INDIAN INSTITUTE OF MANAGEMENT KOZHIKODE

Working Paper

IIMK/WPS/224/ITS/2017/08

February 2017

Ctx-effSAMWMIX: A Contextual Multi-Armed Bandit Algorithm for personalized recommendations

Boby Chaitanya Villari[1]
Mohammed Shahid Abdulla[2]

[1] Doctoral Student, IT & Systems Area, Indian Institute of Management Kozhikode, IIMK Campus P.O, Kerala – 673570, India, E-mail: Bobycv06fpm@iimk.ac.in
[2] Associate Professor, IT & Systems Area, Indian Institute of Management Kozhikode, IIMK Campus P.O, Kerala – 673570, India, E-mail: shahid@iimk.ac.in, Phone: +91 - 495 - 2809254

IIMK WORKING PAPER

Ctx-*effSAMWMIX*: A Contextual Multi-Armed Bandit Algorithm for personalized recommendations

Boby Chaitanya Villari
Doctoral Student of IT & Systems Area, IIM Kozhikode

Mohammed Shahid Abdulla
Associate Professor of IT & Systems Area, IIM Kozhikode

Abstract

Machine Learning algorithms play an active role in modern day business activities and have been put to an extensive use in the marketing domain as well. In Ecommerce domain, these algorithms play an important role in suggesting recommendations to users, be it a merchandise of interest to the user or a news article for a website visitor. Due to the larger variety of available information and multiplicity in the merchandise based data, these personalized recommendations play a major role in the successful business activity that could be a sale in the case of an Ecommerce website or a click on a news article in case of a news website. The personalized recommendation problem, where the challenge is to choose from a set of available choices to cater to a target user group, can be modelled as a Contextual Multi-Armed Bandit problem. In this work we propose Ctx-effSAMWMIX which is based on LinUCB and effSAMWMIX algorithms. We empirically test the proposed algorithm on Yahoo! Frontpage R6B dataset by using an unbiased offline evaluation technique proposed in literature. The performance is measured on Click Through Rate (CTR) which effectively reports the ratio of Clicks the recommended articles obtained to that of total recommendations. We compare the performance of Ctx-effSAMWMIX with LinUCB and a random selection algorithm and also report the results of t-tests performed on the mean CTRs.

Keywords: Contextual Multi-Armed Bandit, Unbiased offline evaluation, personalized recommendations

1 Introduction

Machine Learning (ML) is being actively adopted by organizations that cater to plethora of different tasks. Today, ML is used to diagnose diseases, stop crime, drive cars, predict Tsunamis & draughts and many more. It is common for the business organizations to use modern day marketing applications aided by Machine Learning technologies (ML) to interact with customers and make sensible recommendations. ML algorithms analyze the data, observe patterns from it and could provide predictive and prescriptive suggestions. ML is used across all fields from Marketing analytics to Financial Analytics. Digital Ad-marketplaces use them to serve targeted Ads, Ecommerce companies use them to send across discount coupons to specific customers, Banks and Financial institutions send appealing loan or mortgage offers to targeted customers. The motto of such targeted marketing campaigns is quite apparent – to customize the offers (to consumers with specific interests) such that the personalized marketing would resonate better with the customers. Such an effort could yield a greater customer engagement, loyalty to the brand being marketed and higher customer spending. It's a win-win situation for both consumers and sellers. These techniques will also help the marketers to calculate and optimize Life Time Value of customers, their segmentation, churn prediction and also to identify which marketing campaign investments turn risky or void at what point of time. In the field of Display and Targeted Internet advertising, ML is used to recognize the innate characteristics of the content posted on web pages. Today these ML algorithms, which could even parse the social networking sites, are capable of recognizing brands (logos) from the posted images. Firms like GumGum[1] pioneer in such visual marketing techniques.ML algorithms will help in finding patterns from the data which the marketers would miss otherwise or rather would take a larger time to process. With growing data, reduced prices of electronic data storage and availability of cheaper data processing platforms like Apache Hadoop, ML is already a technological revolution that every business organization is willing to be a part of.

1.1 Machine Learning Algorithms & Contextual Multi Armed Bandits

Though ML is not necessarily a new tool, it has gained prominence in last two decades. Currently, Machine Learning is regarded as a modern day's scientific discipline that involves using of advance mathematics, computer science, statistics and could handle huge volumes of data. It is mainly used to see patterns and make predictions from data. Machine Learning techniques which were previously confined to Supervised learning and Unsupervised learning are now extended to a more dynamic mode of learning called Reinforcement learning.

[1] http://www.gumgum.com/

In Supervised learning, a decision making model is built over a labeled input data which is used for training the computer system (or machine). The model is tested for the accuracy of its predictions when employed over a test data. If the predictions are inaccurate, the model is corrected accordingly. Regression and classification are classical examples of supervised learning. In Unsupervised learning, the input data is not labeled and the model is built in order to deduce structures present in the input data. This could use advanced mathematics to reduce redundancies in data and to organize data using similarity measures. Clustering is a classic example of Unsupervised learning.

