Analyzing the user Action for Web Content Augmentation

Dheeraj Kumar N, Joyce Vinnaṛasi A
Sheethal Chinnappa C,Tejaswini S
Dept of Computer Science & Engineering
T John Institute Of Technology Bengaluru, India

Dr. Mahesh T R
Professor & Head Of Dept CSE
T John Institute Of Technology
Bengaluru, India

Abstract - **Web sites have become a major source for rendering any information desired by the user, be it the current affairs or social media like Facebook, Twitter etc...These websites, in addition to the information desired by the user, also broadcast a wide range of commercial promotions. To captivate the users towards these commercials, it is necessary to have a suggestive mechanism to propose appropriate commercials to the users based on their interests. The users' concern for the commercials is estimated based on the number of clicks on that item and the users' assessment for the same. This forms the ground for the suggestive mechanism. For this mechanism to perform effectively, exclusivity of the users and their fragmentation is emphasized. Also a behavior - driven model is built based on historical user activity by analyzing users' actions so as to build an efficient web content augmenting system.**

Keywords- Suggestive mechanism; Exclusivity; Fragmentation; Augmenting

INTRODUCTION

In modern day, internet plays a vital role in human life. The commercial promotion which are posted on web portal makes the users aware of the product that are available in the market. The organization of the commercial products are benefited as their products are promoted on the web portals. These advertisements captivate the users towards the products and provides the users with a wide range of choices. However it is not possible to post all the advertisements on the webpage. It is necessary to post the appropriate advertisements to the users such that it would grab their attentions. This can be done

by determining the users' interaction with the webpage. The users behaviour can be determined in two ways, implicit and explicit manner. In implicit fashion the users' clicks forms the basis to identify their interest. On the other hand in explicit fashion as the name suggests the interest will be identified by explicitly rating the advertisement. Implementing the same ensures the commercial organization that their adverstiments are posted to the appropriate users.

Often, human copyholder are employed to manually select a set of contents to present from the applicants pool. Although this technique can eliminate low- quality contents and make certain necessities that characterize the websites, which is quite expensive and usually cannot assure that the most attractive and personally accurate contents are suggested to the users especially when there are many number of applicant items. As a result a strong and automated content augmentation becomes crucial for serving users with attractive data in a flexible manner. Exclusivity of the user is an enticing feature for content augmentation to know an individual's interests.

Suggestive mechanism involves the process of collecting and storing data about the candidates' website. The exclusivity of the users are divided into two major aspects: Data Distillation and Associative Distillation.

DATA DISTILLATION

In this method a profile is created for a user based on the data description that are previously being rated by the candidates. The main disadvantage of this technique is its limited facilities to suggest data items
that differ from the previously rated ones.

Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICESMART-2015 Conference Proceedings

ASSOCIATIVE DISTILLATION

This is one of the most broadly used techniques, which determines users previous rating to recognize the similarities and suggest items by giving more prominence to the interest.

Shortcoming:There is a big traffic of users' visiting in every minute, websites portal can attract a large number of user's actions in terms of browsing and clicks on presented data modules. Such user action information can obviously provide strong signals of users' recent interests on the content item. However, it is quite a challenge to incorporate them into the recommendation model in real time or near real time

The purpose of exclusivity is to provide users with personalized experience of highly attractive content, the problem of how to appropriately define user segments to achieve exclusitivity becomes crucial for effective content optimization.

Most of existing studies focus on building the offline suggestion model; those previous approaches may not be good enough data suggestion, and users' interests.

ONLINE STUDY FOR EXCLUSIVE SUGGESTION

In exclusive suggestive mechanism effort is made to determine the users' interest which would be beneficial for posting advertisements on the web portal. However there are challenges in building such a mechanism.

PROBLEM FORMULATION:

To build an effective suggestive model, it is necessary to emphasize certain user engagement metrics like the click through rate. However if this is managed manually it would be very expensive as well as inefficient, since it cannot guarantee that the most attractive and relevant items are suggested to the users. Therefore an attempt is made to design a suggestive mechanism that automatically estimates the content items attractiveness in accordance to the users interest. There are three critical elements for this suggestive model.

1.Online Study: To attract more users to surf and click on items that are posted on the web portal an online study is necessary to model the users behavior

such as clicks and views which implicitly ranks the content item that are posted.

