

Prediction of User Behavior using Web log in Web Usage Mining

Virendra R. Rathod
PG Scholar, Shankalchand Patel college of Engineering

Govind V. Patel
Sankalchand Patel college of Engineering

ABSTRACT

Web usage mining is application of data mining. Web Usage Mining is the automatic discovery of user access pattern from web servers. Web usage mining is consists of preprocessing, pattern discovery, pattern analysis. Web prediction is a classification problem which attempts to predict the most likely web pages that a user may visit depending on the information of the previously visited web pages. In this paper emphasizes is given on the user Behaviour using web log file prediction using web log record, click streams record and user information. Here, two different clustering techniques, namely Fuzzy C-Means Clustering algorithms and Markov model has been investigated to predict the webpage that will be accessed in the future based on the previous action of browsers behavior. But prediction of future request of the user mainly concern with its accuracy and efficiency. The discovered patterns can be used for better web page access prediction. Prediction model are better prediction of next web page the user want to visit. Using web page access prediction, the right advertisement will be placed in the website according to the users' browsing patterns. In Web page prediction, the next action corresponds to predicting the next page to be visited. The previous actions correspond to the previous pages that have already been visited.

Keywords

Web mining, Web usage mining, Web Log File, Prediction, Pre-processing, Fuzzy C Means (FCM) Algorithm, Markov Model.

1. INTRODUCTION

Web usage mining is application of data mining technique to discover automatic discover from particular web site. Web mining can be categories into three categories are shown in Fig.1, web content mining, web structure mining and web usage mining. Web content mining is referring to extract information of content of web pages. Web structure mining is used to analyze the links between web pages through the web structure to infer the knowledge. Web usage Mining is extracting the information from web log file which is accessed by users [1]. The main aim of web usage mining is extract information from web pages which access by users.

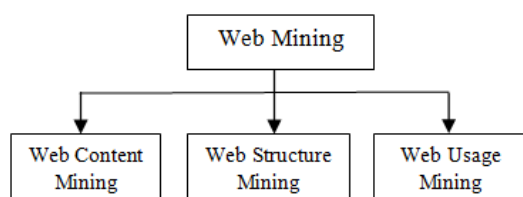


Fig 1: Types of web Mining

Web prediction is a classification problem which attempts to predict next set of pages user can access.

Intelligence Web Usage Mining users on the secondary web data such as web server logs ,proxy server logs browser logs user profiles registration data user session, transactions, cookies user queries ,bookmarks mouse clicks or any other data generated by interaction of user with the web [3]. Prediction is used to predict next set of web pages that are required by the user. Prediction of user's behavior is based on user's previous history and knowledge [2]. Web page is important to predict next page access because million of users generally access the similar web page in a particular Web site [8].

2. BACKGROUND

Web usage mining consists of three main steps: Pre-processing, Pattern Discovery and Pattern Analysis. The following Fig.2 shows this process.

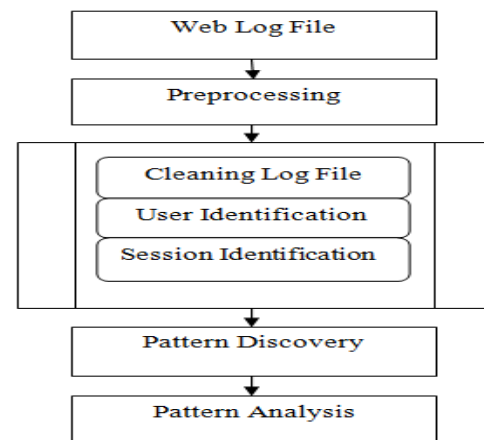


Fig.2: Web Usage Mining Process

2.1 Preprocessing

Preprocessing is first step of web usage mining process in which the unnecessary web log data or noisy data from web log so as to reduce its size. Remove the useless data noise from web log file so reduce it size.

