International Journal of Computer Science Trends and Technology (IJCST) – Volume 2 Issue 1, Jan-Feb2014

RESEARCH ARTICLE OPEN ACCESS

Online Content Explosion for Content Optimization Using Prioritizing Queue with Updated Frequent Pattern Tree

A.Gokulakannan[1] B.Santhoshkumar[2] S.Sridhar[3]

PG Scholar, Department of Computer Science, Sri Krishna College of Engineering and Technology, Coimbatore
Anna University, TamilNadu - India

ABSTRACT

The digital content and modules such as news, sports, entertainment etc. are delivered by the web portal in the internet. Web portal must have the user interested and attracted modules. So it's required to build the recommender system that can achieve online content optimization by user interest as implicit and explicit user rating. These consider the particular period historical activities such as user search and feedback. Online learning framework analysed for personalized content optimization to fully leverage historical activities for that used the behaviour driven user segmentation. This keeps lot of historical activities of the portal web site. To address this issue and improve the accuracy prioritizing queuing model with updated frequent pattern tree is introduced to improve better result.

Keywords- Recommender System, user action, click through rating (CTR) estimation, prioritizing queue, updated frequent pattern tree.

I. INTRODUCTION

Now a day's internet is growing rapidly. Everything present in the form of digital content. All the digital content are presented under some categories for example various type of the news are presented below the news module this included in the portal website like yahoo, msn fig1 represent msn portal website front page has news module which consist various type of news ,trading module trading from the search engine that is used in portal website , since they are multiple content venders and huge of content Web user normally has the short attention on the portal while searching the contents and module from the plenty of module. So using recommender system has to pick user interested and attractive module for portal website.

To address this recommendation first manually editors select the items from the candidate item set. In the sense recommender system can prune low quality of content.

To face this challenge personalized content optimization recommender system applied on portal website. Personalized content recommendation involves the process of gathering and storing information about portal website users. To maintain and analysing the present and past user actions [4] Information filtering is a technology applied that responsible for plenty of information. Based on a profile of user interests and preferences, systems recommend items that may be of interest or value to the user. Information filtering majorly categorized into content based filtering and collaborative filtering.

For filtering contents are gathered from the parallel serving bucket which consist of user clicks and views is known online learning.

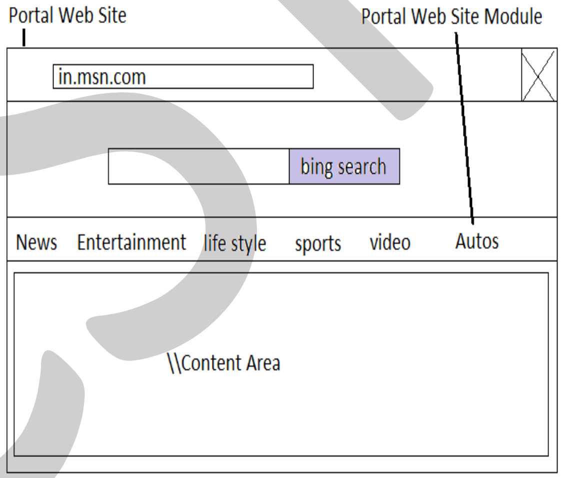

Fig 1: MSN Portal Website

Then each–item model estimate the CTR (Click through rating) from parallel serving bucket. For CTR estimation parallel serving bucket leverage the all the historical data which contain user action and that including user engagement for better estimation. This paper introduce the EMP-prioritizing queuing model contain historical data CTR Estimation. This estimation is based on the EMP- frequent pattern tree.

II. RELATED WORK

About the portal web site recommendation articles published [2], [3], highlight the personalization in the portal website module articles published news personalization of recommendation [12], [13].

Recommender system [1] for Content optimization is two major categories of approaches for content based filtering and collaborative filtering.[recommender] represent these two filtering.

A. Content Based Filtering

Content based filtering based on user profile that is created at the beginning. Profile contains the user taste and interest. Positively rated by the user contents are taken into the content based filtering.

B. Collaborative filtering

Collaborative filtering considers the user browsing details and behaviours. Contain three approach item based, user based hybrid based approaches.