To enable the online study for content augmentation the dual bucket approach is used. In this there are two buckets, Reading bucket and Service bucket. In reading bucket when the user visit a web page this visit can be assigned to reading bucket. Within the reading bucket random items are present, from which advertisement are chosen and are given to the service bucket. The Service bucket as the name suggests, it receives the contents from the reading bucket based on the click through rate and serves it on the web portal.
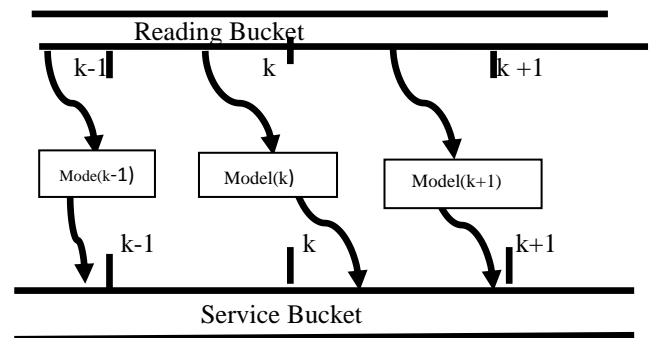
2.Ranking model: To build effective model the best method is to use a dedicated model for each content items to estimate its attractiveness . Using these ranking models we can rank all item by their respective score. And then present the top ranked to the user. The score gives a strong signal of the users interest which is obtained by the click through rate. The user interaction plays a vital role in determining his or her interest. The contents that are viewed by the user would be ranked.

According to the ranking system the contents are shown to the user. The ranking system would vary depending upon the views that the user made. The Commercial organization would make use of the Ranking table in knowing which is the advertisement most liked by the user and also this would give a clear picture for the organization which advertisement are being not liked by the user and hence necessary step would be taken by the organisation to improve the product.T

Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICESMART-2015 Conference Proceedings

TABLE 2
Ranking Table

Actual ranking	Predicted ranking by Model 1	Predicted ranking by Model2
1(clicked)	1	2
2	5	3
3	4	1
4	3	4
5	2	6
6	6	5

3. Exclusivity:

Exclusivity has become very important for content augmentation as its provide users with the appropriate advertisement. Having a separate model to each users is quite a tedious task. Hence to reduce the work load the users with similar interest can be grouped which is termed as user fragmentation.

In user fragmentation, the users with similar interests are grouped together forming a homogenous fragment, where each fragment of users are served with the dedicated suggestive model. This approach is both simple and reliable. Each user group has its own ranking model which are read based on clicks and views only from the users belonging to the respective group and the service of posting the appropriate content to the corresponding group is also based on this model.

To obtain user fragmentation, a set of user attributes is generalized and then clustering techniques are used to categorize the user based on the elicited attributes. There are two major user attributes:

Explicit Attributes: The exclusive information explicitly requested by the website such as age, gender, location etc.

Implicit Attributes: These are various user behaviors tracked by the website such as surfing and online shopping patterns of users on the pages within the websites.

CHALLENGES

1. It is a challenge to group the users appropriately as in the homogenous users i,e users with similar interest should belong to the same group while heterogeneous users should belong to different groups. This becomes difficult when the users have

diverse interest as the high dimensionality of user attribute will lead to ambiguity.

2. The information provided for grouping the users in each user fragment is not sufficient for the correct understanding of users' behavior and this degrades the suggestive mechanism.

ANNOTATION OF USER ACTION

In order to overcome the above stated drawbacks a good understanding and analysis of users' actions including clicks and views, ca effectively benefit the suggestions by better understanding users' general interests and the contents' attractiveness to users.

ACTION ANNOTATION FOR USER FRAGMENTATION

To appropriately group users into different fragments, the most straight forward method is to group users based on their explicit static attributes. However, this method may not be optimal since it ignores large amount of implicit user behavior information. Fragmentation by demographic information Intuitively, users' with same democratic attributes such as age and gender are likely to have similar interest. Accordingly, the best approach for user fragmentation is to group users based on combinations of several demographic attributes which are provided as profiles to the websites.

