LOCATION BASED CBCF METHOD FOR EFFECTIVE PERSONALIZATION OF ONLINE CONTENT

Steffina.T[1] Sivananaitha Perumal.S[2]

[1]PG Scholar, [2]Professor, Department of IT

Dr.Sivanthi Aditanar College of Engineering, Tiruchendur, (India)

ABSTRACT

Many on-line stores provide recommending services. There are two prevalent approaches for building recommender systems — Collaborative Filtering (CF) and Content-based (CB) recommending. CF systems work by collecting user feedback in the form of ratings for items in a given domain and exploit similarities and differences among profiles of several users in determining how to recommend an item. On the other hand, content-based methods provide recommendations by comparing representations of content contained in an item to representations of content that interests the user. User interaction plays a vital role in building effective content optimization, as both implicit user feedbacks and explicit user ratings on the recommended items form the basis for designing and learning recommendation models. In particular, we propose an approach to leverage historical user activity to build behavior-driven user segmentation; then, We introduce an approach for interpreting users' actions from the factors of both user engagement and position bias to achieve unbiased estimation of content attractiveness. Our experiments on the large-scale data from a commercial Web recommender system demonstrate that recommendation models with user action interpretation can reach significant improvement in terms of online content optimization over the baseline method.

Keywords*: Action Interpretation, Content Optimization, Personalization, Recommender Systems, Behaviour-Driven User-Segmentation***

I. INTRODUCTION

There is a rapid growth of the Internet, which has become an important medium to deliver digital content to Web users instantaneously. Digital content publishers, including portal websites, such as MSN (http:// msn.com/) and Yahoo! (http://yahoo.com/), and homepages of news media, like CNN (http://cnn.com/) and the New York Times (http://nytimes.com/), have all started providing Web users with a wide range of modules of Web content in a timely fashion. Therefore, it is necessary for those Web publishers to optimize their delivered content by identifying the most attractive content to catch users' attention and retain them to their portal sites on an ongoing basis. Often, human editors are employed to manually select a set of content items to present from a candidate pool. Although editorial selection can prune low-quality content items and ensure certain constraints that characterize the portal website, such human effort is quite expensive and usually cannot guarantee that the most attractive and personally relevant content items are recommended to users especially when there is a large pool of candidate items. As a result, an effective and automatic content optimization becomes indispensable for serving users with attractive content in a scalable manner. Personalization is also a desirable feature for the content optimization since it can further tailor content presentation to suit an individual's interests rather than take the traditional "one-sizefits- all" approach.

In general, personalized content recommendation on portal websites involves a process of gathering and storing information about portal website users, managing the content assets, analyzing current and past user interactive actions, and, based on the analysis, delivering the right content to each user. Traditional personalized recommendation approaches can be divided into two major categories: content-based filtering and collaborative filtering. In the former method, a profile is generated for a user based on content descriptions of the content items previously rated by the user. However, the main drawback of this approach is its limited capability to recommend content items that are different than those previously rated by users. Collaborative filtering, which is one of the most successful and widely used techniques, analyzes users' ratings to recognize commonalities and recommend items by leveraging the preferences from other users with similar tastes.

To summary, the main contributions in this paper include

. an effective online learning framework for taking advantage of user actions to serve content recommendation in real time or near real time;

. a new approach to leverage historical user activity to build a behavior-driven user segmentation, which results in higher engagement after application to the personalized content optimization; and

. a novel approach of interpreting users' actions for the online learning to achieve better estimation on content items' attractiveness, including taking into account the factors of both user engagement and position bias.

