

Survey on Web Prediction Techniques

Rubab Sharma¹ Akanksha sambya²

Department of CSE SRI SAI University Palampur, India¹

V.P.O Sungal District, Kangra(H.P), India²

ABSTRACT: As we know web usage mining is one of the application of data mining techniques which is used to find out user patterns from web log data. So for this, it contacts the log files for extracting the useful information from the internet. When the user interacts with any websites, then the user's activities are automatically stored in web log server. Various researches had been proposed in this field but this paper shows the prediction of user browsing behaviour using web log data. This paper provides both the overview of current and past techniques used by different authors to reduce the searching time of user and also provide their research work limitations.

Keywords: Web Usage Mining, Web Log, User Browsing Behaviour.

I. INTRODUCTION

When a user accesses the web pages, then some information is stored in their web log. Through this content the web page navigational pattern of user can be dogged. As web usage mining consist of three phases i.e. data pre-processing, pattern discovery, pattern analysis. Firstly, all the web log data is pre-processed to get back logs with least amount of redundancies and user session. Secondly, pattern discovery is used to take out user direction-finding patterns. At last, pattern analyzing algorithm is applied to mine data.

This paper contains, literature survey on predicting user's future request - by using a range of methods and algorithms under web usage mining, which have been projected on this work and also this paper discuss about benefits and limitations of the technique used.

II. RELATED WORK

[1] Study and Implementation Of LCS Algorithm For Web Mining: Vrishali P. Sonanane planned the calculation of user routing patterns using LCS algorithm. They mark out the prediction of user behaviour process which contains different phases i.e. data re-treatment, direction-finding pattern mining, direction-finding pattern modelling, clustering and prediction engine. Using LCS algorithm, we can predict user's future request more accurately. In [2] Evaluation of Web Usage Mining Approaches for User's Next Request Prediction: Mathias Grey, Hatem Huddad recommended a scaffold for a system which predicts the user's next request based on their actions exposed from web log data. In this paper, three web usage mining techniques are combined i.e. association rules, frequent sequence and frequent generalized sequence. By using rule function selection, they identify two prediction strategies i.e. Last Sequence (LS) and Highest Confidence (HC) to select rules matching the pages requested by a user. After all the calculations, the output describes that frequent sequence provides better exactness as compared to association and frequent generalized sequence. In [3] An improved user browsing behaviour prediction using web

log analysis: Vedpriya Dongre, Jagdish Raikwal wished for a system architecture for discovering the hidden navigational patterns using web log data. In this, one or more client is connected to the server and then server generated a log. The data is first pre-processed and then clustering is done using k-mean algorithm to find out matching data. Now this clustered data is passed from the regression analysis, through which the numeric values of data is corrected so as to predict the user future request. In [4] WebPUM: A Web-based recommendation system to predict user future movements: Mehrdad Jalali, Narwat Mustapha, Md. Nasir Sulaiman, Ali Mannat developed their preceding work and renamed their structural design as Web PUM. They applied user routing patterns for the navigation pattern mining phase. They also applied LCS for classifying all the user's deeds to guess user near future request. In [5] A new classification model for online predicting users future movements. Mehrdad Jalali, Narwat Mustapha, Md. Nasir Sulaiman, Ali Mannat anticipated an online and offline phase design.

Both these designs works concurrently. In offline mode, all the web log data is under the data pre-processing module which processes and reformat data to recognize all web access sessions. All the like properties of the users are clustered under navigational pattern mining component. In online phase, the essential information is reorganized and the list of ideas are gathered in the prediction list. The semantic knowledge about essential domain is used to recover the quality of the suggestion. In [6] Improved Web Prediction Algorithm Using Web Log Data: Megha P. Jarkad, Prof.Mansi, Bhonsle listed a system architecture which consists of five steps pre-processing, classification, clustering and background algorithm and using LCS to predict user future request. By using classification, clustering and backtracking algorithm it improve the performance and also lower down the time complication of the proposed system. In [7] User Future Request Prediction Using KFCM in Web Usage Mining: Dilpreet kaur, A.P.sukhpreet kaur anticipated a structural design using Fuzzy Clustering i.e. fuzzy c-means and kernelized

