CTR Prediction for DSP with Improved Cube Factorization Model from Historical Bidding Log

Lili Shan, Lei Lin, Di Shao, and Xiaolong Wang

School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China
{shanll,linl,dshao,wangxl}@insun.hit.edu.cn

Abstract. In the real-time bidding (RTB) display advertising ecosystem, demand-side-platforms (DSPs) buy ad impressions through real-time auction or bidding from ad exchanges for advertisers. Receiving a bid request, DSP needs predict the click through rate (CTR) for ads and determine whether to bid and calculates the bid price according to the CTR estimated. In this paper, we address CTR estimation in DSP as a recommendation issue. Due to the complicated trilateral interactions among users, ads and publishers (web pages), conventional matrix factorization does not perform well. Adopting ideas from high-order singular value decomposition (HOSVD), we extend two dimensional matrix factorization model to three dimensional cube factorization containing users, ads and publishers, and propose an improved cube factorization model to address it. We evaluate its performance over a real-world advertising dataset and the results demonstrate that the improved cube factorization model outperforms the matrix factorization.

Keywords: click through rate estimation, demand-side platform, real time bidding, cube factorization model.

1 Introduction

In the real time bidding (RTB) display advertising ecosystem, ad exchanges aggregate ad impressions from multiple publishers and send them to several demand-side platforms (DSPs) via real time auction. Receiving a bid request, each DSP needs to use bidding algorithms to determine whether to bid the ad impression and search for an optimal bid price for each impression. This bid price must be not higher than the expected cost-per-impression(eCPM) which is equal to the click-through-rate (CTR) for the impression multiplied by the cost-per-click (CPC), or the conversion rate (CVR) multiplied by the cost-per-action (CPA) [1,2]. If a CPC or CPA goal is fixed in advance, the eCPM directly depends on how well the CTR or CVR can be estimated. Due to the difficulty of tracking the conversion actions of audience, CPC is nowadays prevalent cost model how advertiser pays DSP. Therefore, we mainly focus on the approach of CTR estimation for ad impressions used in DSPs.

We consider CTR prediction problem as a recommendation problem that ads need to be recommended for users. Regularized matrix factorization models are known generate high quality rating predictions for recommender systems [3,4]. However,

C.K. Loo et al. (Eds.): ICONIP 2014, Part III, LNCS 8836, pp. 17–24, 2014.
© Springer International Publishing Switzerland 2014

matrix factorization models the bilateral interaction between users and items, that is to say, though three types of attributes are available respectively according to a user, an ad and a publisher in our problem, they would be divided into two groups on *x-axis* and *y-axis* respectively [4]. Without loss of generality, it is supposed that one group is composed of all attributes from a user and the other group is composed of those from an ad or a publisher. So the interactions between the ad and the publisher inside the second group will not to be learned. However the CTR estimation in RTB is a tripartite interaction among pages, users and ads. Because the topic of the page p clicked by the user u reveals the user's intention, whether the user u will click the ad a on the page p is influenced not only by how well the content of the ad a conform the preference of the user u but also by what extent the product in the ad a matches the topic of the page p. For example, if the page p is a web page containing certain expertise, the ads recommending professional publications with respect to that expertise may be more appropriate to be impressed than game ads. In order to learn this tripartite interaction, we propose an improved cube factorization model based on high-order Singular Value Decomposition(HOSVD) to extends two dimensional matrix factorization model containing users and ads to three dimensional cube factorization covering users, ads and publishers. This model learns the interaction among users, ads and pages and outperforms the matrix factorization in addressing CTR estimation in our experiments.

The contribution of this paper is two-fold:

- We address the issue of click-through rate prediction for DSP by introducing improved cube factorization model based on high-order SVD. Our model shows its superior performance in handling sparse data than matrix factorization. Furthermore, it also presents better scalability than matrix factorization.
- We conduct various experiments on large-scale real-world bidding log data to evaluate our model and algorithm. Our results show that our improved cube factorization model is a highly promising direction for CTR estimation for DSP.

2 Related Work

There are a number of published studies on click-through rate prediction for search advertising or web search [5,6,7]. Due to many new challenges different from previous application situations, such as more seriously sparse data, more types of ad slot and more complex possibility of user action etc., it is hard to directly apply these approaches to solve our problem. B. Kanagal etc. [8] propose a novel focused matrix factorization model which learns users' preferences towards the specific campaign products for audience selection in display advertising. However, similar to recommendation, only the relevance of the user preference and the ad campaign need to be considered in audience selection. While for real-time CTR prediction in RTB, the user, the ad and the context of publishing are all factors deservedly taken into account. Jinlong Wu etc. [9] present cube factorization model(CF) based on high-order SVD to transform click-through rate prediction problem in personalization web search into rating prediction issue, and verify its performance on artificial datasets. Due to different context of RTB from web search mentioned above, cube factorization model is hardly applied to directly handle click-through rate prediction problem for DSP. Therefore, we perform adaptation on the cube factorization model in order to make it applicable for our problem and improve its performance on real-world datasets.