Reinforcement learning (RL) is more of a later phenomenon that has gained prominence in last decade. In the field of robotics, RL is in action in applications like self-driving cars to unmanned aircrafts. In marketing, RL is mainly being used for recommendations. In RL, the hidden patterns in the data are leveraged to predict the next action which could be recommending a product to a user, showing a news content to a viewer or showing an ad to the visitor of the website. In contrast to supervised learning, RL algorithms do not need a training input/output samples but the learning happens through trial-and-error based schema. Unlike the unsupervised learning methods which majorly fall towards applications like clustering or segmentation, RL's applications could be used to propose recommendations. In marketing, RL is used in combination with Collaborative Filtering, Bayesian networks etc., to generate recommendations. There are many business example applications of RL based Personalized Recommendation Systems (RS) which range from friend recommendations seen in social networking sites to recommendations provided by the Ecommerce platforms. E-commerce sites like Netflix recommend movies to users. Facebook suggests the tags to the photos being uploaded.

In this work we discuss about implementing a personalized recommendation using a RL based Contextual Multi-Armed Bandit algorithm named Ctx-effSAMWMIX. Personalized recommendation services identify the preferences of users and appropriately show the web content to suit to their preferences. A News article recommendation task is such a problem where the website chooses an article from its repository and shows it to the visitor with an aim to maximize the readership i.e. the visitor clicking the article link to view the same. The success of such a personalized recommendation system (PRS) is measured by Click Through Rate (CTR) (Li, Chu, Langford, & Wang, 2011) which is the ratio of total clicks obtained to total number of times the PRS chose to show news articles to various users. Content based RS and Collaborative RS may not retain high accuracies if large number of visitors (users) or news articles (items) are new to the system. Such an issue is referred to as Cold-start problem(Mary, Gaudel, & Philippe, 2014; Schein, Popescul, Ungar, & Pennock, 2002) and in such situations the recommendation task can be modelled as a Contextual Multi-armed bandit problem (CMAB). In the following sections we briefly review CMAB literature, introduce Ctx-effSAMWMIX, describe the experimental setup and discuss the results.

2 Contextual Multi-Armed Bandits – A brief review

It is easy to comprehend that an algorithm (or agent) makes decisions based on the feedback (observations) from the environment (world). If the observations are incomplete, the decision making process becomes tricky and challenging. Multi-armed bandit (MAB) algorithms suit well for such decision making under uncertainty(Vermorel & Mohri, 2005). In such a decision making situation, there would be a series of trials (iterations) in each of which the agent has to choose from a set of available choices (arms). In case if the agent can use any contextual or side information, that is obtained from the environment, in each of these iterations these MABs can be modelled as Contextual MABs (CMAB). In MAB literature, CMABs are also referred as Associative Bandits, Bandits with covariates (Langford & Zhang, 2008).

Like in MAB, in every iteration, the CMAB agent obtains information on the reward which could be following a distribution completely unknown to the agent but is determined by the contextual information. The agent which is willing to maximize the cumulative reward over the series of iterations will validate its own performance through the reinforcements it receives from the environment. In the current work we model the news article recommendation challenge as a CMAB problem where a news item represents an arm and each user visit represents an iteration. The user demographics provide the necessary contextual information. CMABs proposed in the extant literature (LinUCB(Li, Chu, Langford, & Schapire, 2010) , EXP4.P & EXP4 (Beygelzimer, Langford, Li, Reyzin, & Schapire, 2011)), were modelled similarly as explained below. As surveyed in (Zhou, 2015) , in addition to LinUCB,EXP4 and EXP4.P , a few other CMAB algorithms are proposed in literature. They are EXP4.P with infinite Experts, Epoch-Greedy algorithm, Randomized UCB, ILOVETOCONBANDITS, Thompson Sampling with Linear Regression, Thompson Sampling with Logistic Regression.

In a CMAB setting, There are a set of arms (news articles) A_t available to the algorithm at iteration (user visit) t which is associated with the contextual information vector X_t. Using the previously acquired knowledge and the context in the current iteration (user visit) t, the algorithm chooses to show an arm (news article) a_t and obtains a reward r_t^a. This reward being dependent on the contextual features and the chosen article, the expectation of the cumulative reward is thus a function of X_t and a_t. The algorithm stores the reward information obtained in the current iteration and uses it for decision making in the subsequent iterations. In the following section we discuss the algorithms and the experimental set up.

3 Algorithms, Data & Experiments

Evaluating a Contextual Multi-Armed Bandit (CMAB) algorithm for Online evaluation has always been a challenging task mainly due to limited availability of data. The evaluator ideally desires for datasets that explicitly contain the data which forms the basis for evaluation, like the changes in users' preferences, demographics etc. Thus we evaluated the Ctx-effSAMWMIX on Yahoo! Today News Frontpage (R6B) dataset.(Vanchinathan, Nikolic, De Bona, & Krause, 2014). R6B dataset, released through the Yahoo! Webscope Program, provides user view/click data log for those articles displayed on Yahoo! Frontpage Today module over a 15-day period from October 2 to 16,2011. This R6B dataset, consisting of a total of 28,041,015 user visits to the Yahoo! Frontpage Today module, is larger than its predecessor Yahoo! R6A dataset in the size raw features. Each of these user visits is a single line in the data file whose structure is as shown in quotes below

“1317513293 id-564335 1 |user 1 7 11 36 65 13 22 23 33 32 16 18 38 24 26 17 42 45 35 19 44 25 40 29 75 15 47 43 14 70 30 50 27 21 |id-552077 |id-555224 |id-555528 |id-559744 |id-559855 |id-560290 |id-560518 |id-560620 |id-563115 |id-563582 |id-563643 |id-563787 |id-563846 |id-563938 |id-564335 |id-564418 |id-564604 |id-565364 |id-565479 |id-565515 |id-565533 |id-565561 |id-565589 |id-565648 |id-565747 |id-565822”

