Web User Interest Prediction Framework Based on User Behavior for Dynamic Websites

T.R.Ramesh[1], C. Kavitha[2]

[1]Department of MCA, Sona College of Technology, Salem, Tamilnadu, India
[2]Department of Computer Science, Thiruvalluvar Government Arts College, Rasipuram, Tamilnadu, India
Email:ramesh_tr@rediffmail.com,kavithachellappan@yahoo.com

Abstract:We develop a framework to predict the user interest based on the behavior of user to increase the efficiency of dynamic websites. The content management in the dynamic website is difficult because it varies with the user profiles, i.e. different contents have to be placed for different users according to the user profiles. Various ways have been identified earlier to track the user interest but lacks with the accuracy here we propose a new one which composes both implicit and explicit. We track all behaviors like time of visit, navigation url, web logs, user actions on the web page. Our model uses the web log data of the user and also tracks the implicit behaviors performed by the user. The tracked information are used to identify the user interest and The web users are clusters based on the identified interest which is used by the dynamic websites. The dynamic websites administrator could use the outcome of the cluster for various purposes.
[T.R.Ramesh, C. Kavitha. **Web User Interest Prediction Framework Based on User Behavior for Dynamic Websites.** *Life Sci J* 2013;10(2):1736-1739]. (ISSN:1097-8135). http://www.lifesciencesite.com. 245

Keywords*:* Dynamic Websites, User Behavior, User Interest Prediction, Clustering

I INTRODUCTION

Providing interested information for the web users become more complicated due to the increased number of users. Now a days the people uses web for everything and the website management has the responsibility to provide valuable, useful and interested information to the web user.

For example the online marketing and shopping company sells their products through the web and they need to post the information about the launch of new products to their customers. The challenge here is identifying the interested user, so that they could be informed. Posting the new product launch to all the customers becomes meaningless, because if the user is not interested in the new product then the web resource usage is lost in sending a mail.

Also in the business point of view the marketing management can infer the most interested product and most moving product by using the web logs. Using the web logs of the online Shoppe the business people could infer various information and in various ways for the development. For example, using the visiting history of the customers, we can infer that which is the most interested product and using the time variant data we can infer that which product generates more profit in particular time frame.etc.

The user interest can be predicted using time variant data, the time variant data is one which is collected by the website management for future use either implicit or explicit manner. The interest of the customer may change in timely manner and identifying the user interest in each time frame is important one.

There exist various ways for interest prediction, the user interest can be predicted by the user actions, web log, time spent in a web page etc. User behaviors are classified in two ways as implicit and explicit. Implicit behaviors are time spent, number of visit to the page etc. , where as explicit behaviors are actions performed on the web page like save, bookmark, copy, print etc.

The implicit behavior of the user shows that the user may be interested on the product, because we cannot say that the user is definitely interested on the product only by viewing the web page or spending some time on the web page. But we can say the user has confirmed interest by performing some actions on the web page. We propose such a framework which tracks both implicit and explicit behaviors to predict the user interest for the development of the dynamic web sites.

II BACKGROUND

We explore the earlier methodologies proposed by various researches as follows.

H-UNC (2002) proposed a New Evolutionary Approach to Web Usage and Context Sensitive Associations Mining, which uses a set of Web sessions preprocessed from Web log data for several months. The data was segmented into sessions based on the client IP address and a time-out threshold between two consecutive accesses after removing the irrelevant data. After filtering irrelevant URLs like graphics and requests from Web crawlers, we get a number of unique sessions. The association technique is applied on the identified session details.

Amr Ahmed, Yucheng Low (2011) designed a user behavior inference framework to analyze changes in a user's activity patterns which are particularly useful

for improved prediction and recommendation. An increased interest in car-related web pages may well suggest that the user might be shopping for a new vehicle.

Ryon W.White (2009) proposed a user interest prediction framework which is based on the contextual information. They used five different contextual information sources, which are social, historic, task, collection, and user interaction.

Ingwersen, P. (1994). The element of a cognitive theory for information retrieval interaction is proposed which is a Poly representation of information needs and semantic entities. The historical information employed in user interest modeling is one source of contextual evidence about the current session. Others include time of day, user gender, age, ethnicity, locality, etc.

