ijRASET

INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

www.ijraset.com

Call: ◎08813907089 | E-mail ID: ijraset@gmail.com

Click Stream analysis by using Markov Cluster Algorithm

D. Anusha[1], K. Sri Harshavardhan ReddZ[2], B. Rajesh[3], V. Nagaraju[4]
[1, 2, 3, 4,] Department of IT, LBRCE

Abstract: As technology is evolving continuously the process of online shopping became more easy and flexible. This evolution leads to increase in customer needs in day to daily life. To maximize the business profits and to gain more number of customers, there is a need to analyze the user interactions effectively. But the user actions are not limited to workflow constraint, which adds new problems like complexity in collecting user actions and difficulty in mining these data. An approach is introduced to overcome this problems i.e., session characterization, in which sequence of user interactions are being analyzed by using data mining techniques. Although conventional methods of data mining can get user interests, but it fails to reach optimum accuracy. These results in the introduction of our proposed system called Markov Clustering Algorithm, an implementation of Markov chains on clusters. This has the capacity to mine the user actions effectively.
Keywords: Session characterization, Markov Clustering, clusters

I.   INTRODUCTION

Online shopping became easier with the increase in technology and it also creates more competition between Commercial sites. In order to lead in this competition e-commerce should provide a better customer service, more number of products, and variety of products. To implement this in practice e-commerce should analyse the user's behaviour more accurately. This creates an opportunity for data analysts to mine and analyse the data which was created by the actions of the user in e-commerce. The sequence of actions performed by the user or customer in an e-commerce is considered to be as click streams, flow of actions. The data analysts make good use of this click streams to analyse the user behaviour.

In existing system many conventional methods like model checking, which analyse the data by using some mathematical formulas were used. This methodology is difficult to mine knowledge from the data. This existing system is not focussed on the product recommendation problem, which aids to enhance the structure of e-commerce.

In proposed system, the data is analysed by using Markov clustering algorithm form the click stream data. This technique aids the data analysts to analyse each user behaviour individually. This also solves the problem of product recommendation problem. Hence the e-commerce site can be enhanced more effectively according to the user interests.
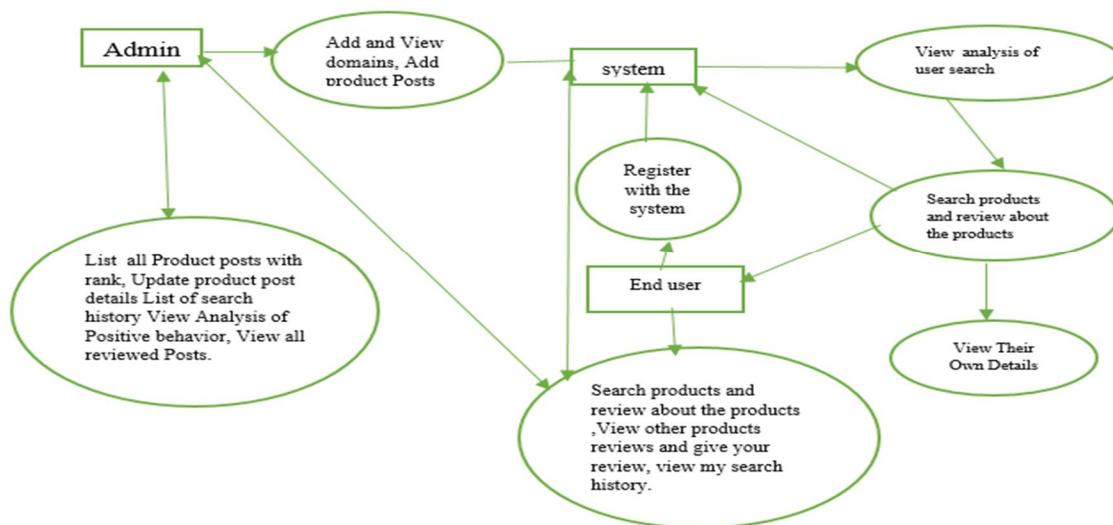

Fig 1. Architecture

II. RELATED WORK

In this paper[1], concept of Significant Usage Patterns(SUP) is proposed to mine specific "user preferred navigational patterns" .This approach is used to extract the valuable user behavioral information from the sessionized click streams, which are generated by using web logs. Several experiments have been conducted by using this technique by J.C. Penney. According to their reviews this Significant Usage Patterns can distinguish customers based on their actions. In this paper [2], a technique called visualization of navigational patterns is used based on model relied clustering. This model based clustering approach partitions the types of users. They analyze the paths through which user navigations are done within the same cluster. By focusing on the user requested web pages this technique can be implemented. In this paper[3],they implemented a technique called novel path clustering method which identifies the user interests based on client's clicked links, page viewing time, order and cache references. The information that is captured is used to discover knowledge of similar user interests. In this paper[4], Non-negative matrix factorization is applied to dimensionality reduction of the URL session matrix. It also uses the k-means algorithm to partition the user session vectors into several clusters. The several experiments conducted on this approach proves that user interests can be mined effectively.

