IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 10, October 2016

Improved Next Web Page Recommendation using Multi-Attribute Weight Prediction

Prof. Umesh A. Patil[1], Mr. Avinash Kunnure[2], Mr. Vaibhav Herwade[3], Mr. Vijaykumar Dawarpatil [4], Mr. Satyawan Magar [5]

Assistant Professor & Head of the Department, Computer Science and Engineering, D. Y. Patil Technical Campus, Faculty of Engineering & Faculty of Management, Talsande, Kolhapur, India [1]

UG Student, Computer Science and Engineering, D. Y. Patil Technical Campus, Faculty of Engineering& Faculty of Management, Talsande, Kolhapur, India[2,3,4,5]

Abstract: Web page prediction is the web usage mining by performing pre-processing of the data from a web site. The need for predicting the user's needs in order to improve the usability and user maintenance of a web site is more than marked now a day's lacking proper guidance, a visitor often wanders aimlessly without visiting significant pages,loses attention, and leaves the site earlier than expected. When they access the network, a large amount of data is generated and is stored in Web log files which can be used efficiently as many times user repeatedlysearched the same type of Web pages recorded in the log files. These series can be considered as a web access pattern, helpful to find the user behaviour. Through this personalized information, it's quite easy to predict the nextset of pages user might visit based on the previously searched patterns, thereby reducing the browsing time of auser.

Keywords: Web usages mining, recommendation, web log analysis, session based predication, K-NN algorithm.

I. INTRODUCTION

Web page prediction is the web usage mining by performing pre-processing of the data from a web site. Web prediction is a classification problem whichattempts to predict the most likely web pages that a user mayvisit depending on the information of the previously visitedweb pages. The needfor predicting the user's needs in order to improve the usability and user maintenance of a web site. Web usage mining is widely used to discover the usage patterns from web log files. It deals with web log data which are taken from web servers, proxy server or client's cache. The proposed webrecommendation system is concept by which theprevious or historical user navigation data is analyzed and based on the navigation technique; the next web pageaccess is predicted. The proposed recommendation system has some relevant concepts such as behaviour analysis of user access patterns, personalizationof data and predictive modelling. The behaviour of users accessed data is extractedusing the K-mean clustering algorithm. Then search the similar user behavioursfrom the web log using KNN algorithm which analyze data in distance based function and most nearest patterns are listed with the help of user frequent patterns. From nearest frequent pattern, the timebased data clustering is also prepared to amount of time spent on a particularURL in the entire log file. After evaluation of these parameters namely usernavigational frequency and time based frequency a combine weight for all the URLs areevaluated. These weightsare further sorted and by the rank ofweights the next most possible webpage is predicted. This project is contributed that improves the recommendation accuracy based on the session of user's web access. It provides more appropriate recommended web page to the active user.

II. LITERATURE REVIEW

An automatic web usage data mining and recommendation system based on current user behaviour through his/her click stream data onthe newly developed Really Simple Syndication (RSS) reader website, in order toprovide relevant information to the individual without explicitly asking for it. The K-Nearest-Neighbour (KNN) classification method has been trained to beused on-line and in Real-Time to identify clients/visitors click stream data, matching it to a particular user group and recommend a tailored browsingoption that meet the need of the specific user at a particular time.

III.PROPOSED WORK

This project focus is to develop a next web page recommendation system using web access logs.The data in the log files of the server about the actions of the users can not be usedfor mining purposes in the form as it is stored. For this reason the data should be pre-processed to improve the efficiency andease of the mining process. The main task of data pre-processingis to prune noisy and irrelevant data. The proposed web page recommendation system contains theK-means algorithm which is used to group of data according to theuser IP address for finding

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 10, October 2016

the similar access patterns of theuser sessions. Additionally for classification and prediction theKNN algorithm is implemented. The KNN algorithms to analyze data in distance based function and most nearest similar patterns are listed which is belongs from the otheruser therefore the proposed model also predicts the rarelyaccessed patterns. Thus to make the recommendations webusages data is personalized, based on URL frequencies, usernavigational frequenciesand timebased data analysis. Additionally to combine these parameters aweighted technique is used. A combine weight for all the URLs is evaluated. According to the obtained weights theURLs are sorted and the maximum weight is selected as prediction of recommendation system.