Table 1: Components of a data line in Yahoo! R6B dataset

Tuple data	**What does the tuple represent**
1317513293	Timestamp which is considered as a Unique user
id-564335	Arm or Article Id
1	User Click Status (1 if article obtained Click and 0 otherwise)
|user	String indicating the start of User's contextual features
1 7 11 36 65 13 22 23 33 32 16 18 38 24 26 17 42 45 35 19 44 25 40 29 75 15 47 43 14 70 30 50 27 21	Binary vectors indicating user’s contextual features(136 dimensional vector).These carry information such as user's age, gender etc.
|id-552077 |id-555224 |….|id-565747 |id-565822	The list of arms(articles) that are available to user during that particular visit where from the article which is shown is chosen

Table 1 explains about each tuple in the data line. The term “user visit” refers to an event that logs user related information such as users’ age, gender, demographics and other features related to users’ clicking behavior .Extant literature (Chapelle & Li, 2011; Li et al., 2010; Li et al., 2011; Tang, Jiang, Li, & Li, 2014) has used Yahoo! Frontpage News datasets to evaluate CMAB algorithms. This work performs experiments on the complete dataset which is split in to 15 different files, utilizing the available data to the fullest.

This work utilizes the Replay methodology(Li et al., 2011) where a historical user event log is used as a proxy for the online evaluation of the CMAB algorithms. It is to be noted that this Replay methodology provides an unbiased evaluation of the CMAB algorithm. In these experiments a CMAB algorithm which is under contention is evaluated over a performance measure called Click Through Rate

(CTR). CTR is the average reward and is calculated as $CTR = \frac{1}{n}\sum_{t=1}^{n} r_t$ where n is the number of trials and r_t is the reward obtained in trial t.The CMAB algorithm that obtains a higher CTR is preferred to the rest in comparison. Also, since the Yahoo! Data corresponds to picking a News article at random, the work reports the overall CTR of a Random Selection algorithm. The Ctx-effSAMWMIX algorithm is compared with LinUCB and the Random Selection (RandSel henceforth) Algorithms. The pseudo codes and working methodologies of each of the three CMAB algorithms are put below.

3.1.1 The Random Selection (RandSel) algorithm

The Random Selection algorithm works by randomly selecting a News Article (an Arm) from the available list of Arms. This algorithm emulates the process in which Yahoo! displayed the news articles to its user database during the data collection period. This algorithm is applied and the CTRs are noted so that we can compare the improvements in CTR obtained by the proposed CMAB algorithm.

Algorithm 1: Random Selection (RandSel) Algorithm

$\boldsymbol{for}\ t - 1,2,3, \dots..T\ \boldsymbol{do}$
  $Observe\ \&\ note\ the\ arms\ present\ A_t$
  $Count\ the\ arms\ A_t\ and\ store\ as\ N_{A_t}$
  $Pick\ a\ random\ number\ r_t\ between\ 1\ \&\ N_{A_t}$
  $Choose\ arm\ a_t = r_t^{th}\ element\ of\ A_t$
$\boldsymbol{end\ for}$

3.1.2 The LinUCB algorithm

The LinUCB algorithm is proposed in the seminal work (Li et al., 2010) is extended in to Ctx-effSAMWMIX which is proposed in this work. We have put the LinUCB algorithm below for any reference.

Algorithm 2: LinUCB Algorithm (Li et al., 2010)

$With\ Inputs: \alpha \in R_+$
$\boldsymbol{for}\ t - 1,2,3, \dots..T\ \boldsymbol{do}$
  $Observe\ features\ of\ all\ arms\ a \in A_t : x_{t,a} \in R^d$
  $\boldsymbol{for\ all}\ a \in A_t\ \boldsymbol{do}$
    $\boldsymbol{if}\ a\ is\ new\ \boldsymbol{then}$
      $A_a = I_d\ (\ d - dimensinal\ identity\ matrix)$
      $b_a = 0_{dX1}(\ d - dimensinal\ zero\ vector)$
    $\boldsymbol{end\ if}$
    $\boldsymbol{\theta}_a = A_a^{-1} b_a$
    $p_{t,a} = \boldsymbol{\theta}_a^T x_{t,a} + \alpha \sqrt{x_{t,a}^T A_a^{-1} x_{t,a}}$
  $\boldsymbol{end\ for}$
$Choose\ arm\ a_t = argmax_{a \in A_t}\ p_{t,a}\ with\ ties\ broken$
$arbitrarily\ and\ observe\ a\ real - valued\ payoff\ r_t$

12. $A_{a_t} = A_{a_t} + x_{t,a_t} x_{t,a_t}^T$
13. $b_{a_t} = b_{a_t} + r_t \, x_{t,a_t}$
14. $\boldsymbol{end\ for}$

3.1.3 The Ctx-effSAMWMIX algorithm

The Ctx-effSAMWMIX is formulated using LinUCB and effSAMWMIX (Boby & Abdulla, 2016) where LinUCB is the prior whose output is fed to effSAMWMIX but with a modification when a new arm (news article) is added. The input parameter α is same as given in (Li et al., 2010) while d represents the distance between the means of two of the closest arms. This is the same d parameter required for effSAMWMIX algorithm proposed in (Boby & Abdulla, 2016) . For example a $d = 0.1$, as set in our experiments indicates that the means of the rewards of closest arms could have a maximum difference of about 10%.

