

See discussions, stats, and author profiles for this publication at: <http://www.researchgate.net/publication/270880736>

Web Navigation Model based on linear probabilistic approach

ARTICLE · AUGUST 2011

DOWNLOADS

16

VIEWS

16

3 AUTHORS, INCLUDING:



Diksha Dani

Institute of Public Enterprise

6 PUBLICATIONS 1 CITATION

SEE PROFILE



Web Navigation Model based on linear probabilistic approach

Dr.Arvindeer Kaur
USIT, GGSIPU
New Delhi, India
arvinderkaurtakkar@yahoo.com

Diksha Dani* & Kamna Singh
CSE Dept.
IPEC College, Ghaziabad, India
dikshadani@yahoo.com

Abstract: All web applications are made up of a set of pages. Navigation is one of the most important aspects of a web design. Therefore one of the primary concerns of a web application is to manage the navigation between these pages. User's web behavior is unpredictable. The user's eye scans across a page to decide what link to click on. So it is necessary to study and track the user navigation behavior in order to predict the next page accesses. Web navigation prediction is one of the most important topics for discussion in research area of Web navigation pattern mining. This paper proposes a probabilistic web navigation model that predict the sequence of web pages likely to be accessed by a users after entry into the website. It is based on linear probabilistic approach. The model resembles markov model.

Keywords: navigation model, navigation prediction, web

I. INTRODUCTION

The intent of users of search engine can be divided into three categories navigational, transactional and informational. People with informational intent want to find information primarily for research purpose. This includes a set of pages for the site giving him the information he needs. People with a transactional intent, have a purpose to reach a site where further interaction will happen. A user with a navigational intent is focused on finding a specific website related pages. The users switch from one page to next page of the website by clicking on the links according to their need.

Prediction of user's consecutive web page accesses is a big challenge for researchers in the web engineering area. If people cannot navigate through the site, they will quickly leave. Thus, designing effective navigation on the Web site is crucial. To make navigation more effective if we are able to estimate next user's request with sufficient accuracy, then based on this information we could modify behavior of a web systems to accommodate user's needs and meet his expectations. Moreover, response time reduction by loading the user's request before the actual access can be provided by web prefetching that will further enhance web navigation efficiency.

The Web can be seen as a structure containing information about hyperlinks, Web usage information, and Web contents in itself. Web site usage data, which contain records of how user has visited a Web site, can be used to identify collective user behavior in using the Web site, and use it as a base information for its predictive model.

The data of Web usage of the Web sites is extracted from the web server log files and page tagging. A Web log file is a collection of records of user requests of documents on a Web server updated on frequently with new information. Typical web log record contains following fields: IP address of the computer from which the request was made, User ID (Identification), date and time stamp of the request, URL of the requested document, status indicating whether the request was successful, document size, referring URL, and name and version of the browser and operating system being used for making the request.

However, due to the influence of caching on the user side and proxy server, not all the requests are recorded in Web log files. But in page tagging the JavaScript is automatically run every time the page is loaded so there are fewer worries about caching. Also Page tagging can report on events which do not involve a request to the web server, such as interactions within Flash movies. Logfiles contain information on failed requests; page tagging only records an event if the page is successfully viewed.

With huge amount of information available, only the web usage mining is being extensively used in research. Web usage mining is a process of extracting useful information from server logs i.e users history. It is the process of finding out what users are looking for on the Internet. It consists of web log preprocessing, request pattern discovery and pattern analysis.. There are several approaches to build predictive models from web usage data.

II. RELATED WORK

There are various web navigation modeling approaches for web application that have been found in the literature. The problem of modeling and predicting a user's browsing behavior on a Web site has been addressed by many researchers by different approaches. They try to capture the navigation paths that user traverse and explore to get the desired information.

The web navigation models that are proposed include [10-13], such as the models based on extended UML notation [14], Statecharts [15], etc

The navigation process of a user on a Web site can be modeled as a Markov chain. Markov models have been extensively used to model Web users' navigation behaviors on Web sites. Jianhan Zhu et al proposed a hierarchical clustering algorithm called CitationCluster to cluster conceptually related Web pages on a Web site based on cocitation and coupling similarities between Web pages defined on the transition probabilities on their in-links and outlinks, respectively. The clustering results are then used to construct a conceptual hierarchy of the Web site and then Markov model based link prediction is integrated with the

hierarchy to assist user navigation in a prototype called ONE [1].

Naïve Bayesian method approach is used for modeling and predicting users' navigation behavior using Web server logs as source data, for Web usage mining by Mahdi Khosravi et al [9].

Wen-long LIN et al presented novel website structure optimization model for more effective web navigation is proposed [5]. A web structuring model was developed by Wookey Lee et al and also designed a top-k tree algorithm to provide users with recommendations and directions for Web navigation [6].