1) *Item Based Approach:* User taste is constant apart from that slightly focuses the item that is mostly referred neighbourhood users. Recommendation preferred by item.

2) *User Based Approach:* User performs the main role in the user based approach. Filtering performed based on number of user behaviours and browsing details.

3) *Hybrid approach:* For the better performance and accuracy content based and collaborative based filtering method is combined and that gave better result. The combination of approaches can proceed in different ways [3].

Build the filtering model and join the model. Utilize some rules of content-based filtering in collaborative approach. Utilize some rules of collaborative filtering in content based approach. Create a unified recommender system that brings together both approaches.

Most of the existing works are building the offline content optimization model. In this offline model gather the historical activities of user and user taste based on that content are optimized. Since offline model can't provide better result. Hence online content optimization were introduced that consider the collaborative filtering especially hybrid approach to leverage the user activities.

[9] represent online learning content modelling Estimating Click through rating for portal web site homepage gets millions user visit per day, In the offline model based on historical details can estimate the CTR. But the online model needs exploratory analysis of CTR. Here CTR estimated at 5 minute time intervals. Serving bucket collect those aggregated histories. In proposed parallel serving bucket established to improve online content learning.

EMP (estimating most popular) model one of the online models for content optimization [11].EMP model is based on the log-odd record .these record gathered from the serving buckets. 't' is the initial time of user action then 't1' is after 5 min contents are get started in the serving bucket. For this logistic transformation applied to estimate most popular (EMP) site.

III. ONLINE LEARNING FOR PERSONALIZED CONTENT OPTIMIZATION

Here online learning frame work is introduced and further by each-item model, personalization and segmentation are performed for content optimization.

A. Online Learning Frame Work

This framework has the parallel serving bucket which has random learning bucket and serving bucket. When the user visit the portal website content are gathered into the random learning bucket that contents are web links, clicks and views ,user details. Then random learning bucket estimate the CTR every time interval and sampled by the serving bucket.CTR is the strong signal of the user interest and behaviour. Serving bucket build each-item model and segmentation for the recommendation from random learning bucket within the every cycle interval.

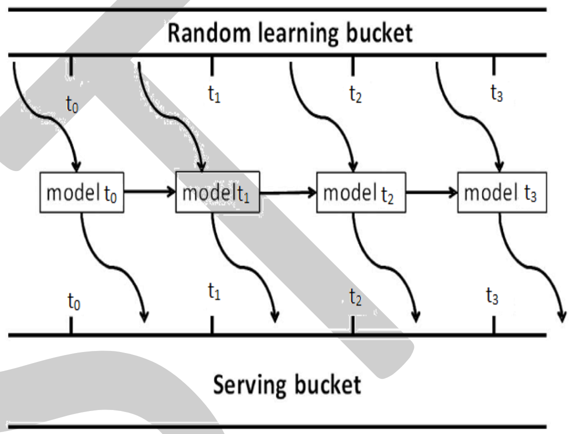

Fig 2: Online Learning Framework

From fig 2 when user entered into the portal web site every five minutes both buckets are updated simultaneously (time interval [t, t+1]).In the serving bucket each-item module build every 5 min updated corresponding item by user views and clicks according random learning bucket time interval [t,t+1].

B. Each Item Model

In this model used to build the effective recommender system for the portal website. Here the each candidate item are consider for the evolution for the better score of CTR. Exiting model[1] ,[3]consider only for direct CTR. This method Et value applied to the Updated frequent pattern tree from that prioritizing queue estimate the CTR and EMP (estimate most popular value. Assume [t, t+1] is the time in travel between the two model from random learning bucket.

$$E_{t+1} = \frac{\mathbf{r}_t E_t + c[t, t+1]}{\mathbf{r}_t + n[t, t+1]} \qquad (1)$$

Where r t is the sample size of the number of user equation (2) is applied in equation(1), which is updated as.,

$$r_t = w\, r_{t-1} + n[t-1, t] \qquad (2)$$

C. Segmentation

To integrate user segmentation into the online learning approach as, users are divided into a small number of groups, each of which has its exclusive online learning and serving process. In other words, each user group has its own per-item models which are learned based on clicks and views only from the users belonging to the corresponding group, and the serving results using these models are also only applicable to the users belonging to the corresponding group.