White, R.W. (2005) proposed a study of factors affecting the utility of implicit relevance feedback. He suggests that the overlap between numerous contexts associated with the current session can be used to locate pertinent items. The algorithm supports the development of rudimentary user interest models that are based solely on the interaction context. All these models lags with the execution time and efficiency , to improve the efficiency of these models we propose a new framework which identifies the user interest based on the user behavior for the development of the dynamic websites.

III PROPOSED SYSTEM

The user interest prediction framework has various modules to track the user behavior on the web pages. Our framework records the web pages visited by the user at each time frame and the actions whatever performed by the user is also recorded to the data base. The recorded log is used to cluster the user session and interest is identified on each session. Finally the user interest is predicted by projecting the user interest from the cluster details.

The proposed framework contains the following phase for user interest prediction namely:

A. Implicit Tracking

The web pages visited by the user at each time frame are recorded in the background without knowing the user. At each web page the user may spend some time that also recorded in the background in implicit manner. The recorded information is used to infer the interest of the user at later stage. For example the user may visit a particular web page and spend some 15 seconds, and in another web page he may spend 20 seconds which infer that the second web page visited is more interested one. These details are collected in each time frame like in monthly, 15 days, 7 days, daily manner. At each session the web page visited and time spent, date of visit, time of visit is recorded to the data base. We have designed a specific web browser to track the user actions.

B. Explicit Tracking:

Whenever the user visits the web page the user may perform some action like saving the web page, book marking the web page, copying some content of the web page or printing the content. These actions are called explicit actions, which are recorded to the data base with the web page visited, date and time of visit. These explicit details are used to identify the user interest on each day, week, and monthly manner.

Algorithm

Step1: Read the user input as url or query.
Step2: if input == url
Read the web page and show it to the user interface.
Initialize the timer.
Else
Submit the query to any web browser.
Receive the result.
Show it to the user interface.
Initialize the timer.
Step3: Receive the User Action.
Step4: Store to the database.
Step5: if user query or url changes
Calculate the time Spent.
Store time spent details to the database.
Re-initialize the timer.
Step6: End.

C. Web log Session Clustering

The recorded information from the earlier phases are retrieved from the data base and grouped into different categories. Initially at each session a single user details are retrieved and the unique web pages visited by the user is identified. The single web page may be visited by the user at many times in single session , the total time spent and number of time visit is calculated based on which average visiting time is calculated. At each time of visit the user may perform various actions, that is also calculated as total number of actions performed for each unique web page. Using the calculated values of average visiting time and total number of actions performed on each web page, the interest weight value is calculated to identify the interest at each session or time frame. Finally a single interest is projected from the available interest in each session.

Algorithm

Step1: for each session-Si read recorded user visit details(Vs).
Step2: read time spent details(Ts)
Step3: read actions performed details(As).
Step4: at each time frame or session(Ss)

- Identify unique web urls (Us).
- Calculate total number of visit made by user -

Tv.

- Calculate total time spent on the web page Ts.
- Calculate average time spent- avts.

Step5:

- Arrange web pages according to avts.
- Calculate action value using total actions performed by user(Av).
- Select top three (Sw) web pages which has more avts.
- Select top web pages from (Sw) based on total actions.

Step6: return list.

Average Time spent avts =∑Ts/ ∑Av

∑Av- Sum of Actions performed on the web page.

∑Ts- Sum of Time Spent on the web page.

Sw- sorted web pages.

D. User interest prediction

In this phase the interest at each session is projected using the calculated values of earlier phase. The persistent interest is identified from the cluster result, by identifying the repeated interest. The interest which is persistently occurs in each session cluster is the most interested item and is identified and projected based on the total actions performed.

Algorithm

Step1: read clustered information

Step2: read calculated total actions details

Step3: read top web pages selected at each session.

Step4: Identify repeated interest at each session.

Step5: select the interest which has more action value(Av).

Step6: project the interest.

Step7: end.

Interest In= (∑Si)μ

∑Si- shows the set of interest identified in each session.

μ- Projection operation which projects a single interest which is common in all the session.

IV RESULTS AND DISCUSSION

The methodology we proposed has many advantages, and we used both web log data and user behavior details. We tracked three months of details to evaluate our framework and we tracked all the action performed by the user and visit details. The proposed methodology identify the user interest exactly and efficient manner. We used time variant data so that we can infer other information from the data set.