II. RELATED WORK

Content optimization is defined as the problem of selecting content items to present to a user who is intent on browsing for information. There are many variants of the problem, depending on the application and the different settings where the solution is used, such as articles published on portal websites [3], [2], news personalization [12], recommendation of dynamically changing items (updates, tweets, etc.), computational advertising [7], and many others. Chu et al. [10] and [11] recently proposed user behavior feature-based models for personalized services at individual and segmentation levels, respectively. Those personalized models are shown to outperform several demographic segmentation models. However, they did not analyze the quality of each of more than 1,000 user behavior features. In our work, we take advantage of user click information to select a subset of user behavior features with high quality. Some work discussed user behavior models based on controlled user studies [21], while other studies focused on large-scale log analysis [14]. YourNews [5] allows users to customize their interest profiles through a user model interface. These studies on user behaviors show the benefit from customization, but also warn of the downside impact on system performance. In our application, we take advantage of user behavior information without explicitly soliciting it from users.Das et al. [12] and Liu et al. have made earlier effort to enhance news recommendation based on users' click behaviors. Beyond them, my work will propose a more comprehensive study on the effects of users' behaviors for online content optimization, and our study will be expanded into any content module.

III. PROPOSED FRAMEWORK

3.1 Online Learning

To enable online learning for content optimization, we introduce a parallel-serving-buckets approach. Fig.1 illustrates the flowchart of this online learning approach for the system. The term bucket is used to denote a part of the whole users visit traffic on portal websites. Different buckets yield different strategies to serve recommendation. Specifically, in our parallel-serving-buckets framework, we divide the whole users visit traffics into two parallel buckets serving simultaneously in the system: random learning bucket and serving bucket. When a user visits the portal website, this visit event can be randomly assigned into either the random learning bucket or the serving bucket.

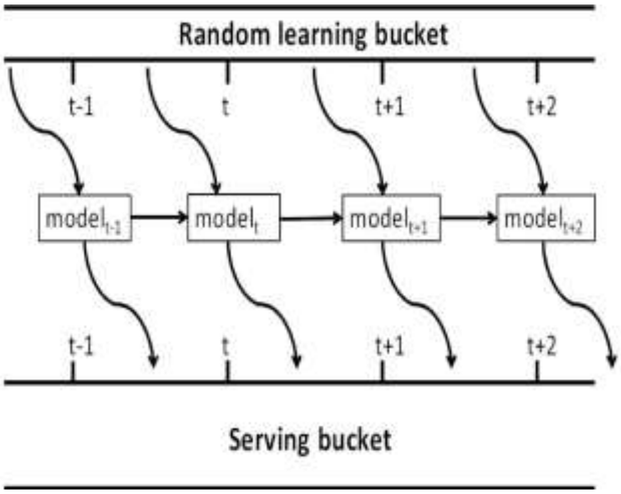

Fig 1: Online learning Framework

Within the random learning bucket, a certain number of items are randomly sampled from the pool of candidates to serve as recommended items for each user visit. In our parallel-serving-buckets approach, as shown in Fig. 1, all the models in both buckets are updated simultaneously every 5 minutes (i.e., the time interval [t ,t + 1] equals 5 minutes in Fig. 2). In general, within the serving bucket, each per-item model, at a certain time point t + 1, is adapted to the observations (users views and clicks) the corresponding item from the random learning bucket during the time interval [t ,t + 1]. The updated models are then applied to the candidate items in serving bucket and the items are displayed by the ranking scores in descending order.

3.2 Per-Item Model

To build effective online recommendation model, the straightforward but reliable method is to apply a dedicated model for each candidate content item to estimate its attractiveness/relevance score. Using these dedicated per-item models, we can rank all items by their respective recommendation scores in the descending order and present the top ranked ones to users. To adapt per-item model in our online learning framework, it is essential to employ an effective method for updating per-item models. In this paper, we employ an estimated most popular (EMP) model.