Algorithm 3: Ctx-effSAMWMIX Algorithm

$With\ Inputs: \alpha \, , d \,\epsilon\, R_+$
1. $\boldsymbol{for}\ t - 1,2,3, \ldots \ldots T\ \boldsymbol{do}$
2. $Observe\ features\ of\ all\ arms\ a \,\epsilon\, A_t : x_{t,a} \epsilon\, R^d, store\ the\ armCount\ N_t$
3. $\boldsymbol{for\ all}\ a \,\epsilon\, A_t\ \boldsymbol{do}$
4. $\boldsymbol{if}\ a\ is\ new\ \boldsymbol{then}$
5. $A_a = I_d\ (\ d - dimensinal\ identity\ matrix)$
6. $b_a = 0_{dX1}(\ d - dimensinal\ zero\ vector)$
7. $\phi_a = 1/N_t$
8. $\boldsymbol{end\ if}$
9. $\boldsymbol{\theta}_a = A_a^{-1} b_a$
10. $p_{t,a} = \boldsymbol{\theta}_a^T x_{t,a} + \alpha \sqrt{x_{t,a}^T A_a^{-1} x_{t,a}}$
11. $\boldsymbol{end\ for}$
12. $Calculate\ \phi_s = \sum \phi_a \quad a \,\epsilon\, A_t$
13. $\boldsymbol{if}\ \phi_s \sim = 1,\ normalize\ \phi_a\ i.e.\ \phi_a = \grave{\phi_a} = \frac{\phi_a}{\phi_s}\ for\ all\ a \,\epsilon\, A_t$
14. $\underline{Utilizing\ the\ effSAMWMIX\ algorithm: Note\ p_{t,a}\ as\ rewards\ vector\ G_t\ \&\ N_t = N}$
15. *Calculate*
    a. $C_0 = N + 1;\ \sigma^2 = 2 * N;$
    b. $\eta_0 = \frac{1}{C_0} log\left(\frac{1 + C_p * d}{\sigma^2}\right)$
    c. $t_0 = ((4 + d) * N + d)/d^2$
16. $\boldsymbol{for}\ i\ = 1, \ldots, N\ \boldsymbol{do}$
    a. $Obtain\ reward\ X_{t=i}^i$
    b. $Initialize\ \phi_t^i\ = \eta_0 * \left(\frac{t_0}{N}\right) * \left(\frac{X_{t=i}^i}{\frac{1}{N}}\right)$
    c. $Initialize\ pull\ count\ for\ \ arm\ a^i\ as\ p_i$=1
17. $\boldsymbol{end\ for}$
18. $\boldsymbol{for}\ t\ = (t_0 + 1 + N), \ldots, (t_0 + T)\ \boldsymbol{do}$

a. Obtain random probability r

b. Choose an arm i as winner if $\sum \phi_t^i > r$ and store reward $G_{t-t_0}^i = a_t^*$ and normalize the reward using its probability $\widehat{X}=a_t^*/\phi_t^i$

c. Update $p_i = p_i + 1$

d. $\boldsymbol{for}\ d_t = 1, \dots, \frac{t-t_0}{T-N}\ in\ steps\ of\ d_{tstep}\ \boldsymbol{do}$

   i. Calculate $k_t = d + d_t$;

   ii. $\gamma_{tt} = ((4 + k_t) * N + k_t)/(t * k_t^2)$

   iii. $C_t = \left(\frac{N}{\gamma_{tt}}\right) + 1$ and $\sigma_{tt}^2 = 2 * N/\gamma_{tt}$

   iv. $\eta_t = \frac{1}{C_t} \log\left(\frac{1+C_t*k_t}{\sigma_{tt}^2}\right)$

   v. $\sum \phi_{tt}^i = \sum \phi_t^i + e^{\sum \eta_t \widehat{X_t^i}}$

   vi. If $e^{\sum \eta_t*(d_t*\widehat{X_t^i}) + \eta_t*(d_t*\widehat{X_t^{a^i}})} > \sum \phi_{tt}^i$ then assign $d_{tt} = d_t - d_{tstep}$

e. $\boldsymbol{end\ for}$

f. *Assign*

   i. $K_t = d_t + d_{ttrue}$

   ii. $\gamma_t = ((4 + k_t) * N + k_t)/(t * k_t^2)$

   iii. $C_t = \left(\frac{N}{\gamma_{tt}}\right) + 1$ and $\sigma_t^2 = 2 * N/\gamma_{tt}$

   iv. $\eta_t = \frac{1}{C_t} \log\left(\frac{1+C_t K_t}{\sigma_t^2}\right)$

g. $Now\ update\ \phi_{t+1}^j\ using\ \phi_{t+1}^j = (1-\gamma_t)\frac{e^{\sum \eta_t \widehat{X_t^j}}}{\sum_{j=1}^N e^{\sum \eta_t \widehat{X_t^j}}} + \frac{\gamma_t}{N}$

$\boldsymbol{end\ for}$

$Choose\ arm\ a_t = argmax_{a \epsilon A_t}\ \phi_{t,a}\ with\ ties\ broken$
$arbitrarily\ and\ observe\ a\ real-valued\ payoff\ r_t$