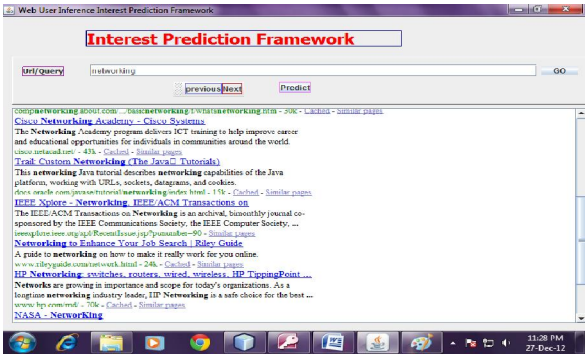


Figure 1: Shows the Interface Designed

The Figure1 shows the user interface designed for the purpose of algorithm and it shows the result of a search engine for the query submitted by the user through our framework the user can submit both the query or the url which the user wants to visit.

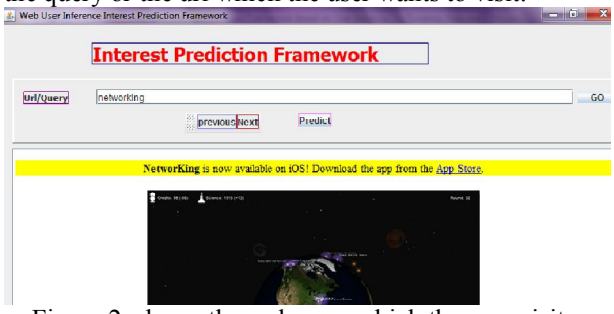


Figure 2: shows the web page which the user visits

Table1. User Visit Details

VisitDetails

Query	Url	Mdata
data mining	http://www.thearling.com/text/dmwhite/dmwhite.htm	<a href="http:/
data mining	http://www.thearling.com/text/dmwhite/dmwhite.htm	<a href="http:/
data mining	http://www.thearling.com/text/dmwhite/dmwhite.htm	<a href="http:/
data mining	http://www.autonlab.org/tutorials/	<a href="http:/
data mining	http://www.webopedia.com/TERM/D/data_mining.html	<a href="http:/
data mining	http://www.webopedia.com/TERM/D/data_mining.html	<a href="http:/
data mining	http://www.webopedia.com/TERM/D/data_mining.html	<a href="http:/
data mining	http://www.webopedia.com/TERM/D/data_mining.html	<a href="http:/
networking	http://www.nasa.gov/multimedia/3d_resources/scan.html	<a href="http:/

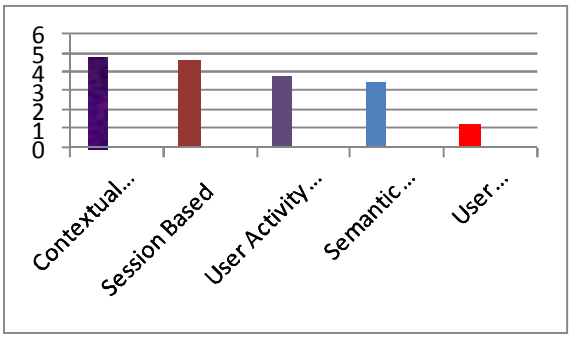
The Table 1 shows the recorded details of the user like the query submitted, web page visited, Meta data of the page.

Table 2 User Visit Details and Action Details

VisitDetails

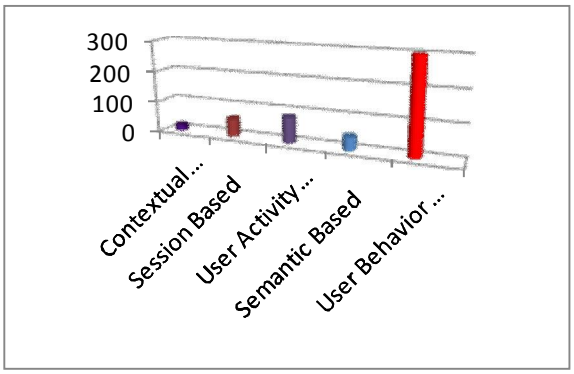
Mdata	Asave	ACopy	ABookmark	TSpent	Dvisit
<a href="http:/	1	0	0	52	22/12/2012
<a href="http:/	0	1	0	29	22/12/2012
<a href="http:/	0	0	1	77	23/12/2012
<a href="http:/	1	0	0	25	24/12/2012
<a href="http:/	0	1	0	95	25/12/2012
<a href="http:/	1	0	0	48	26/12/2012
<a href="http:/	0	1	0	33	27/12/2012
<a href="http:/	0	0	1	65	27/12/2012
<a href="http:/	1	0	0	53	27/12/2012

The Table 2 Shows the details of action performed by the user in particular web page and the time spent.