fuzzy c-means algorithm. Firstly, web log data is fetched and then it enters the pre processing phase. After pre processing phase, fuzzy clustering algorithm is applied for user’s future request prediction and then results were calculated. Through this, conclusion is made that KFCM is more robust than FCM and also creates superior clusters for prediction. In [8] A New Web Usage Mining Approach for Next Page Access Prediction: Yogesh rajaram bhalerao, Prof. P.P. rokade deliberate a system design to predict user navigation pattern using statistical classifier and modern techniques. In this, dataset is pre-processed from various users and potential user is found out on the basis of the time spent by them in a website. Afterwards, the utmost likelihood classification algorithm is applied for prediction.

[4]. M. Jalali, N. Mustapha et al ,” WebPUM: A Web- based recommendation system to predict user future movements”, in international journal Expert Systems with Applications 37 (2010) 6201–6212 .
[5]. Megha P. Jarkad, Prof. Mansi, Bhonsle: “Improved web prediction algorithm using web log data” International journal of innovative research in computer and communication engineering. Vol. 3, Issue 5, May 2015.

III. SUMMARY OF LITERATURE SURVEY

S.NO	AUTHORS	METHOD	DISADVANTAGE
1	Vrishali P. Sonavane	LCS algorithm.	The LCS algorithm is used to alter distance between 2 sequences, so it is much quicker when the difference is small.
2	Mathias Grey, Hatem Haddad	Association rules, Frequent Sequence and Frequent Generalized Sequence.	The disadvantage of association mining is that it does not essentially use the idea of sequential distance but also frequent sequences cannot calculate direction-finding patterns for data sets.
3	Vedpriya Dongre, Jagdish Raikwal	K-means and Regression analysis.	The parameters are not discussed like time occupied for prediction and also memory taken to store the data.
4	Mehradad Jalali, Narwan Mustapha, Md. Yasin Sulaiman, Ali Mannat	Online-Offline phase of architecture, LCS Algorithm, clustering.	A websites which contain million of pages were miserably affected as memory needed to store web server pages is quadratic in number of pages. In this graph partitioned clustering algorithm is used instead of calculating weights of the web pages frequencies directly can be used for prediction.
5	Megha P. Jarkad, Prof.Mansi, Bhonsle	Classification, Clustering & Backtracking Algorithm	
6	Dilpreet kaur, A.P.rokharjee kaur	Fuzzy Clustering i.e. fuzzy c-means and kernelized fuzzy c-means algorithm.	The disadvantage in this work is that FCM is less robust and can't apply on the bulky sets.

IV. CONCLUSION

This survey shows how various authors use many different algorithms and techniques under web usage mining. A lot of research work had been done on user future request prediction. This paper focus on how prediction time can be increased without losing prediction accuracy. So, for this various algorithms of pattern discovery techniques have been applied. In future, the research work can be extended for a small number of previous log files and prediction accuracy level can be improved by using different techniques of data mining.

ACKNOWLEDGMENT

I express my deep sense of gratitude to **Sri Sai University, Palampur** for giving an opportunity to prepare this paper. I would like to express a deep sense of gratitude and thanks to Department of Computer Science and Engineering, **Sri Sai University, Palampur** for generous guidance, help and useful suggestions. I express my thanks to my advisor, who guided me and provided with all the useful information for presenting this paper.

REFERENCES

[1]. Vrishali P. Sonavane,”Study And Implementation Of LCS Algorithm For Web Mining”, International Journal of Computer Science Issues, Vol. 9, Issue 2, No 3, March 2012.
[2]. Mathias Gery, Hatem Haddad ,“Evaluation of Web Usage Mining Approaches for User’s Next Request Prediction” WIDM’03 Proceedings of the 5th ACM
[3]. Vedpriya Dongre, Jagdish Raikwal ,”An improved user browsing behavior prediction using web log analysis”, International Journal of Advanced Research in Computer Engineering and technology (IJARCCCE), Vol. 4, Issue 5, , May 2015.