An online prediction model based on LZ78 and LZW algorithms that are adapted for modeling the user navigation paths in web was proposed by Alborz moghaddam et al [7]. Antonio Maratea et al proposed a heuristic algorithm, called Trust-The-Unknown, that mimics human behavior in an unknown environment and and predict the next page category visited by a user [8].

Janez Brank et al modelled the user's habitual behaviour to analyze usage patterns and pro-actively present links for Web revisitation [2]. Silvia Gordillo et al presented an approach for modeling, analyzing and composing navigational concerns in Web applications to address the impact of crosscutting concerns in design models, improve the discovering of meaningful design artifacts and improve traceability of design decisions. [3].

Jose Borges proposed new method to evaluate the summarization ability of the model, making use of the Spearman foot rule metric to assess the accuracy with which a model represents the information content of a collection of user Web navigation sessions [4].

III. PROPOSED MODEL

A. Conceptual Model:

This model is based on the reasoning that past history can be used to predict the future events. For this we need to extract the usage pattern of users path in a session and use it to predict the sequence of pages that user is likely to visit. A session here is the time of navigation from entry to the website till the exit from the website. The model is based on the access pattern of the population. The quality of the web page access is improved by prefetching the documents that have high likelihood of access.

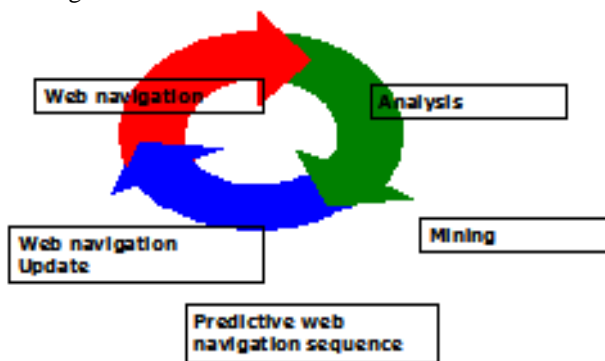


Figure 1: web navigation linear approach model

Step 1: Good navigation starts with the very first page that visitors see so Let h_i the any landing page of a particular website .

$$H_i = \{h_1, h_2, \dots, h_n\}$$

Where n is the number of pages in the site

Step 2: Let r_j be the next page surfed by the user after step 1.

$$R_j = \{r_0, r_1, r_2, \dots, r_n\}$$

Where $r_0 = \{\phi\}$ i.e user has left the site after previous page and

Let us define an array $k[]$ which stores the number of users hitting the page after h_i , for each page.

$$K1 = \{r_{11}, r_{12}, r_{13}, \dots, r_{1n}\}$$

Where each r_{1i} total number of users hitting a particular page k_i in time t_i , where I varies from 1 to n.

Similarly

$$k2 = \{r_{21}, r_{22}, \dots, r_{2n}\}$$

.

.

.

.

$$kn = \{r_{n1}, r_{n2}, r_{n3}, \dots, r_{nn}\}$$

Where r_n is the any page after any page r_{n-1} .

Step3: Site navigation predictive sequence is given by

$$Pr(s) = h_i + \sum_{j=0}^n r_j$$

Step 4: Computation of any page r_j

Let us also define a variable x_1 which stores the node or the page having the maximum number of users $r_1[ij]$ ie $x_1 = \max[k_1[r_1[j]]]$. This x_1 becomes our next node in the navigation sequence

Step 5: Similarly as above, the probable sequence of navigation by any user for a particular site can be written as $\{h_i, \max[k_1[r_1[j]]], \max[k_2[r_2[j]]], \dots, \max[k_n[r_n[j]]]\}$

It can be written as

$$\{h_i, x_1, x_2, \dots, x_n\}$$

Where $x_n = \max[r_i[ij]]$ n is the number of pages in the website. And h_i is the landing page and x_1, x_2, \dots, x_n is the predictive sequence or the path that any user will take while surfing the site after landing on a page of the site.

Using the above algorithm we can form an equation

$$Y = a_1/N * h_i + a_2/N * r_1 + \dots + a_n/N * r_{n-1}$$

Where a_1 is the number of users surfing the page h_i for time t and N is the total number of users of the landing pages. a_2 is the number of users surfing the page r_2 and N is the total number of users of the next page after h_i .