3.3 User Segmentation

To introduce personalization for online content optimization, we propose employing user segmentation-based approach, in which homogeneous groups of users are entailed by a priori segmentation, where each segment of users are served with the dedicated recommendation model. There are a few other categories of personalization approaches; however, the user segmentation approach yields advantages in terms of both simplicity and reliability, especially for real-world commercial recommender systems. To integrate user segmentation into the online learning approach , users are divided into a small number of groups, each of which has its exclusive online learning and serving process. To obtain this user segmentation, we propose generalizing a set of user features and then applying clustering techniques to group users based on extracted features. We collect two major categories of user features that is available to the portal website owner: 1) Explicit features: the personal information explicitly requested by the portal website, such as age, gender, location, preferences, and so on. 2) Implicit features: various types of users' behaviors tracked by portal website, such as browsing and purchasing patterns of users on the pages within this website, and so on. Both of these two categories of user features can implicitly represent users' preferences and recent interests over Web content. In our recommender system, each user is represented as a vector of features, whose dimensionality can be more than 1,000 in our experiments.

IV. USER ACTION INTERPRETATION

We propose to use historical user click information to select discriminant features for user segmentation. Then, we study how to improve online learning based on more accurate interpretation of users' actions in terms of clicks and views with considering both user engagement and click position bias.

4.1 Action Interpretation for User Segmentation:

To appropriately divide users into different segments, the most straightforward method is to group uses based on their explicit static features, such as demographic information. However, this heuristic rule-based method may not be optimal since the generated segmentation is ad hoc and it ignores large amounts of implicit user behavior information which can better reflect users' interests. We propose to take advantage of the rich user behavior information, especially the history of users' clicks on the portal website, to obtain a user segmentation that results in a better serving content optimization. In particular, we will introduce two different clustering techniques for leveraging such important information.

Segmentation by Demographic Information

Unsupervised Clustering

4.2 Action Interpretation for Online Learning:

The online learning algorithm relies heavily on user clicks and views, which are critical for developing effective content optimization. For a candidate item, its CTR is estimated based on the number of clicks and views for this item (1), which implies that correct interpretation of user actions is important since click/view samples are derived from the user actions logged by the portal website. Along this direction, we address two important factors in this section, including user engagement and position bias.

We identify three categories of events regarding user engagement:

. **Click event**. Click event is an event where the user clicked one or more items in the module after she opened the web page. Note that one click event consists of the user's click on one item and her views on other items along with it. Obviously, click events is useful for CTR estimation.

. Click-other event. Click-other event contains at least one action on other application/modules in the interface (such as clicking items displayed by other modules, doing search in search box, etc.). Obviously, click-other events should be excluded from being used for CTR estimation.

. Nonclick event. Besides click events and clicks-other events, there are also nonclick events in which users had no action such as click or search after they opened the web page. For a nonclick event, unlike click event or click-other event, it is not straightforward to determine whether or not the user actually examined the module under study as usually the system cannot track user's eyes. However, based on user's historic behaviors, it is still possible to deduce if the user intends to examine the module or not.

There are at least two factors that may lead to such position bias:

1) an item displayed at different positions may have different probabilities of being examined by users; and

2) if a user examines an item at bottom positions, the probability that she clicks this item is lower than the case that this item is displayed at top positions. This is because when the item is displayed at bottom positions, users may have less confidence that this item is high quality. We call this phenomenon position decay factor.

V. LOCATION BASED SEARCH

Location Detection is a system designed to capture, store, manipulate, analyze, manage, and present all types of spatial or geographical data. Location-based searching is one of the popular tasks . A location-based query consists of a topic and a reference location. Unlike general web search, in location-based search, it is expected to find and rank documents which are not only related to the query topic but also geographically related to the location which the query is associated with.

VI. EXPERIMENTS

We design experiments to validate our proposed approaches. The experimental results showed that the success of user segmentation for personalization is due to the fact that the proposed clustering algorithms actually group users by interests and preferences that are implicitly demonstrated by their behaviors. Once the interest patterns are determined by clustering algorithms, a user will be assigned to a segment by her profile features. Although K-means algorithm and tensor segmentation algorithm yield similar precision performances, the K-means algorithm is much more preferred due to its efficiency. User engagement is another important factor. In our user segmentation model, precise action interpretation is critical for online learning as the samples in each segment are relatively sparse. In the real product, we have tested the user segmentation model using just click events. Similar to the offline results ,the online CTR result is also significantly improved over the nonsegmentation model using all events. Position bias is a sophisticated factor. Using fixed position weights is not very effective to further improve CTR estimation. For a click event, a strong signal is that the clicked item should be more attractive to the user than the rest of nonclicked items. However, for our per-item model approach, the CTR of each item is estimated independently by a Poisson process assumption so that such comparison information in click events are lost. While per-item model is easy for product implementation, we need to further study new models which can utilize such competing preference information.