$A_{a_t} = A_{a_t} + x_{t,a_t} x_{t,a_t}^T$

$b_{a_t} = b_{a_t} + r_t\, x_{t,a_t}$

$\phi_{a_t} = \phi_a$

$\boldsymbol{end\ for}$

Ctx-effSAMWMIX, like LinUCB, is a General Purpose CMAB algorithm which can also be applied to cases other than News Article Recommendations. For each CMAB algorithm, the performance is tested over the entire data which is spanned over 15 files each code named File-002 to File-015.Except LinUCB, the other two algorithms are randomized. For Ctx-effSAMWMIX, the d parameter is set to 0.1 , meaning that the rewards of the closest arms could differ by 10% (Boby & Abdulla, 2016). If observed carefully, even in LinUCB, an arm is randomly selected in a case with same UCB values for multiple arms in a particular iteration. Thus the randomness could vary the performance of the algorithms (even though only slightly for LinUCB). In order to statistically test the algorithm, each of the above three algorithms is run for 30 times over each of the 15 data files. Thus the mean, standard deviation, minimum & maximum of overall CTR is obtained 15 times (See Table 2 for descriptive statistics of the same.).

Table 2: CTR obtained by CMAB algorithms on data files

	Random Algorithm				LinUCB Algorithm				Ctx-effSAMWMIX Algorithm			
File Number	Mean	Standard Deviation	Minimum	Maximum	Mean	Standard Deviation	Minimum	Maximum	Mean	Standard Deviation	Minimum	Maximum
File-002	0.0341	0.0018	0.0330	0.0357	**0.0468**	0.0018	0.0445	0.0500	**0.0480**	0.0016	0.0457	0.0506
File-003	0.0388	0.0004	0.0368	0.0398	**0.0545**	0.0004	0.0537	0.0551	**0.0606**	0.0108	0.0513	0.0921
File-004	0.0329	0.0014	0.0320	0.0346	**0.0453**	0.0014	0.0430	0.0472	**0.0532**	0.0168	0.0064	0.0684
File-005	0.0385	0.0010	0.0366	0.0401	**0.0584**	0.0010	0.0567	0.0607	**0.0646**	0.0064	0.0556	0.0748
File-006	0.0412	0.0016	0.0385	0.0427	**0.0739**	0.0016	0.0718	0.0772	**0.0780**	0.0194	0.0567	0.1013
File-007	0.0420	0.0012	0.0369	0.0437	**0.0545**	0.0027	0.0516	0.0604	**0.0618**	0.0241	0.0426	0.1226
File-008	0.0337	0.0007	0.0324	0.0347	**0.0441**	0.0007	0.0425	0.0452	**0.0639**	0.0245	0.0447	0.1478
File-009	0.0316	0.0012	0.0303	0.0330	**0.0424**	0.0012	0.0407	0.0449	**0.0566**	0.0134	0.0441	0.0750
File-010	0.0345	0.0007	0.0333	0.0359	**0.0499**	0.0005	0.0487	0.0507	**0.0537**	0.0039	0.0450	0.0566
File-011	0.0356	0.0008	0.0335	0.0370	**0.0467**	0.0005	0.0457	0.0478	**0.0669**	0.0171	0.0464	0.0964
File-012	0.0399	0.0007	0.0384	0.0411	**0.0594**	0.0014	0.0560	0.0631	**0.0596**	0.0175	0.0423	0.0965
File-013	0.0402	0.0007	0.0377	0.0420	**0.0562**	0.0007	0.0551	0.0582	**0.0577**	0.0141	0.0417	0.1050
File-014	0.0385	0.0016	0.0374	0.0417	**0.0478**	0.0016	0.0451	0.0508	**0.0614**	0.0109	0.0497	0.0793
File-015	0.0368	0.0017	0.0349	0.0389	**0.0449**	0.0017	0.0432	0.0523	**0.0654**	0.0200	0.0458	0.1121
File-016	0.0337	0.0006	0.0325	0.0346	**0.0503**	0.0006	0.0494	0.0517	**0.0630**	0.0122	0.0499	0.0963

The Ctx-effSAMWMIX algorithm has always had a better average performance than LinUCB algorithm on all the 15 data files but with a greater standard deviation. LinUCB is evaluated with parameter $\alpha = 0.1$ which gave a stable performance(Tang et al., 2014).To conclude that the mean rewards(CTR) of Ctx-effSAMWMIX is statistically different from those of LinUCB and Random algorithms, t-tests (assuming unequal variances) are performed on the data. The results are put in the Tables 3 & 4 below.

Table 3: Two Sample t-test of mean CTR of Ctx-effSAMWMIX against Random Policy

	Against Random Policy(at alpha =0.1)			
Dataset File	**p-value in 1-tail t-test**	**Result**	**p-value in 2-tail t-Test**	**Result**
File-002	1.11E-35	Pass	2.22E-35	Pass
File-003	1.53E-11	Pass	3.06E-11	Pass
File-004	6.23E-08	Pass	1.25E-07	Pass
File-005	3.99E-21	Pass	7.98E-21	Pass
File-006	5.98E-10	Pass	1.20E-09	Pass
File-007	5.35E-05	Pass	0.000107	Pass
File-008	6.84E-08	Pass	1.37E-07	Pass

File-009	2.34E-11	Pass	4.68E-11	Pass
File-010	3.96E-23	Pass	7.93E-23	Pass
File-011	2.88E-11	Pass	5.76E-11	Pass
File-012	4.79E-07	Pass	9.59E-07	Pass
File-013	2.53E-07	Pass	5.06E-07	Pass
File-014	3.15E-12	Pass	6.29E-12	Pass
File-015	6.07E-09	Pass	1.21E-08	Pass
File-016	5.67E-14	Pass	1.13E-13	Pass