IV. ACTION INTERPRETATION FOR ONLINE LEARNING

Online learning algorithm fully based on user clicks and views .its CTR is estimated based on the number of clicks and views for this item, which implies that correct interpretation of user actions[1] is important since click/view samples are derived from the user actions logged by the portal website.

C. Click Event

Click Event is when user enters into the browser all the clicks are considered. Click event is engaging the user efficiency. This click event is considered for the CTR Estimation.

D. Click Other Event

Click other event consider the other events of website like search box, dialog box etc..., for the better estimation of CTR.

E. Non-Click Event

Apart from user click event and non-click event there is non-click event .After user enter to the website which has not been clicked by the user is non click event. However we can't consider this for CTR Estimation .To reduce this analyse the past historical activity of user.

V. UPDATED FREQUENT PATTERN TREE

This Updated frequent tree count values are used to estimate the ranking of each module. This tree constructed from service bucket. Here the root is the portal website first siblings are the portal module.

In Fig 3 'h' is the hight of the tree and 'n' is number of node for the parent tree CTR estimaion from the each item module based on this module confidance is estimated.

Portal website is the root of the tree.While user enter in to the "www.msn.com" tree has sibiling as the module then each and every click of module formed as the tree for the root first sibilings hight ,nuber of node under the sibiling are calculated .Then CTR estimation calculated by the each item module.

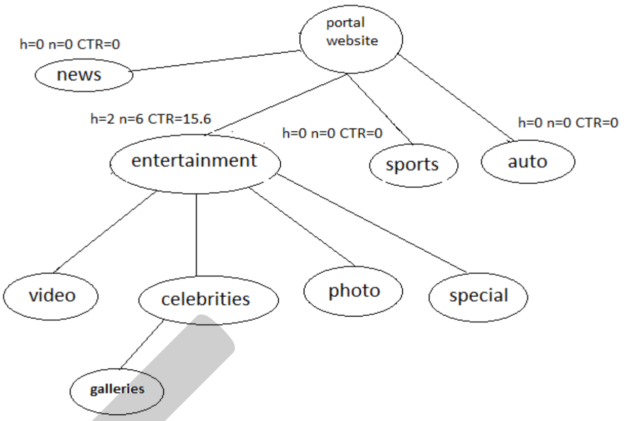

Fig 3 Updated Frequent Pattern Tree

Here tree wil estimate the count for the each module of first sibilings.

$$\text{Count}_{ij} = (h/n) + h \tag{3}$$

The equation (3) will appy to the prioritizing queue model.

This value is consider for the estimation of most popular model.EMP-prioritizing queue will be estimate the most popular module.support count value fixed follwed count greater then the support count value applied to the prioritizing queue.

Algorithm 1: Construction of updated frequent pattern tree FP-Tree

Input : From parallel serving bucket and prioritizing queue

Output: Updated FP-tree, each time frequency stored in prioritizing queue.

Step1: Sort the items in the from the serving bucket based on clicks and views.

Step 2: Create the root of the tree R. Since it is a prefix tree, R= NULL.

Step 3: For each serving bucket.

do

Step 4: Let form the tree for each fist siblings K and L is rest of the items under the first siblings.

Step 5: If the root R has a direct child node L, such that L's item_name = K's item_name, .

Step 6: For each item in K do the following steps up to K is empty.

Step7: **for** each sibling frequency value is calculated. Then for each module count stored in prioritizing queue.

Fig 4: Uptated Frequent Pattern Tree Algorithm

VI. PRIORITIZING QUEUE

In this queue contain the ranking based information .this updated by based on the Updated frequent pattern tree. In this EMP-prioritizing queue initially estimate CTR_k

$$\text{Then } CTRk = Count_{ij} + P[t+1] \quad (4)$$

Then previous and current value gathered and estimates the most popular .based on most popular value ranking model created for the portal website.

$$CTR_{k+1} = Count_{i+1,j} + P[t+1] + CTR_{k+1} \quad (5)$$

Finally obtain the CTR value which value is higher that will be present in prioritizing queue this will be applied to the sub modules

Algorithm 2:EMP-Prioritizing queue

Input : MFPT count values and CTR value.