B. Procedure for Site Navigation:

Input : h_i -> first random page that is hit by the user of the website or entry page.

r_j -> is the next page after h

r_{1i} -> total no. of users on particular page k

x_i -> $\max[r]$

N -> Total no. of Users navigate a particular page in a time t after page h

$P(S_i)$ -> probability of site navigation

Pseudocode:

Step1

For each entry page h_i in the website

for($i=1; i \leq N; i++$)

{

while ($current_url = r_i$)

{

for($j=0; j \leq N; j++$)

{

$r_i[j] = r_{ij}$

}

```

}
}
    if  $r_j \neq 0$ 
        Then
            for( $j=0; j \leq n; j++$ )
                 $P(S) = h_i + \sum r_j$ 
Step 2: Construct  $k_i [f_{1N} f_{2N} \dots f_{NN}]$  //Where  $f_{in}$  is the
frequency of visit of each page after previous page.
    determine  $\max k_i [f_{in}] = x_i$ 
next_url=page_having(xi)
seq (i) = {  $h_i$ , next_url( $x_i$ ),.....next_url
( $x_N$ ) }
     $Y = [(x_i * h_i / N) + (x_1 * r_1 / N) + \dots + (x_{n-1} * r_{n-1} / N)]$ 
    Update(analysis)
Step 4: If(entry page is the exit page) then
{
     $P(S_i) = h_i$ 
     $r_i = \text{lastpage}(h_i)$ 
    seq=  $h_i$ 
}
     $Y = x_i * h_i / N$ 
    Update (analysis-1 )

```

C. Explanation of Pseudocode:

Let us suppose user arrives on any page i of the website (i. e. $hit=i$). Therefore the value of ' i ' varies from 1 to p , where p is the total no. of pages in the particular website.

Next the user navigate the next page .If the user does not navigate the next page, it means that $j=0$ and the first navigated page will be the last page that the user navigates($r_i=\emptyset$).

If the user navigates next page of the website means $r=1$ so we can say that value of ' r ' varies from 0 to n . Here it is important to understand that each time user navigates a page, the next page available for navigation in the sequence are all the page of the website except the one in which the user is currently in. For example let us suppose that a website has 4 pages. A number is taken for convenience of understanding. let us number the pages 1,2,3,4 for better understanding. Suppose any user arrives on a page 1 of the website. Now next page in the sequence will be pages 2,3,4 but not 1.

Therefore in this case sequence of navigation will be

$$P(S_i) = h_i + \sum r_j$$

Now the frequency of each next page visited by the user from the current page is observed and the maximum frequency is marked as the next page of the sequence of the current page. Therefore we conclude the navigation sequence is

$$\text{Seq}(i) = \{ h_i, x_i, \dots, x_N \}$$

The class diagram of the web navigation can be represents as

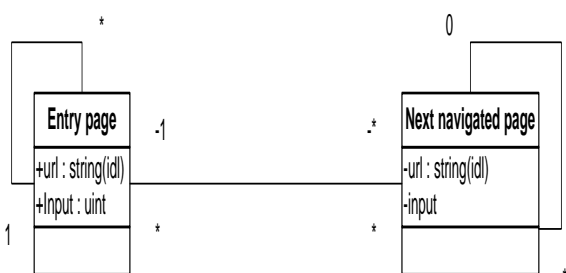


Figure2: web navigation class diagram

Above figure explains that if the entry page is h_i so the sequence of navigation will be started from 1 to many pages. Now for the next navigation of the web page by the user it may be user finish his navigation at the very first page or entry page or user continues navigation to entry page to other pages. Therefore we conclude that next navigated page sequence will be started from 0 to many pages. The relationship between the entry page and next navigated page will be one to many.