Table 4: Two Sample t-test of mean CTR of Ctx-effSAMWMIX against LinUCB Policy

	Against LinUCB Policy(at alpha =0.1)			
File Number	**p-value in 1-tail t-test**	**Result**	**p-value in 2-tail t-Test**	**Result**
File-002	5.93E-03	Pass	1.19E-02	Pass
File-003	5.87E-03	Pass	1.17E-02	Pass
File-004	4.30E-03	Pass	8.60E-03	Pass
File-005	6.37E-06	Pass	1.27E-05	Pass
File-006	1.72E-01	**Fail**	3.45E-01	**Fail**
File-007	5.53E-02	Pass	0.110502	**Fail**
File-008	4.81E-05	Pass	9.62E-05	Pass
File-009	1.39E-06	Pass	2.77E-06	Pass
File-010	5.44E-06	Pass	1.09E-05	Pass
File-011	2.08E-07	Pass	4.17E-07	Pass
File-012	4.70E-01	**Fail**	9.41E-01	**Fail**
File-013	1.93E-01	**Fail**	3.87E-01	**Fail**
File-014	2.34E-07	Pass	4.68E-07	Pass
File-015	2.47E-06	Pass	4.95E-06	Pass
File-016	2.40E-06	Pass	4.80E-06	Pass

The mean CTRs of Ctx-effSAMWMIX and Random Algorithm are from different distributions as the t-Test rejects the null hypothesis that both the means are the same. But we observed that in files named File-006, File-007, File-012 & File-013, Ctx-effSAMWMIX's mean cumulative CTR is close to that of LinUCB and the t-test fails at $\alpha = 0.1$. In these 4 data files, the t-test fails to reject the hypothesis that the mean CTRs of Ctx-effSAMWMIX & LinUCB are the same. Though Ctx-effSAMWMIX has performed better with a higher mean CTR than LinUCB in all the 15 dataset files, the (final) cumulative CTRs of LinUCB and Ctx-effSAMWMIX are closely spaced in these 4 (out of 15) dataset files. It is observed that LinUCB's performance improves with the number of iterations which is the lines of parsed data in the case of these datasets(Tang et al., 2014). This indicated that analyzing the CTR at multiple iterations (parsed line counts) would help in evaluating these algorithms' performances in a better way.

Also, it makes gives a clear idea (to the reader), if the relative performance of Ctx-effSAMWMIX to a Random algorithm or LinUCB are depicted graphically. From the dataset documentation it could be seen that, a news article (arm) is randomly chosen and shown to the user. A measure like R-CTR would help in evaluating how better or worse would the resulting CTR be if Instead of showing an article at random, a CMAB is employed. For doing the same, a measure called Relative CTR (R-CTR) is obtained by dividing the CTR obtained by Ctx-effSAMWMIX with Random Algorithm and with LinUCB. The graphical results are given in Figure 1 while final R-CTRs are given in Table 5.

Table 5: R-CTRs Performances of CMAB algorithms against each other

Dataset File	Ctx-effSAMWMIX Policy with respect to Random	Ctx-effSAMWMIX Policy with respect to LinUCB
File-002	1.4013	1.0177
File-003	1.5044	1.0707
File-004	1.6014	1.1674
File-005	1.7000	1.1174
File-006	1.8413	1.0254
File-007	1.4916	1.1541
File-008	1.6447	1.2581
File-009	1.8521	1.3824
File-010	1.5669	1.0841
File-011	1.8040	1.3737
File-012	1.5917	1.0686
File-013	1.4065	**1.0062**
File-014	1.6298	1.3132
File-015	1.8593	**1.5207**
File-016	1.9021	1.2776

The graphs of R-CTRs measured on the 15 dataset files are put in Appendix A. In the following Figure 1 we put the graphs where Ctx-effSAMWMIX performed the best (on File-015) and weakest (on File-013) when compared to LinUCB.

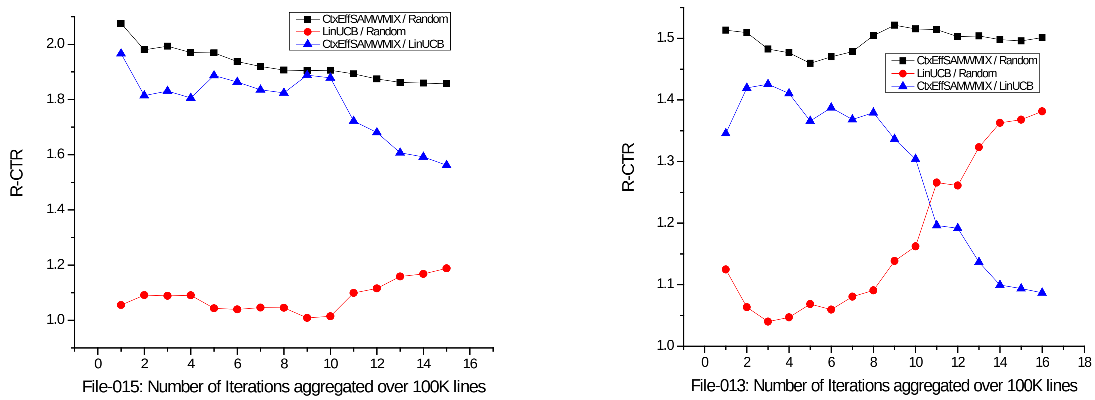


Figure 1: Strongest(File-015) & Weakest (File-013) R-CTRs of Ctx-effSAMWMIX algorithm measured over iterations