ip	datetime	path
157.55.39.76	20/Jan/2016:05:08:55	GET /members/chetanarora/activity/following/ HTTP/1.1 404
157.55.39.201	20/Jan/2016:05:09:05	GET /members/sumit1992m/ HTTP/1.1 404
157.55.39.201	20/Jan/2016:05:09:17	GET /members/mahimaa/activity/following/ HTTP/1.1 404
157.55.39.201	20/Jan/2016:05:09:24	GET /members/mahnaaz2005/ HTTP/1.1 404
66.249.79.90	20/Jan/2016:05:09:25	GET /members/rponraj1988/media/5796/ HTTP/1.1 404
107.167.107.80	20/Jan/2016:05:09:25	GET /feed/ HTTP/1.1 404
66.249.79.88	20/Jan/2016:05:09:28	GET /members/vidyadhar
157.55.39.76	20/Jan/2016:05:09:40	GET /members/purvesh1971k/ HTTP/1.1 404
157.55.39.201	20/Jan/2016:05:09:49	GET /members/maniprabha/media/18574/ HTTP/1.1 404
157.55.39.201	20/Jan/2016:05:09:59	GET /members/swapnilg1990/activity/following/ HTTP/1.1 404
66.249.79.88	20/Jan/2016:05:10:06	GET /members/bhavesk1997/media/photo/ HTTP/1.1 404
157.55.39.201	20/Jan/2016:05:10:11	GET /members/manish1989k/activity/following/ HTTP/1.1 404
66.249.79.88	20/Jan/2016:05:10:24	GET /members/monuo1989/media/video/ HTTP/1.1 404
157.55.39.201	20/Jan/2016:05:10:39	GET /members/tanya123/ HTTP/1.1 404

Fig 3: Cleaning Log File

Fig.3 Preprocessing in first step of cleaning web log data is getting data from www.stellanstechsoft.com web server using NetBeans 8.0.2 tool and java program. Then using this web log file after user identification (identifying by different users Through IP address) and session identification (identifying different session) is done. The data preprocessing includes three basic steps is Data Cleaning, User Identification, Session Identification. Each user has a separate session. This can be done using simplest method as to expiration time. Default time for user session identification is thirty minutes.

2.2 Pattern Discovery

Second step of web usage mining process in which the cleaned log file generated in which in the Preprocessing step is used to discover web usage patterns. We used pattern discovery for fuzzy c means algorithm.

2.3 Pattern Analysis

Final step of web usage mining process in which the pattern discovered in the second step are further analyzed to generate more useful information related to the user behaviour pattern [3].

3. LITERATURE REVIEW

Kaushal Kishor Sharma, Kiran Agrawal proposed web usage mining to predict to the user is accessing the web page using novel algorithm and named it as predicting user navigation pattern using Clustering and Classification from web log data to predict user's navigation [3]. Web page prediction is helpful in many applications such as web advertisement area. We used web log CSV file from www.stellanstechsoft.com then cleaning of web log data find session based FCM frequency and page based frequency using user surfing through the website. Web page prediction is improving accuracy of web page access prediction. Using web page access prediction, the right advertisement will be placed in the website according to the users' browsing patterns. Also, web page access prediction helps the web administrators to restructure the Web site. By predicting the Web page, we can improve the browsing speed and navigation paths [7]. Prediction of user's behavior using web log in web usage mining for next page prediction based on previous history and knowledge of user. There are many problems in web page prediction using web log record in web usage mining. For better web page prediction in used clustering technique, algorithm and prediction model. A lot of the research work has been already done in next web page prediction using web log in web usage mining. Prediction is problem in which we need to predict next set of web pages which the user will visit knowing all the previous visits of the user in the session [2]. There are various techniques developed for better prediction of user's behavior and to give user next set of web pages. Meera Narvekar, Shaikh Sakina Banu

proposed a system which uses several Prediction Model such as Markov Model Hidden Markov Model and Hybrid model for prediction. However, Markov Model cannot predict a session [2]. This is used for prediction using Dumpsters rule. In this paper system consist of two tier frame work where in tier one is done offline where training of database is done using backward and forward probability. A main advantage of Markov Model is its efficiency and performance in terms of building a model a model. Anshul Bhargav and Munish Bhargav proposed framework which is based on three step: Preprocessing, pattern discovery and user classification. In this paper mainly focused on doing users classification. In the paper discusses about the Preprocessing, Pattern Discovery and Pattern Analysis. In this paper he proposed a novel algorithm and clustering and classification from web log data to predict user navigation. This paper introduces an efficient hybrid predictive model, which is combination of Markov Model and Bayesian theorem. In these paper users browsing behavior will be observed at two levels to meet the nature of the portal. Next using Bayesian theorem in the level two predictive the users browsing page is more effective and accurate. In this paper two stage prediction model used for next web page prediction of web pages [1].