Few published literature adopts the method related to factorization models to handle our problem. Therefore, we adopt the similar method to the solution proposed by Tianqi Chen etc. [4] as the baseline. Tianqi Chen etc. [4] combine feature-based factorization models and achieve first place in Track1 of KDD Cup 2012.

3 CTR Prediction for DSP with Improved Cube Factorization Model

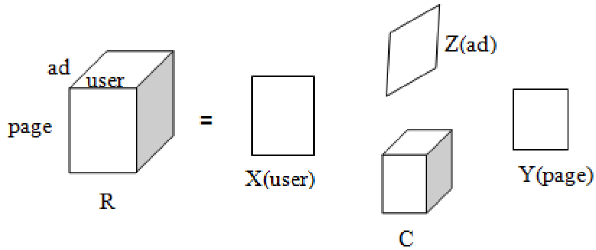

Fig. 1. Third-Order HOSVD

3.1 Problem Setup and Formulation

A bid request that an ad exchange sends to a DSP is denoted by $bid = \{user : u, page : p\}$ which indicates that the user u clicks the page p. The DSP has an ads set $A = \{a_1, a_2, \ldots, a_n\}$ whose member needs to be displayed. These data are all aggregated in the DSP side when a bid request arrived. The goal of the DSP bidding algorithm is to determine which ad in A has the highest probability of being clicked by the user u on the page p. The random variable X is used as the notation of click event outcome, and X equals to 1 if the user u click the ad a_k on page p, -1 if not. Mathematically, it is formulated as:

$$a^* = \arg\max_{k=1,\ldots,n} prob(X = 1 | u, p, a_k) \tag{1}$$

In which, a^* is the most optimal ad for bid. A bid price will be calculated according to the CTR estimation of a^* and submitted to the ad exchange for bidding.

We regard the problem as a cube complement issue with users, publishers and ads on x, y and z axes respectively. Accordingly, the value of the element (x, y, z) represents the quantity $prob(X = 1 | u, p, a)$. Our objective is to estimate different quantities $prob$ according to different triples (u, p, a).

3.2 High-order SVD Based Factor Model

Lathauwer etc. [10] extend two-dimensional matrix Singular Value Decomposition (SVD) to high-order tensor and obtain high-order SVD (HOSVD). According to HOSVD, if given N = 3, then a tensor R can be factored into $R = C \times_1 X \times_2 Y \times_3 Z$ (Fig. 1). Matrix factorization closely related to SVD is used to solve matrix filling problem. Similarly, we apply third-order SVD-based factor model to address three-dimensional cube complement problem. Specifically, it is supposed that *x-axis, y-axis and z-axis* of R represent users, pages and ads respectively(Fig. 1), then according to HOSVD, the value of element(x, y, z) indicates the probability $\hat{r}_{upa}$ of the user u will click the ad a on the page p and can be estimated by the equation as follow:

$$\hat{r}_{upa} = \sum_{l=1}^{L} u_l \sum_{m=1}^{M} p_m \sum_{n=1}^{N} a_n c_{lmn} \tag{2}$$

Where, parameters L, M, and N are numbers of latent factors respectively corresponding to u, p and a.

3.3 Estimating CTR with Improved Cube Factorization Model for DSP

In order to alleviate sparsity of the click event of the user, the tags assigned to the user are incorporate into our model to represent the user. Since tags are composed of three different types of user attributes including gender, personal follows and purchase behaviors, it is good not to normalize the coefficients of tags in our experiments. Details are shown in equation (3) where $T(u)$ is the set of tags the user u has.

$$u = \sum_{t \in T(u)} x_t \tag{3}$$

One benefit of the factorization model is its flexibility in dealing with various data aspects. However, much of the observed variation in click events is due to effects associated with users, publishers or ads, known as biases or intercepts, independent of any interactions [9]. For example, some users show higher tendency in clicking ads than others, and some ads also receive more clicks than others. Therefore, a first-order approximation of the bias involved in r_{upa} is as follows:

$$b = b_u + b_p + b_a \tag{4}$$

$$b_u = \sum_{t \in T(u)} b_t \quad or \quad b_u = \frac{\sum_{t \in T(u)} b_t}{|T(u)|} \tag{5}$$

The notation b denotes the bias involved in r_{upa}. The notations b_u, b_a and b_p involved in b indicate the observed deviations of the user u, the ad a and the publisher p respectively. We examine two ways to combine tag bias as user bias. Details are shown in equation (5).