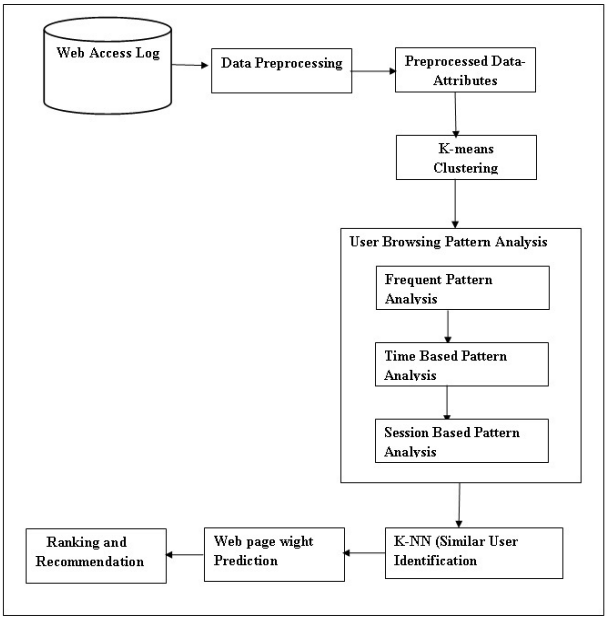

Fig. 1. Proposed Model

The highest weightshows the higher probability of visiting a web page after thecurrent navigated web page. This project is contributed that improves the recommendation accuracy based on the session of user's web access. It provides more appropriate recommended web page to the active user.

IV. IMPLEMENTATION

Steps:
A. Module 1: Data Pre-processing
B. Module 2: User Browsing Pattern Analysis
C. Module 3: Time based Browsing Pattern Analysis
D. Module 4: Similar User Browsing Pattern Analysis
E. Module 5: Next Web Page Recommendation

A. Module 1: Data pre-processing
a. Input: Raw Web Server Log File
b. Output: Pre-processedLog file
Web logs contain multiple records and information. When userenters on to the web, what data he accessed by him, howmany times, which websites are visited mostly, IP addressof that users system and many more information. Web logs also contain the error or failure entries, some access recordswhich are generated by search engine. Hence datapre-processing step perform data cleaning, formatting andgrouping operation. In data cleaning all unwanted entries areremoved and only that entries are extracted which are useful for recommendation operation.

c. Attributes Selection
During the pre-processing of log files the selected or targetedattributes are extracted and preserved in a database. Theseattributes are used for computing the different parameters onwhich the prediction of next web page is performed. It contains different kinds of attributes i.e. IP address, time stamp, requested URL, browser information and others. Among them some of the data is required for developing the proposed recommendation system and not all the attributes are used.

B. Module 2: User Browsing Pattern analysis
a. Input: Pre-processedLog file
b. Output: Frequent Pattern for each web log user
c. k-means Clustering
Each user accessed data is extractedusing the K-mean clustering algorithm. K-means clustering isapplied on the data to prepare group of data according to theuser IP address. Each user IP address is represented as centroid in clustering. From the log file IP address from each entry and the corresponding access pattern is processed and merged to the closest centroid. Finally number of groups is obtained based on the IP address that contains the individual user's web browsingpattern.
d. Frequent pattern analysis
The individual user's web browsingpattern is identified; find the mostfrequent accessed web pages for each user. The frequency of the individual web pagesaccessed by the user the following formula can be used.

C. Module 3: Time based Browsing Pattern Analysis
a. Input: PreprocessedLog file
b. Output: URL Access Time and URL Session for each web log user
The pre-processed log data, the timestamp is analysed for a individual user browsing pattern. The amount of time spent on a particular web page is calculated using the following formula,
c. Session based Browsing Pattern Analysis
In this module the session based navigational pattern is analysed. A session is a list of web pages accesses from a given user during a period of time. For the task of identifying the list of web pages visited during a user's session at morning, afternoon or evening likewise. It provides more appropriate recommended web page to the active user.

D. Module 4: Similar User Browsing Pattern Analysis
a. Input: Frequent Pattern of Web log User
b. Output: Extract Similar user Browsing Pattern

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 10, October 2016

The k-NN classification algorithm to identify the target users search pattern, matching it to a web logs user group. It takes target user previous logs (frequent pattern) as a inputand find out which user access the same pattern, from that datait predicts the users interest. The neighbours (similar user) of target user browsing pattern is evaluated by measure the Euclidean Distance between the target user frequent pattern and allthe web log user frequent pattern. From k-minimum distance from web log user, most nearest pattern of active user will extracted.

E. Module 5: Next Web Page Recommendation
a. Input: Nearest Browsing Pattern
b. Output: Recommended the Next Web Page
c. Multi-Attribute URL Weight Prediction
From similar nearest frequent pattern are identified, the strength of nextupcoming URL is computed based the multi-attribute browsing patten. The multi-attribute parameters namely usernavigational frequency, time based URL and session based URL to combine weight for all the URLs areevaluated.

These weights are further sorted and by the rank ofweights the next most possible web page is predicted. According to the current user inputpattern system generate the prediction of next web page.