4. RELATED WORK

We Proposed Web using mining technique to predict browsing behavior of user may visited based on previous history and knowledge using FCM(Fuzzy C-Means)Algorithm and Markov Model. First we used web log CSV file is collect from www.stellanstechsoft and then preprocess step is apply and remove irrelevant data from web log file. Then after fuzzy c means (FCM) algorithm apply to the pattern discovery and pattern analysis. We used Markov model used to next page prediction and better web page prediction accuracy. We proposed methodology of next page prediction by using pattern discovery and pattern analysis. Most of the data needed for web log analysis resides on web server proxy servers, enterprise logs, web clients etc [3] the first step is cleaning or preprocessing which removes unwanted data from web log file and reduces the size of web log file. In second step, users are classified into potential users and non potential users and only potential users are only considered for further processing. In third step clustering is done using graph partitioned algorithm and in last step user future request is predicted using Fuzzy C-Means (FCM) algorithm and Markov Model. Web page prediction is used to predict next set of web pages that are required by user. Prediction of user's behavior is based on user's previous history and knowledge [2]. We Proposed Web Usage mining technique to predict the browsing behavior of user using fuzzy Clustering methods such as Fuzzy C-Means.FCM based clustering algorithm is provided better result than that of k-means algorithm.In this, web log file data is collected and then preprocessing step is performed to clean irrelevant data and required attributes are chosen from log file. After that fuzzy clustering methods are implemented and user future requests are predicted. Web Prediction Problem can be broadly classified into two different categories, namely, point-based and path-based prediction models. Path-based prediction is based on user's previous and historic path data while point-based prediction is based on currently observed actions [5].In Web page Prediction the next action corresponds to the predicting next page to be visited. The previous action correspond to the previous page that have already been visited [5].For Example, In second order Markov model prediction of next web page the user visits is computed based only on two web pages previously visited [5] the system

improves prediction accuracy without compromising prediction time. The system consist of two tiers wherein tier one is used for training while tier one is used for training while tier two is used for prediction.

5. PROPOSED METHODOLOGY

The Proposed approach consists of following steps.

Step 1: Data is collected from web log file.

Step 2: Then Preprocessing applied to the web log data.It is performed in the second phase.

Step 3: After Preprocessing data is cleaning by remove useless data from web log file.

Step 4: Then User Identification (Identifying by different users Through IP address)

Step 5: In this Step Session Identification (Identifying different session)

Step 6: In this step then comes Pattern Discovery followed by Pattern Analysis.

Step 7: Then apply FCM (Fuzzy C-Means) Clustering algorithm and Markov Model is used for next page prediction.Markov Models have been widely used for predicting next Web-page from the users' navigational behavior recorded in the Web-log.

Step 8: In this step Pattern analysis is done. (On the basis of Pattern analysis Prediction is done of the User's Future request)

In the above steps are being described diagrammatically.

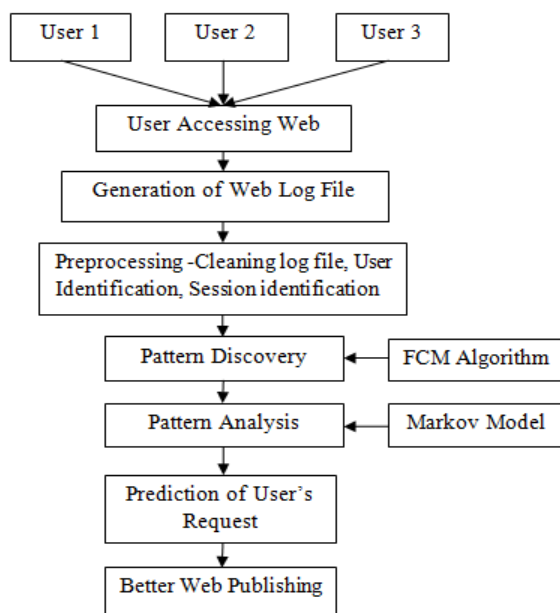


Fig 4: Methodology of Proposed approach

In this diagram there are eight steps in first step the user is access on web site. Then generation of web log files and web usage mining for Preprocessing is required. Then Comes Pattern discovery follow by Pattern analysis. After pattern discovery is applying fuzzy c means algorithm.