For the sake of enhancing efficiency for training, we borrow the way how matrix factorization approach is developed from SVD via absorbing the singular value matrix and deal with the tensor C in the similar means. Our experimental results show that extremely similar performance is achieved whether or not the tensor C is considered. Final estimation formulation with bias extend is as follows:

$$\hat{r}_{upa} = \sum_{l=1}^{L}\left(\sum_{t\in T(u)} x_{tl}\right)\sum_{m=1}^{M} p_m \sum_{n=1}^{N} a_n + b_u + b_p + b_a \tag{6}$$

The parameters u, p, a, b_u, b_p, b_a are learned by minimizing the squared error function of train set as follows:

$$L = \frac{1}{2}\sum_{(u,p,a)\in S}\left[\left(r_{upa} - \hat{r}_{upa}\right)^2 + \lambda\left(\sum_{t\in T(u)} \|x_t\|^2 + \|p\|^2 + \|a\|^2 + b_u{}^2 + b_p{}^2 + b_a{}^2\right)\right] \tag{7}$$

Where, S is the set of samples for training, λ is a regularization coefficient.

4 Experimental Evaluation

4.1 Experimental Setup

To evaluate our proposed models, we use the second season log data of Global bidding algorithm competition released by the DSP company iPinYou[1] recently. The train set contains bidding, impression, click, and conversion logs collected from eighteen advertising campaigns during seven days. A set of bidding logs from the following three days is used for offline testing purpose. We split the training dataset into two parts according to the impression date and use the last two days' data as validation set. There are totally 14,758,859 impression records and 11,117 click records in the whole dataset which contains 18 advertising campaigns, 74 ad creatives, 12,456,794 users, 45 user tags, 28505 domains. In order to alleviate the sparseness of the click sample in train set, we choose the combination of the tags the user u possesses to represent the user. The domain of the web page is used to characterize the publisher p. Finally, the ad campaign is selected to stand for the ad a.

To verify the effectiveness of our approach, we use feature-based matrix factorization approach [3], [4] as the baseline, which estimates the CTR as equation (8) and is denoted by the notation FMF. In equation (8), $T(u)$ is the set of tags the user u possesses, and the sum vector of latent factors of the tags the user u have is adopted to describe the user u.

$$\hat{r}_{u,a,p} = \left(\sum_{t\in T(u)} x_t\right)(a+p) + \frac{\sum_{t\in T(u)} b_t}{|T(u)|} + b_a + b_p \tag{8}$$

[1] `http://contest.ipinyou.com/`

We employ the area under the ROC curve (AUC) to compare our model with the baseline. AUC is a commonly used metric for testing the quality of CTR prediction.

To update the model, we conduct stochastic gradient descent training. The number of iterations for SGD inference process is set as 100. We also design experiments to select the appropriate number of the factors.

4.2 Experimental Results and Discussions

Impact of the Number of Factors.
Firstly, we investigate the influence of the number of factors *L*, *M* and *N* on our models, because the time cost of the training algorithm is directly associated with these parameters. Fig. 2 presents the results where horizontal axis is the training set size from one day to five days. As shown in the figure, our model shows relatively stable performance on diverse number of factors from 2 to 10. The number 4 in the horizontal axis means: L=M=N=4. In inference algorithm, the time complexity of calculating e_{upa} or updating u, p and a is $O(|S|LMN)$. That is to say, an appropriate choice of this parameter such as 4 can achieve an optimal balance between better prediction quality and less training time.

Coincidentally, when the number of factors is 4, the performance of the comparing algorithm (FMF) is also optimized. So, unless otherwise specified, four factors are used in the following experiments.

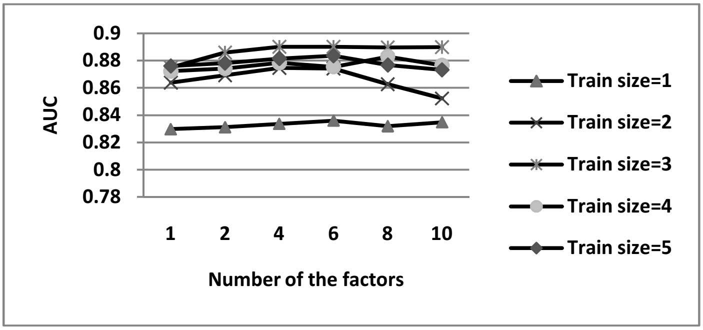

Fig. 2. Impact of the Number of Factors

Combination Approach of Tag Bias for Users.