Output: Ranking model.

Step1: at the time t_0 .initial values of each module 0.

Step2: At the time $t_{1.}$ Module gets ranked based on Updated frequent pattern tree.

Step3: For estimate Updated frequent pattern tree at time t_2 .t_1 value send to the MFPT.

Step4: Then continue the step3:

Fig 5: EMP-Prioritizing Queue Algorithm

VII. RESULT

The data set for online learning frame work are gathered directly from the "www.msn.com".

link	Time (mm.ss)	user
http://www.msn.com/	00.16	gokul
http://entertainment.in.msn.com/	00.36	gokul
http://entertainment.in.msn.com/#	00.08	gokul
http://entertainment.in.msn.com/southcinema/	03.40	gokul
http://www.msn.com/	00.21	gk
http://news.in.msn.com/	01.28	gk
http://news.in.msn.com/topstories/	01.02	gk
http://news.in.msn.com/crimefile/	01.49	gk

Fig 6: Online Learning Frame Work

Each user link and the time interval of staying on the particular module store in the learning frame work.

While browsing the user on portal web site online learning frame work gathered user clicks and views.

TABLE1
USER CLICKS AND VIEWS.

Module	Clicks and Views
News	3
Entertainment	6
Sports	4
Auto	2
Life Style	0

Then action Interpretation model Estimate the CTR by the Equation (1) followed by user action interpretation constructed. Enhancement of Updated frequent pattern tree and Prioritizing queue work has been started and gathered sets are applied.

VIII. CONCLUSION AND FUTURE WORK

This method builds the parallel serving bucket to explore and personalized the user activities in the web portal. Updated Frequent pattern tree and prioritizing queue utilize the all historical activities to produce better result than user action model. Comparison of the various model will shows the better result in this paper.

In the case of future work prioritizing queue is need to be establish and also possible to enhance GPS of position of user can consider that will more effective of personalized content optimization.

REFERENCE

[1]Jiang Bian,Anlei Dong,Xiaofeng He,Srihari Readdy and Yi Chang "*User Action Interpretation for Online Content Optimization*" ieee Transaction on Knowledge and data engineering , vol.25,no 9, September 2013.

[2] Dolly Sigroha,Chhavi Rana "*Survey Paper on Analysis of Various Recommendation Algorithms*"(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (2) , 2012,3406-3408

[3] Daniar Asanov, "*Algorithms and methods in recommender system* ," in In: Berlin Institute of Technology, 2010.

[4] H.-N. Kim, I. Ha, K.-S. Lee, G.-S. Jo, and A. El-Saddik, "*Collaborative User Modeling for Enhanced Content Filtering in Recommender Systems*," Decision Support Systems, vol. 51, no. 4, pp. 772-781, 2011.

[5]J. Wang, A.P. de Vries, and M.J.T. Reinders, "*Unifying User-Based and Item-Based Collaborative Filtering Approaches by Similarity Fusion*," Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2006.

[6]J. Liu, P. Dolan, and E.R. Pedersen, "*Personalized News Recommendation Based on Click Behavior*," Proc. 15th Int'l Conf. Intelligent User Interfaces (IUI), 2010.

[7]Le.M. de Campos, J.M. Ferndez-Luna, J.F. Huete, and M.A. Rueda-Morales, "*BAYESIAN Bayesian Framework For User Intrest Prediction*" Int'l J. Approximate Reasoning, vol. 51, no. 7, pp. 785-799, 2010.

[8]C. Liu, R.W. White, and S. Dumais, "*Understanding Web Browsing Behaviors Through Weibull Analysis of Dwell Time*," Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2010.

[9]M. Richardson, E. Dominowska, and R. Ragno, "*Predicting Clicks: Estimating the Click-Through Rate for New Ads*," Proc. 16th Int'l World Wide Web Conf. (WWW), 2007.

[10]D. Agarwal, B.-C. Chen, and P. Elango, "*Spatio-Temporal Modelsfor Estimating Click-Through Rate*," Proc. 18th Int'l World Wide Web Conf. (WWW), 2009.

[11]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.

[12]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.