The number of iterations are aggregated over 100,000 lines of data and the cumulative CTR up to that iteration is used to calculate R-CTR. For example, the legend Ctx-effSAMWMIX/LinUCB indicates that the curve shows the R-CTR measured as follows

$$\mathrm{R-CTR(CTX-effSAMWMIX/LinUCB)} = \frac{\mathrm{CTR\ (CTX-effSAMWMIX)}}{\mathrm{CTR(LinUCB)}} \qquad (1)$$

From Figure 2 (in Appendix A) , it is observed that Ctx-effSAMWMIX consistently performs better than LinUCB while the performance of LinUCB improves with the number of iterations. The blue line with ▲ legend falls towards 1.0 as the number of iterations increase. This indicates that the performance of LinUCB improves considerably with the number of iterations while that of Ctx-effSAMWMIX falls only slightly and approaches that of LinUCB. Hence even though the t-test fails on a few dataset files, Ctx-effSAMWMIX's performance is seen to be better than that of LinUCB.

The reason for the betterment in the performance is straight forward. It is known that in a News article recommendation system, newer articles are added to the list of articles (arms) and some of the previously shown articles could be taken off from the list. In case of Ctx-effSAMWMIX, the ϕ values get normalized in order to summate to 1 ensuring a reasonable importance being given to the newer article than in case of LinUCB.

Ctx-effSAMWMIX utilizes the UCB values obtained from LinUCB and then provides them as an input to the effSAMWMIX algorithm Since Ctx-effSAMWMIX builds on the LinUCB platform and further utilizes effSAMWMIX to obtain the best arm, it requires further computations. Ctx-effSAMWMIX takes approximately 15 % more CPU time than LinUCB. Since $\mathrm{R-CTR}\left(\mathrm{CTX}-\frac{\mathrm{effSAMWMIX}}{\mathrm{LinUCB}}\right) \geq 1$ holds true on all the dataset files, indicating that Ctx-effSAMWMIX has performed better and obtained a better CTR than LinUCB, we could consider the computational overload as a reasonable bargain. Considering that Yahoo! R6B is a benchmark dataset, it is safe to infer that using Ctx-effSAMWMIX is advantageous for the practitioner as it could require a lesser number of iterations than LinUCB.

4 Conclusion

In this work, we propose a Ctx-effSAMWMIX and empirically evaluate its performance on Yahoo! R6B dataset which is designed to apply the unbiased offline evaluation approach (Li et al., 2011). In parallel, we implemented LinUCB and Random Selection algorithms so as to understand and compare the performance of the proposed algorithm. The current CMAB approach to personalized web services is based on LinUCB and hence inherit the advantages of the same i.e. the requirement of simulator building to test an online algorithm like CMAB is avoided by the utilization of an unbiased offline evaluation technique. Ctx-effSAMWMIX has performed better than LinUCB, the base algorithm on which it is formulated. It is observed in previous researches that LinUCB betters CTRs with the increase in the parsed data lines i.e. iterations, which is expected for any MAB. Ctx-effSAMWMIX achieves better CTRs quicker than LinUCB. This betterment is attributed to the fact that in R6B dataset, when a new arm is added, Ctx-effSAMWMIX first initializes the probability weight of the newest arm as $\phi_a = 1/N_t$ and then will normalize the probabilities of all the available arms in that iteration. N_t is the number of arms available in that iteration t. This gives a better weight (importance) of choosing the newer arm when compared to initializing with a zero weight as in the case of LinUCB. In the case of news article recommendation problem, it is expected that newer articles could generate interest in the visitors and hence is natural to expect a click from the visitor. In addition to the input parameter α that is required in the case of LinUCB, Ctx-effSAMWMIX only requires an input parameter d which indicates how closely the mean rewards of the arms could be separated. In our experiments we set the parameter to be 0.1 which enforces a stringent test to the algorithm as it is told to choose between arms whose mean rewards are very close and only differ by about 10%. Ctx-effSAMWMIX is computationally efficient as well with only a 15% CPU time overload when compared to LinUCB but achieves better CTRs faster (in terms of data lines parsed). We have evaluated the algorithms on the complete dataset split in to 15 dataset files and averaged the results. On each of these dataset files, we conducted paired two sample t-tests on mean CTRs obtained by LinUCB and Ctx-effSAMWMIX. The results support to reject the hypothesis that difference between the mean CTRs is zero which is an intended and satisfactory result.

5 References

Beygelzimer, A., Langford, J., Li, L., Reyzin, L., & Schapire, R. E. 2011. ***Contextual Bandit Algorithms with Supervised Learning Guarantees***. Paper presented at the AISTATS.

Boby, C. V., & Abdulla, M. S. 2016. Working paper-IT & Systems. ***IIM Working paper Series***.

Chapelle, O., & Li, L. 2011. ***An empirical evaluation of thompson sampling***. Paper presented at the Advances in neural information processing systems.

Langford, J., & Zhang, T. 2008. ***The epoch-greedy algorithm for multi-armed bandits with side information***. Paper presented at the Advances in neural information processing systems.