Table 1. Impact of Combination Approach of Bias for Tags

Train Size(days)	Averaging	Summing
1	0.8336	0.7689
2	0.8746	0.7807
3	0.89	0.7773
4	0.8784	0.74
5	0.8815	0.7298

Since the tags are combined to express the user, the bias according to the tags also needs to be combined. We examine two combination solutions: averaging and summing. The results are shown in table 1 and the former outperforms the later over diverse size of datasets. That means the sum of the bias of all tags may be so strong and causes overfitting. Unless otherwise specified, we use the averaging solution to represent user bias in the following experiments.

Prediction Quality.
Finally, we compare the AUC quality of our improved cube factorization model to the baseline system. Fig.3 shows the results of different methods on the data set. Our model not only presents better robustness on diverse size of train dataset than the baseline but also outperforms the baseline approach. This figure demonstrates that it is more reasonable to consider the CTR estimation for DSP as a cube complement problem than two-dimensional matrix fill problem. Because our model properly considers the information interaction among three types of objects: users, ads and publishers. Therefore, the improved cube factorization model shows its superior ability in addressing this problem.

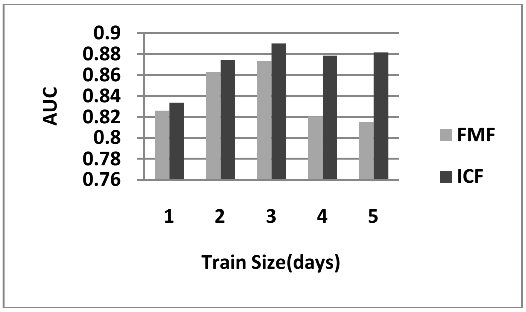

Fig. 3. Prediction Quality of Different Model

In addition, almost all experimental results show that the estimations based on the last three days' history behavior data of users are the most effective results. This reveals that short-term interests of users conduct a greater impact on our problem than long-term interests.

5 Conclusions

In this paper, we focus on the CTR prediction problem in RTB for DSP. We propose an improved cube factorization model to address this issue. In order to alleviate the sparsity of positive samples, the tags of the audience are combined to characterize users' preference. Biases respectively according to users, ads and publishers are also added to the final estimation to model the first-order approximations. For efficiency's sake, the tensor C is absorbed by other parameters without loss of effectiveness. Compared to matrix factorization, our model has superior ability in modeling the interaction among three-dimensional objects: users, ads and publishers, and also outperforms the matrix factorization in real-world data set. Furthermore, our model also shows relatively stable performance both on diverse number of factors and on different size of train set.

Acknowledgments. This work is supported by the National Natural Science Foundation of China (61100094 & 61300114), Specialized Research Fund for the Doctoral Program of Higher Education (20132302120047), and Key Basic Research Foundation of Shenzhen (JC201005260118A).

References

1. Lee, K., Orten, B.B., Dasdan, A., et al.: Estimating conversion rate in display advertising from past performance data: U.S. Patent Application 13/584, 545[P] (2012)
2. Chen, Y., Berkhin, P., Anderson, B., Devanur, N.R.: Real-time bidding algorithms for performance-based display ad allocation. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1307–1315. ACM (August 2011)
3. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. Computer 42(8), 30–37 (2009)
4. Chen, T., Tang, L., Liu, Q., Yang, D., Xie, S., Cao, X., Wu, C., Yao, E., Liu, Z., Jiang, Z.: Combining Factorization Model and Additive Forest for Collaborative Followee Recommendation, KDD CUP (2012)
5. Graepel, T., Candela, J.Q., Borchert, T., Herbrich, R.: Web-scale bayesian click-through rate prediction for sponsored search advertising in microsoft's Bing search engine. In: International Conf. on Machine Learning (2010)
6. Chen, Y., Yan, T.W.: Position-normalized click prediction in search advertising. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 795–803. ACM (2012)
7. Wang, T., Bian, J., Liu, S., Zhang, Y., Liu, T.-Y.: Psychological advertising: exploring user psychology for click prediction in sponsored search. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 563–571. ACM (2013)
8. Kanagal, B., Ahmed, A., Pandey, S., Josifovski, V., Garcia-Pueyo, L., Yuan, J.: Focused matrix factorization for audience selection in display advertising. In: Data Engineering (ICDE), pp. 386–397 (April 2013)
9. Wu, J.: Collaborative Filtering On the Netix Prize Dataset, Peking University doctoral dissertation, pp. 87–104 (May 2010)
10. Lathauwer, L.D., Moor, B.D., Vandewalle, J.: A multilinear singular value decomposition. SIAM J. Matrix Anal. Appl. 21(4), 1253–1278 (2000)