TABLE 1
User fragmentation based on demographic attributes

Fragment	Age Range & Gender
f-u20	10<age<=20,gender=female
f-u40	20<age<=40,gender=female
f-u80	40<age<=80,gender=female
m-u20	10<age<=20,gender=male
m-u40	20<age<=40,gender=male
m-u80	40<age<=80,gender=male
unk	Unknown age or gender

GROUPING USERS

When users browse on the web, there is plenty of information about users behavior on the content displayed to them. However it is quite a challenge to keep track of each and every individual behavior. Since every minute huge number of users visit and maintaining such huge amount of information is tedious task. Therefore it becomes important to group

Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICESMART-2015 Conference Proceedings

the users based on their interest. The users who have similar interest are grouped and hence reducing the effort drastically.

EFFECTS OF USER FRAGMENTATION

The most popular advertisements are determined based on the maximum number of clicks made by the users. However to such users who don't show their interest implicitly by clicking on the advertisement, the rating plays a major role. The users grouping enables less traffic since the details of every users need not be tracked individually instead the group details will be preserved.

CONCLUSION

In this paper we have studied a few important topics towards exploring analyzing the user action for web content augmentation. We build an exclusive content augmentation system using dual bucket framework. In this framework, we introduce action annotation for both more effective user fragmentation and better understanding on the in formativeness of different user actions. We leverage users' click actions to group homogenous users' into the same fragment.

In the future, we are interested in exploring more information about exclusivity of the users such as users' geographic location and click behaviors from web search and to take advantage of them to benefit web content augmentation.

REFERENCES

[1] D. AGARWAL, B.-C. CHEN, AND P. ELANGO, "SPATIO-TEMPORAL MODELS FOR ESTIMATING CLICK-THROUGH RATE," PROC. 18TH INT'L WORLD WIDE WEB CONF. (WWW), 2009.

[2] D. AGARWAL, B.-C. CHEN, AND P. ELANGO, "FAST ONLINE LEARNING THROUGH OFFLINE INITIALIZATION FOR TIME-SENSITIVE RECOMMENDA- TION," PROC. 16TH ACM SIGKDD INT'L CONF. KNOWLEDGE DISCOVERY AND DATA MINING (KDD), 2010.

[3] D. AGARWAL, B.-C. CHEN, P. ELANGO, N. MOTGI, S.-T. PARK, R. RAMAKRISHNAN, S. ROY, AND J. ZACHARIAH, "ONLINE MODELS FOR CONTENT OPTIMIZATION," PROC. NEURAL INFORMATION PROCESSING SYSTEMS FOUNDATION CONF. (NIPS), 2008.

[4] E. AGICHTEIN, E. BRILL, AND S. DUMAIS, "IMPROVING WEB SEARCH RANKING BY INCORPORATING USER BEHAVIOR INFORMATION," PROC. 29TH ANN. INT'L ACM SIGIR CONF. RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL (SIGIR), 2006.

[5] J. AHN, P. BRUSILOVSKY, J. GRADY, D. HE, AND S.Y. SYN, "OPEN USER PROFILES FOR ADAPTIVE NEWS SYSTEMS: HELP OR HARM?," PROC. 16TH INT'L WORLD WIDE WEB CONF. (WWW), 2007.

[6] D. BILLSUS AND M. PAZZANI, "ADAPTIVE NEWS ACCESS," THE ADAPTIVE WEB - METHODS AND STRATEGIES OF WEB PERSONALIZATION, SPRINGER- VERLAG, 2007.

[7] A. BRODER, "COMPUTATIONAL ADVERTISING AND RECOMMENDER SYSTEMS," PROC. SECOND ACM INT'L CONF. RECOMMENDER SYSTEMS (RECSYS), 2008.

[8] G. BUSCHER, L. VAN ELST, AND A. DENGEL, "SEGMENTATION-LEVEL DISPLAY TIME AS IMPLICIT FEEDBACK: A COMPARISON TO EYE TRACKING," PROC. 32ND ANN. INT'L ACM SIGIR CONF. RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL (SIGIR), 2009.

[9] W. CHU AND Z. GHAHRAMANI, "PROBABILISTIC MODELS FOR INCOMPLETE MULTI-DIMENSIONAL ARRAYS," PROC. 12TH INT'L CONF. ARTIFICIAL INTELLI- GENCE AND STATISTICS, 2009.

[10] W. CHU AND S.T. PARK, "PERSONALIZED RECOMMENDATION ON DYNAMIC CONTENT USING PREDICTIVE BILINEAR MODELS," PROC. 18TH INT'L WORLD WIDE WEB CONF. (WWW), 2009.