	13	8	9

In order identify the impact of advertisement, we check for the most likelihood estimate. In order to identify, we find the PDF of the test user and map these with the PDF of existing values, if the likelihood matches, we say the impact of advertisement.

4. EXPERIMENTATION

In order to experiment the data, we have obtained the log file, cleaning of the data is done. The data set used consists of the following fields

The data is cleansed, by removing preprocessing techniques and the processed data is utilized for experimentation. The experimentation is carried out on PIV system, under dot net environment. The data is first modelled using GMM and the PDF are obtained. In order to identify the impact of the advertisement, the MLE match is obtained and the impact is calculated. Based on the maximum likelihood estimates, the advertisement with MLE has more number of hits.
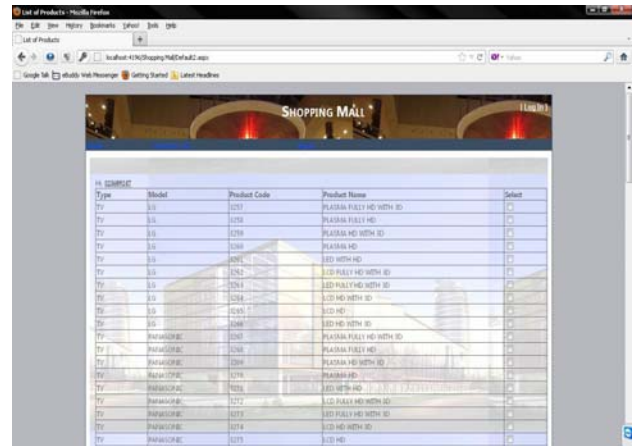
Results:

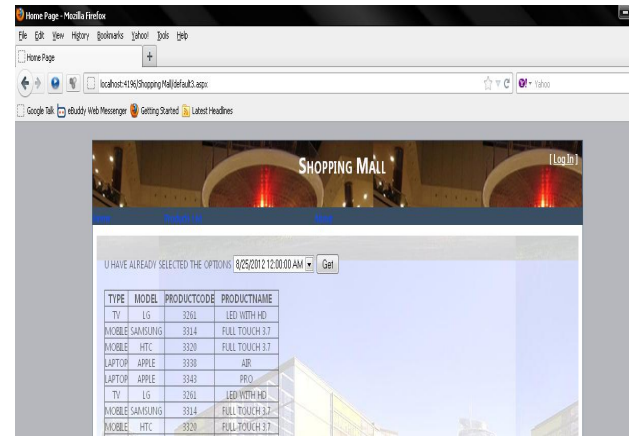Fig1. List of Products for Selection
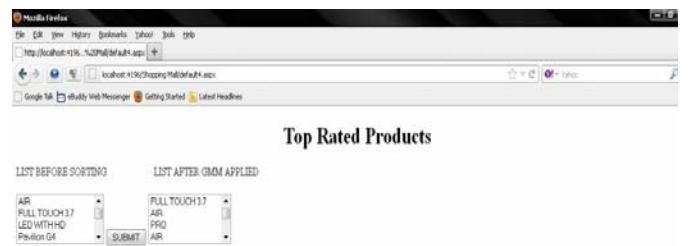

Fig2. Selections of Products by Registered Users

Fig3. Top Rated Products (GMM)

5. CONCLUSION

In this paper, a novel approach for identifying the user's navigation pattern is analyzed using GMM.The developed model helps to analyze the mostly viewed sites by the user. This model is best suited to understand the navigation pattern as well as the user's interest.

REFERENCES

[1] Tasawar Hussain, Dr. Sohail Asghar, Dr. Nayyer Masood, "Web Usage Mining: A Survey on Preprocessing of Web Log File", 2010.

[2] Dimitrios Pierrakos, Georgios Paliouras, "Personalizing Web Directories with the Aid of Web Usage Data", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 22, NO. 9, Sep 2010.

[3] D. Pierrakos, G. Paliouras, C. Papatheodorou, and C.D.Spyropoulos, "Web Usage Mining as a Tool for Personalization: A Survey," User Modeling and User-Adapted Interaction, vol. 13,no. 4, pp. 311- 372, 2003.

[4] J. Srivastava, R. Cooley, M. Deshpande, and P.T. Tan, "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data," SIGKDD Explorations, vol. 1, no. 2, pp. 12-23, 2000.

[5] Tasawar Hussain, Dr. Sohail Asghar, Dr. Nayyer Masood, "Web Usage Mining: A Survey on Preprocessing of Web Log File", 2010.

[6] G. Xu, Y. Zhang, and Y. Xun, "Modeling User Behaviour for Web Recommendation Using lda Model," Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence and Intelligent Agent

[7] B. Mehta and N. Wolfgang, "Unsupervised Strategies for Shilling Detection and Robust Collaborative Filtering," User Modeling and User-Adapted Interaction, vol. 19, nos. 1/2, pp. 65-97, 2009.

[8] Resnick et al. 1994; Shardanand & Maes 1995; Breeze et al. 1998 Collaborative Filtering and Social Filtering.

[9] B. Sarwar, G. Karypis, J. Konstan, and J. Reidl. Item-Based Collaborative Filtering Recommendation Algorithms.

[10] Springer Berlin/Heidelberg Web Personalization Techniques for Ecommerce –

[11] Xin Jin, Yanzan Zhou, Bamshad Mobasher, 2005."A Maximum Entropy Web Recommendation System: Combining Collaborative And Content Features," Acm Kdd.

Ms.M Bangaru Lakshmi received the B.Tech Degree from JNTU, Kakinada in 2009 and She is currently pursuing M.Tech in the Department Of Computer Science and Engineering in Pragati Engineering College, Of JNTUK Affiliation, Her research interests include Data Mining and Computer Networks.

Mr.A Nageswara Rao received the B.E degree from CBIT affiliated to OU in 1998, Hyderabad, India and the M. Tech degree in Computer Science Technology from Central University in 2000, and he is currently pursuing the Ph. D in Data Mining from Rayalaseema University A.P. He is working as a Head of the Department for CSE in Pragati Engineering College, Kakinada,A.P,India.

Mr.G.Venu Gopal received the AMIE degree from Institute of Engineering India in 2002, Kolkata, India and the M. Tech degree in Computer Science Technology from Allahabad university in 2006, he is currently pursuing the Ph. D in Data Mining from Rayalaseema University A.P. He is working as a Associate Proffesor in the Department of CSE in Narayana Engineering college, Gudur, Nellore,A.P,India.