V. ALGORITHM

1. Start.
2. System will access the log files of user and do pre-processing to find most visited web pages and also removing an error entries from log files(error such as 404 page not found, connection failed, internal server error etc.).
3. Finds the frequent pattern of web logs user using formula:

Frequency (Webpages) =1/ N ∑i=1toN (pagecount$_i$)

N-Total number of pages accessed by specific userPagecount$_i$-URL Frequency

4. Calculation of URL Access Time and URL Session for each web log user

Time (Webpages)=1/N∑i=1toN(webpagetime$_i$)

N-Total amount of time web pages accessed by specific userwebpagetime$_i$-URL Accessing TimeSession Formula

Session1(Webpages)=1/N1∑i=1toN1(sessionpagecount$_i$)

N1-Total number of pages accessed by session1sessionpagecount$_i$-URL Frequency for session1

Session2(Webpages)=1/N2∑i=1toN2(sessionpagecount$_i$)

N2-Total number of pages accessed by session2
Sessionpagecount$_i$-URL Frequency for session2

Session3(Webpages)=1/N3∑i=1toN3(sessionpagecount$_i$)

N3-Total number of pages accessed by session3
sessionpagecount$_i$-URL Frequency for session3

5. Calculate the weight of webpages and using formula:

Weight Webpages)=w1*Freq(Webpages)+w2*Time(Webpages)+w3* session(Webpages)

6. Using weight (webpages) recommend the web pages to the users
7. End.

VI. CONCLUSION AND FUTURE WORK

Data pre-processing is an important task of Web log mining application. Therefore, data must be processed before applying data mining techniques to discover user access patterns from web log. The Proposed web recommendation system is concept by which theprevious or historical user navigation data is analysed andbased on the most likely navigated technique the next web pageaccess is predicted.This project is contributed that improves the recommendation accuracy based on the session of user's web access. It provides more appropriate recommended web page to the active user.
In future work we suggest use association rule with the" Improved Next Web Page Recommendation using Multi-Attribute Weight Prediction" to improve accuracy of web page recommendation.

REFERENCES

[1]. Arvind Verma, Balwant Prajapat, "User Next Web Page Recommendation using Weight based Prediction", International Journal of Computer Applications (0975 8887) Volume 142, No. 11, May 2016.
[2]. K. Srinivas, P. V. S. Srinivas, A. Govardhan, V. Valli Kumari, "Periodic Web Personalization for Meta Search Engine", IJCST Vol. 2, Issue 4, Oct-Dec. 2011
[3] Neha Sharma &Pawan Makhija, Web usage Mining: ANovel Approach for Web user Session Construction,Global Journal of Computer Science and Technology: ENetwork, Web & Security,Vol. 15, Issue 3, 2015.
[4] Haidong Zhong, Shaozhong Zhang, Yanling Wang, Shifeng Weng and Yonggang Shu, "Mining Users Similarity from Moving Trajectories for MobileEcommerce Recommendation, International Journal ofHybrid Information Technology Vol.7, No.4,pp.309-320, 2014.
[5] Zahid Ansari, A. VinayaBabu, Waseem Ahmed and Mohammad Fazle Azeem, "A Fuzzy Set TheoreticApproach to Discover User Sessions from WebNavigational Data", IEEE Recent Advances in IntelligentComputational Systems.
[6] I. Petrović, P. Perković and I. Štajduhar, "A Profile- and Community-Driven Book Recommender System, 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2015.
[7] Lina Yao and Quan Z. Sheng, Aviv Segev, Jian Yu, "Recommending Web Services via CombiningCollaborative Filtering with Content-based Features", IEEE 20th International Conference on WebServices,2013.
[8] Quanyin Zhu, Hong Zhou, Yunyang Yan, Jin Qian and PeiZhou, "Commodities Price Dynamic Trend AnalysisBased on Web Mining", Third InternationalConference on Multimedia Information Networking andSecurity, 2011.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 10, October 2016

BIOGRAPHIES

Prof. Umesh A. Patil is working as Assistant Professor & HOD, Computer Science and Engineering at D. Y. Patil Technical Campus, Faculty of Engineering& Faculty of Management Talsande, Kolhapur, India from last 6 Years. He had published 7 International Papers & Attend 2 National Conferences. His area of interest is Operations Research, Data Mining, and Theory of Computation & Computer Algorithms.

Mr. Avinash Kunnure UG Student, Computer Science and Engineering, D. Y. Patil Technical Campus, Faculty of Engineering Talsande, Kolhapur, India. His area of interest is Operations Research, Data mining, Java, and Data Structure & Computer Algorithms.

Mr. Vaibhav Herwade UG Student, Computer Science and Engineering, D. Y. Patil Technical Campus, Faculty of Engineering Talsande, Kolhapur, India. His area of interest is Operations Research, C, C++, Data Structure.

Mr. Vijaykumar Dawarpatil. UG Student, Computer Science and Engineering, D. Y. Patil Technical Campus, Faculty of Engineering Talsande, Kolhapur, India. . His area of interest is Operations Research, Assembly language, C, C++, Data Structure.

Mr. Satyawan Magar. UG Student, Computer Science and Engineering, D. Y. Patil Technical Campus, Faculty of Engineering Talsande, Kolhapur, India. His area of interest is Operations Research, C, Networking.