VII. CONCLUSION

In this paper, we have studied a few important topics towards exploring user action interpretation for online personalized content optimization. We build a online personalized content optimization system using the parallel-serving-buckets framework. In this framework, we introduce action interpretation for both more effective user segmentation and better understanding on the informativeness of different user actions. In particular, we leverage users' click actions to group homogeneous users into the same segment; then, we explore the effects of a couple types of user engagement factors as well as the position bias on the online learning procedure. Large scale evaluations on both offline data set and online traffic of a commercial portal website demonstrate that we can significantly improve the performance of content optimization by integrating all of these user action interpretation factors into the learning process. To explore personalization, we use users' geographic location and studying how to taking advantage of it to benefit content optimization.

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 Recommendation," 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 Intelligence 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.

[11] W. Chu, S.T. Park, T. Beaupre, N. Motgi, and A. Phadke, "A Case Study of Behavior-Driven Conjoint Analysis on Yahoo! Front Page Today Module," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD), 2009.

[12] A. Das, M. Datar, A. Garg, and S. Rajaram, "Google News Personalization: Scalable Online Collaborative Filtering," Proc. 16th Int'l World Wide Web Conf. (WWW), 2007.

[13] L.M. de Campos, J.M. Ferndez-Luna, J.F. Huete, and M.A. Rueda- Morales, "Combining Content-Based and Collaborative Recommendations: A Hybrid Approach Based on Bayesian Networks," Int'l J. Approximate Reasoning, vol. 51, no. 7, pp. 785-799, 2010.

[14] D. Downey, S. Dumais, D. Liebling, and E. Horvitz, "Understanding the Relationship between Searchers' Queries and Information Goals," Proc. 17th ACM Conf. Information and Knowledge Management (CIKM), 2008.

[15] E. Gabrilovich, S. Dumais, and E. Horvitz, "Newsjunkie: Providing Personalized Newsfeeds via Analysis of Information Novelty," Proc. 13th Int'l World Wide Web Conf. (WWW), 2004.

[16] L.A. Granka, T. Joachims, and G. Gay, "Eye-Tracking Analysis of User Behavior in WWW Search," Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2004.

[17] Z. Guan and E. Cutrell, "An Eye Tracking Study of the Effect of Target Rank on Web Search," Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI), 2007.

[18] J.L. Herlocker, J.A. Konstan, A. Borchers, and J. Riedl, "An Algorithmic Framework for Performing Collaborative Filtering," Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 1999.

[19] T. Hofmann and J. Puzicha, "Latent Class Models for Collaborative Filtering," Proc. 16th Int'l Joint Conf. Artificial Intelligence (IJCAI), 1999.

[20] R. Jin, J.Y. Chai, and L. Si, "An Automatic Weighting Scheme for Collaborative Filtering," Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2004.

[21] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay, "Accurately Interpreting Clickthrough Data as Implicit Feedback," Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2005.

ABOUT AUTHOR

	Steffina.T received the B.Tech degree in Information Technology from Anna University, Tiunelveli, in 2011. Currently, she is pursuing M.Tech degree in Information Technology from Anna University, Chennai. Her research areas of Interests include Data Mining, Operating system, and Networking.
	Sivananaitha Perumal.S received the B.E degree in Electrical and Electronics Engineering from Anna University, Chennai, in 1996 and he received M.E in Power Systems from Anna University, Chennai, in Dec 1997. He pursued Doctrate in Electrical from Anna University, Chennai, in 2012. His research areas of Interests include Image Processing and Soft computing.