Li, L., Chu, W., Langford, J., & Schapire, R. E. 2010. ***A contextual-bandit approach to personalized news article recommendation***. Paper presented at the Proceedings of the 19th international conference on World wide web.

Li, L., Chu, W., Langford, J., & Wang, X. 2011. ***Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms***. Paper presented at the Proceedings of the fourth ACM international conference on Web search and data mining.

Mary, J., Gaudel, R., & Philippe, P. 2014. Bandits Warm-up Cold Recommender Systems. ***arXiv preprint arXiv:1407.2806***.

Schein, A. I., Popescul, A., Ungar, L. H., & Pennock, D. M. 2002. ***Methods and metrics for cold-start recommendations***. Paper presented at the Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval.

Tang, L., Jiang, Y., Li, L., & Li, T. 2014. ***Ensemble contextual bandits for personalized recommendation***. Paper presented at the Proceedings of the 8th ACM Conference on Recommender Systems.

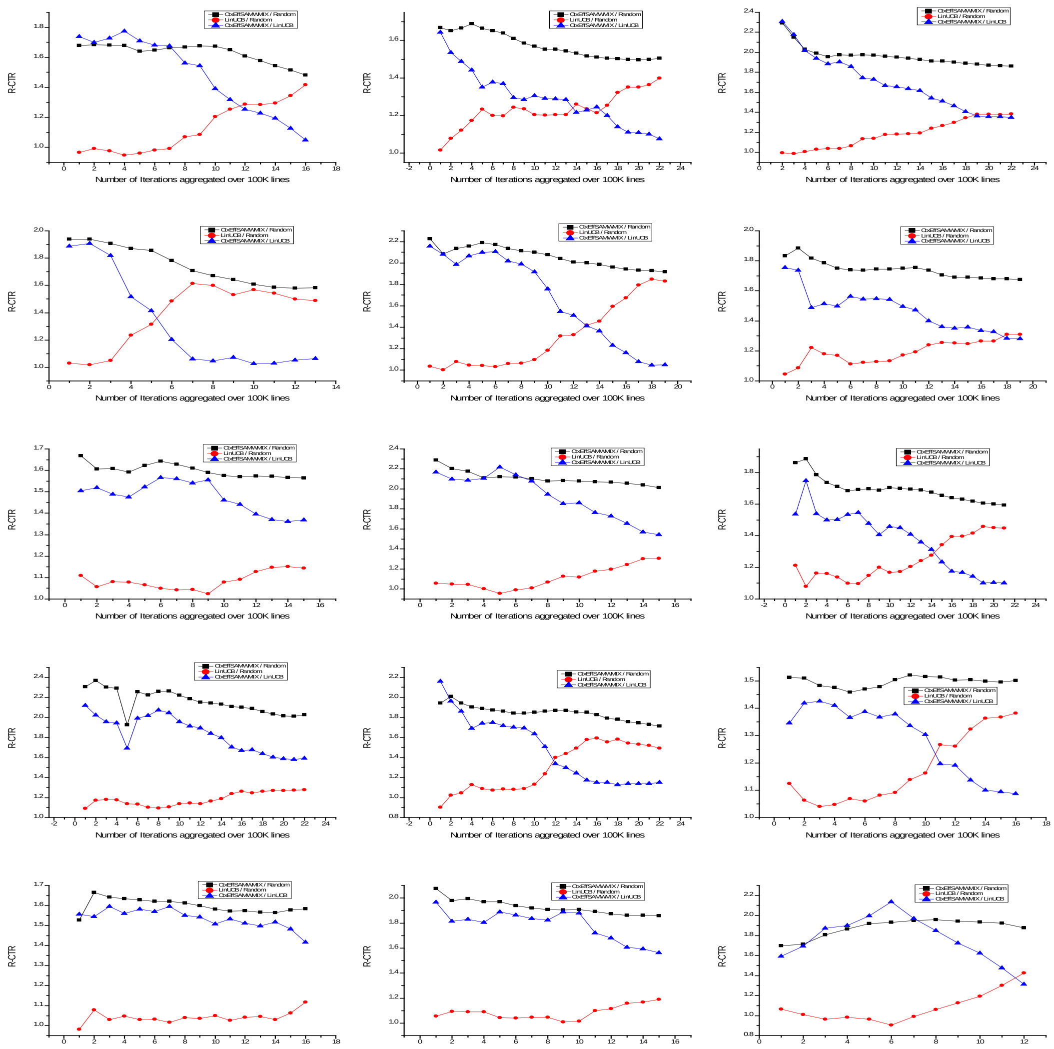
Vanchinathan, H. P., Nikolic, I., De Bona, F., & Krause, A. 2014. ***Explore-exploit in top-n recommender systems via gaussian processes***. Paper presented at the Proceedings of the 8th ACM Conference on Recommender systems.

Vermorel, J., & Mohri, M. 2005. ***Multi-armed bandit algorithms and empirical evaluation***. Paper presented at the European conference on machine learning.

Zhou, L. 2015. A survey on contextual multi-armed bandits. ***arXiv preprint arXiv:1508.03326***.

6 Appendix A

Figure 2: R-CTRs of Ctx-effSAMWMIX algorithm measured over iterations



Research Office

Indian Institute of Management Kozhikode

IIMK Campus P. O.,

Kozhikode, Kerala, India,

PIN - 673 570

Phone: +91-495-2809238

Email: research@iimk.ac.in

Web: https://iimk.ac.in/faculty/publicationmenu.php