Graph 1: shows the false interest prediction

The Graph 1 shows that the rate of false interest prediction of all the algorithms we compared. It shows that our proposed system has very less rate of false interest prediction.



Graph 2: shows the number volume of data used for prediction

The Graph 2 shows the volume of time variant data used for prediction. It shows that the proposed system used more amount of time variant data used. We used nearly 300 days of data used for single user interest prediction.

5. CONCLUSION

We have tracked 2 million records to process and evaluate our algorithm. Our algorithms are very efficient in interest prediction and inference theory, our methodology produces very good results in very little time. The proposed system has very little time complexity and could be adopted to infer any kind of business data.

VI. REFERENCES

[1] R. Cooley, B. Mobasher, and J. Srivastava, "Web Mining: Information and Pattern Discovery on the World Wide Web," Proc. Ninth IEEE Int'l Conf. Tools with AI (ICTAI '97), pp. 558-567, 1997.

[2] Amr Ahmed, Yucheng Low " Scalable Distributed Inference of Dynamic User Interests for Behavioral Targeting " , School of Computer Science, CMU Carnegie Mellon University, ACM-(2011).

[3] Ryen W. White ." Predicting User Interests from Contextual Information" Microsoft Research, ACM-2009.

[4] Ingwersen, P. (1994). Polyrepresentation of information needs and semantic entities: elements of a cognitive theory for information retrieval interaction. *Proc. SIGIR*, 101-110.

[5] O. Nasraoui and R. Krishnapuram, "A New Evolutionary Approach to Web Usage and Context Sensitive Associations Mining," Int'l J. Computational Intelligence and Applications, special issue on Internet intelligent systems, vol. 2, no. 3, pp. 339-348, Sept. 2002.

[6] O. Nasraoui, C. Cardona, C. Rojas, and F. Gonzalez, "Mining Evolving User Profiles in Noisy Web Clickstream Data with a Scalable Immune System Clustering Algorithm," Proc. Workshop Web Mining as a Premise to Effective and Intelligent Web Applications (WebKDD '03), pp. 71-81, Aug. 2003.

[7] P. Desikan and J. Srivastava, "Mining Temporally Evolving Graphs," Proc. Workshop Web Mining and Web Usage Analysis (WebKDD' 04), 2004.

[8] O. Nasraoui, C. Rojas, and C. Cardona, "A Framework for Mining Evolving Trends in Web Data Streams Using Dynamic Learning and Retrospective Validation," Computer Networks, special issue on Web dynamics, vol. 50, no. 14, Oct. 2006.

[9] M.A. Maloof and R.S. Michalski, "Learning Evolving Concepts Using Partial Memory Approach," Working Notes AAAI Fall Symp. Active Learning 1995, pp. 70-73, 1995.

[10] M.A. Maloof and R.S. Michalski, "Selecting Examples for Partial Memory Learning," Machine Learning, vol. 41, no. 11, pp. 27-52, 2000.

[11] D. Agarwal and S. Merugu. Predictive discrete latent factor models for large scale dyadic data. KDD, 2007.

[12] A. Ahmed and E. P. Xing. Dynamic non-parametric mixture models and the recurrent chinese restaurant process In SDM, pages 219~230. SIAM, 2008.

[13] A. Ahmed and E. P. Xing. Timeline: A dynamic hierarchical dirichlet process model for recovering birth / death and evolution of topics in text stream. In UAI, 2010.

[14] A. Asuncion, P. Smyth, and M. Welling. Asynchronous distributed learning of topic models. In NIPS, pages 81~88. MIT Press, 2008.

[15] D. Blackwell and J. MacQueen. Ferguson distributions via polya urn schemes. The Annals of Statistics, 1973.

[16] D. Blei, A. Ng, and M. Jordan. Latent Dirichlet allocation. JMLR, 3:993~1022, 2003.

15/5/2013