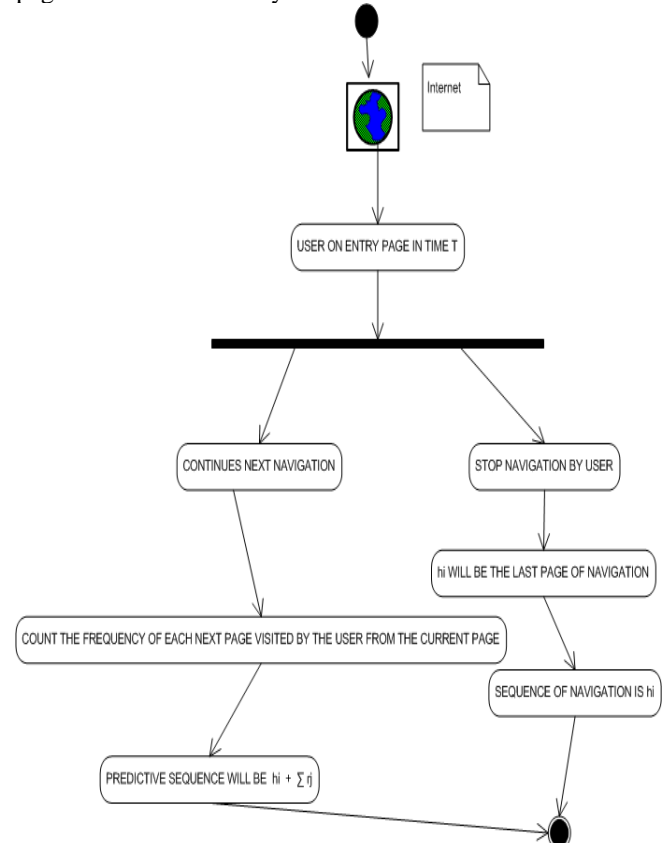


Figure 3: activity diagram of web page sequence

IV. CONCLUSION

Modeling and predicting user surfing paths involves tradeoffs between model complexity and predictive accuracy. In this paper we have proposed a model for predicting web navigation pattern. The dynamic manner of proposed method makes it suitable for dynamic web sites. But the complexity increases with the number of pages. For further work, we plan to implement it and validate it.

V. REFERENCES

- [1] Jianhan Zhu, Jun Hong, John G. Hughes(2002): "Using Markov Models for Web Site Link" HT'02, June 11-15, College Park, Maryland, US .
- [2] J. Brank , N. M. Frayling , A. Frayling, G. Smyth(2005). "Predictive Algorithms for Browser Support of Habitual User Activities on the Web", Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence (WI'05)
- [3] S. Gordillo, G. Rossi1, A. Moreira, J. Araujo ,C. Vairetti1, M. Urbieto(2006). "Modeling and Composing Navigational Concerns in Web Applications. Requirements and Design Issues". Proceedings of the Fourth Latin American Web Congress (LA-WEB'06)

- [4] Jose ´ Borges and Mark Levene (April 2007). “ Evaluating Variable-Length Markov Chain Models for Analysis of User Web Navigation Sessions” . IEEE Transactions On Knowledge And Data Engineering, Vol. 19, No. 4,
- [5] Wen-long LIN ,Ye-zheng LIU(2008) “A Novel Website Structure Optimization Model for More Effective Web Navigation1” ,In First International Workshop on Knowledge Discovery and Data Mining (WKDD).
- [6] Wookey Lee ,Carson K.-S. Leung(2009), “ Structural Top-k Web Navigation with Inclusive Query” . Published in Proceeding ICIT '09 Proceedings of the IEEE International Conference on Industrial Technology
- [7] A Moghaddam, E. Kabir(2009), “Dynamic and memory efficient web page prediction model using LZ78 and LZW algorithms” In the proceeding of 14th International CSI Computer Conference, CSICC 2009.
- [8] A. Maratea, A. Petrosino(2009), “An Heuristic Approach to Page Recommendation in Web Usage Mining”, IEEE Proceedings of the 14th International CSI Computer Conference (CSICC'09).
- [9] M. Khosravi, M.J Tarokh,(2010). “Dynamic Mining of Users Interest Navigation Patterns Using Naive Bayesian Method” IEEE International Conference on Intelligent Computer Communication and Processing (ICCP).
- [10] Alfaro, L. (2001). “Model Checking the World Wide Web”. In: Berry, G., Comon, H., Finkel,A. (eds.) CAV 2001. LNCS, vol. 2102, pp. 337–349. Springer, Heidelberg
- [11] Baumeister, H., Knapp, A., Koch, N., Zang, G(2005) “Modelling Adaptivity with Aspects”. In: Lowe, D.G., Gaedke, M. (eds.) ICWE. LNCS, vol. 3579, pp. 406–416. Springer, Heidelberg (2005).
- [12] Ceri, S., Daniel, F., Demaldé, V., Facca, F.M. (2005) “An Approach to User-BehaviorAware Web Applications”. In: Lowe, D.G., Gaedke, M. (eds.) ICWE 2005. LNCS,vol. 3579, pp. 417–428. Springer, Heidelberg.
- [13] Winckler, M., Palanque, P.: Statewebcharts: (2003). “ A Formal Description Technique Dedicated To Navigation Modelling of Web Applications”. In: Jorge, J.A., Jardim Nunes, N., Falc˜ ao e Cunha, J. (eds.) DSV-IS 2003. LNCS, vol. 2844, pp. 61–76. Springer, Heidelberg
- [14] Koch, N., Baumeister, H., Hennicker, R., Mandel, L(October 2000) “ Extending UML to Model Navigation and Presentation in Web Applications”. In: Workshop on Modelling Web Applications in UML, UML 2000, New York, UK
- [15] Han, M., Hofmeister, C. (2006) “Modeling and Verification of Adaptive Navigation in Web Applications”. In: ICWE 2006, Palo Alto, California, USA, July 11-14, pp. 329–336.