III. EXISTING SYSTEM

In these existing system conventional methods like model checked and linear temporal models were used. The problem with this approach is difficulty in analyzing because of complex mathematical formulations. This technique uses the session wise weblogs to extract the user interests from the sequence of actions. To analyze the user interests individually new approach should be recommended.

IV. PROPOSED SYSTEM

The traditional methods like clustering and model based approaches were failed to analyze the user interests effectively. Hence we are promoting a new approach called Markov Clustering Algorithm which uses a specific process that repeats the several steps of convergence and inflation. This technique can take the random walks for the related item of a similar cluster. By using this technique the partitioning of data can be performed easily. This increases the accuracy of the knowledge discovering.

V. IMPLEMENTATION

A. Admin

In this module, the Admin has to login by using valid user name and password. After login successful he can do some operations such as List of users and authorize, Add and View domains, Add product Posts ,List all Product posts with rank, List of search history, Update product post details, View analysis of user search, View analysis of negative behavior, View Analysis of Positive behavior, View all reviewed Posts, View Analysis of Positive and Negative behavior in Chart

B. User

In this module, there are n numbers of users are present. User should register before doing some operations. After registration successful he has to login by using authorized user name and password. Login successful he will do some operations like View My Profile, Search products and review about the products, view other products reviews and give your review, view my search history, Search for top N Product posts
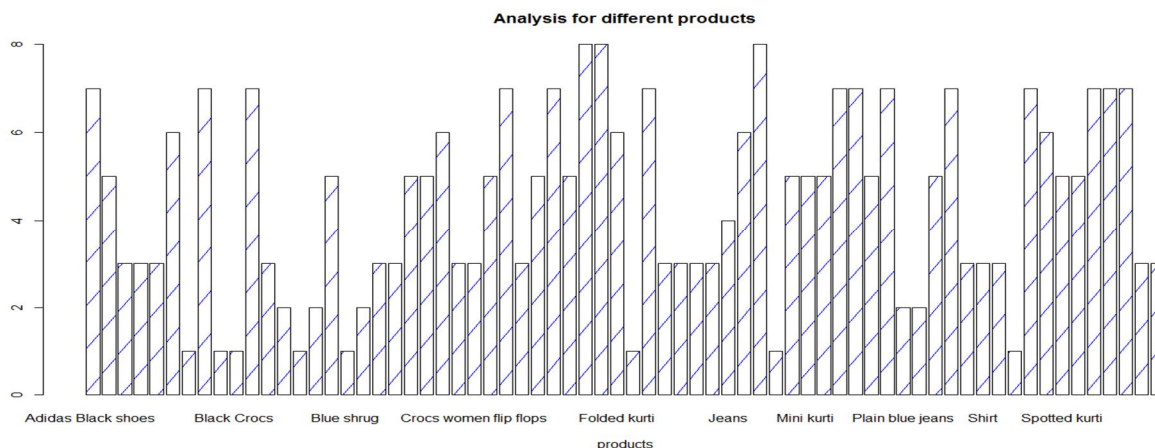

Fig 2 Analysis of user for different products

VI. CONCLUSION

In the case of open systems, where the sequences of interactions (stored as system logs) are not constrained by a workflow, process mining techniques whose objective is to extract interesting information.So, the idea of capturing user interactions with respect to sessions helps the analysts to determine user interests more accurately. In this paper we generated a barplot to analyze the user behavior for different products.

REFERENCES

[1] L. Lu, M. Dunham, and Y. Meng, ``Mining signi_cant usage patterns from clickstream data,'' in Proc. 7th Int. Conf. Knowl. Discovery Web Adv. Web Mining Web Usage Anal., Berlin, Germany, 2006, pp. 1_17

[2] Cadez, I.V., Heckerman, D., Meek, C., Smyth, P., White, S.: Visualization of Navigation Patterns on a Web Site Using Model Based Clustering. In: Proc. of 6th ACM SIGKDD Int'l. Conf. on Knowledge Discovery and Data Mining (2000).

[3] C. Shahabi, A. Z. Zarkesh, J. Adibi, and V. Shah. Knowledge discovery from users web-page navigation. In Proc. of 1997 Int. Workshop on Research Issues on Data Engineering (RIDE'97), Birmingham, England, April 1997.

[4] Jianjiang Lu, Baowen Xu, H. Yang. Matrix dimensionality reduction for mining Web logs. Proceedings IEEE/WIC International Conference on Web Intelligence (WI 2003), 405-408

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)