Then Markov Model is used for next page prediction. In last step is Pattern Analysis on the basis of pattern analysis Prediction is done of user's future request. Fuzzy C-Mean clustering algorithm is one of the most widely used fuzzy

clustering algorithms. This algorithm works by assigning membership to each data point corresponding to each cluster centre on the basis of distance between the cluster centre and the data point^[9]

Let $X = \{x_1, x_2, x_3 \dots, x_n\}$ be the set of data points and

$V = \{v_1, v_2, v_3 \dots, v_c\}$ be the set of centers.

Step1. Randomly select 'c' cluster centers.

Step2. Calculate the fuzzy membership ' μ_{ij} ' using:

$$\mu_{ij} = 1 / \sum_{k=1}^c (d_{ij} / d_{ik})^{(2/m-1)}$$

Step3. Compute the fuzzy centers ' v_j ' using:

$$v_j = \left(\sum_{i=1}^n (\mu_{ij})^m x_i \right) / \left(\sum_{i=1}^n (\mu_{ij})^m \right), \forall j = 1, 2, \dots, c$$

Step4. Repeat step 2) and 3) until the minimum 'J' value is achieved or $\|U(k+1) - U(k)\| < \beta$.

Where,

'k' is the iteration step.

' β ' is the termination criterion between [0, 1]

' $U = (\mu_{ij})^{n \times c}$ ' is the fuzzy membership matrix.

'J' is the objective function^[9]

6. CONCLUSION

From my study various research papers on basis web usage mining we conclude that web usage mining is extracting the information from web server log file which access by users. Web usage mining is done using three main steps Preprocessing, Pattern discovery, Pattern Analysis. Web Usage Mining is process of applying data mining techniques to the discovery of usage patterns from Web log data. We used web log CSV file getting from www.stellanstechsoft and then preprocess step is apply and remove irrelevant data from web log file. Then after fuzzy c means clustering algorithm is find session based FCM frequency and page based FCM using most frequently user surfing through web sites .We find each user find session based and page based FCM frequency based on most time spent on web pages and frequent surfing through web sites. Web page prediction is problem we may try to predict the next web pages that user may go on web site based on the knowledge of previously visited web pages. We have discussed about prediction of user behavior based on web log is generated on web server based on previous history or behavior or knowledge of user next web page prediction. FCM is mainly used to analyze the access patterns of the organization. We used FCM based clustering algorithm provides best results for overlapped data sets comparatively better than that of k-means algorithm. My research work improve of web page access prediction accuracy by combing both Markov Model and with use FCM algorithm and Markov model and Fuzzy Clustering algorithm can work together and provide better web prediction results without compromise with accuracy. In Web page prediction, the next action corresponds to predicting the next page to be visited. The previous actions correspond to the previous pages that have already been visited.

7. REFERENCES

- [1] V.V.R. Maheswara, Dr. V. Valli Kumari, efficient hybridpredictive model to analyze the visiting characteristics of Web User using Web Usage Mining, 2010 IEEE.
- [2] Meera Narvekar,Shaikh Sakina Banu,"Predicting User's Web Navigation Behavior Using Hybrid Approach ", ELSEVIER Science Direct,2015.
- [3] Kaushal Kishor Sharma,Kiran Agrawal," A Hybrid Approach for Predicting User's Future Request",2014,IEEE.
- [4] Anshul Bhargav, Munish Bhargav ,"Pattern Discovery and Users Classification through Web Usage Mining" 2014 IEEE.
- [5] Priyanka S. Panchal,Prof. Urmi D. Agravat,"Hybrid Technique for User's Web Page Access Prediction based on Markov Model",4th ICCCNT 2013, IEEE.
- [6] Mamoun Awad and Issa Khalil,"Prediction of User's web-browsing behavior Application of Markov Models" 2015, IEEE.
- [7] Abdelghani Guerbas, Omar Addam, Omar Zaarour ,"Effective web log mining and online navigational pattern prediction", Science Direct, 2013.
- [8] Phyu Thwe, "Using Markov Model and Popularity and Similarity-based Page Rank Algorithm for Web Page Access Prediction",ICAET'2014.
- [9] Fuzzy C-Means by Balaji K and Juby N Zacharias.
- [10] Dilpreet Kaur1, A.P. Sukhpreet Kaur,"User Future Request Prediction Using KFCM in Web Usage Mining", IJARCCCE, 2013.
- [11] Mayank Kalbhor , Kuntl Jain, "Fuzzy Based Hybrid Approach for User Request Prediction Using Markov Model",IEEE,2015.