Transfer Learning for Behavioral Targeting

Tianqi Chen[1], Jun Yan[2], Guirong Xue[1], Zheng Chen[2]

[1]Computer Science and Engineering
Shanghai Jiao Tong University
Shanghai, 200240, China
{tqchen,grxue}@apex.sjtu.edu.cn

[2]Microsoft Research Asia
No.49 Zhichun Road, Haidian
Beijing, 100190, China
{junyan,zhengc}@microsoft.com

ABSTRACT

Recently, Behavioral Targeting (BT) is attracting much attention from both industry and academia due to its rapid growth in online advertising market. Though a basic assumption of BT, which is, the users who share similar Web browsing behaviors will have similar preference over ads, has been empirically verified, we argue that the users' ad click preference and Web browsing behavior are not reflecting the same user intent though they are correlated. In this paper, we propose to formulate BT as a transfer learning problem. We treat the users' preference over ads and Web browsing behaviors as two different user behavioral domains and propose to utilize transfer learning strategy across these two user behavioral domains to segment users for BT ads delivery. We show that some classical BT solutions could be formulated in transfer learning view. As an example, we propose to leverage translated learning, which is a recent proposed transfer learning algorithm, to benefit the BT ads delivery. Experimental results on real ad click data show that, BT user segmentation by the approach of transfer learning can outperform the classical user segmentation strategies for larger than 20% in terms of smoothed ad Click Through Rate(CTR).

Categories and Subject Descriptors

H.4.m [**Information Systems**]: Miscellaneous

General Terms

Algorithms

Keywords

Behavioral Targeting, User Segmentation, Transfer Learning

1. INTRODUCTION

Behavioral Targeting for online ads is attracting increasingly attention due to its huge market potentials. It aims to utilize users' web browsing behaviors to segment users and then deliver targeted ads to each user segment. A basic assumption of BT is that users who have similar Web-browsing behaviors will also have similar preference over ads[3]. However, we argue that users' Web browsing behavior doesn't exactly equals preference over ads. We propose to consider users' ad click preference as target domain we want to study and Web browsing behaviors as auxiliary domain to help our understanding in target domain. We formulate the BT problem as a problem of transfer learning.

In this extended abstract, we first show that the classical BT solution is a simple case of transfer learning. In addition, we propose to leverage translated learning, which is a recently proposed transfer learning algorithm, for BT problem.

Experimental results show that through considering BT as a transfer learning problem, classical transfer learning algorithms such as translated learning can outperform the classical BT solution in terms of ad CTR. The goal of this extended abstract is not to propose a novel algorithm for BT user segmentation, we aim to show that through considering BT as transfer learning problem, BT ads utility has big potential to be further improved.

2. LIMITATION OF CLASSICAL BT

As introduced in [3], in classical BT ads delivery solution, we firstly segment users according to user behaviors. Then ads are delivered to related segments. The user segmentation plays a key role in BT. The goal of user segmentation is to segment users who share similar preference over ads into the same group. In this work, we only take queries as example user Web browsing behavior. We denote U^{ad} and U^{query} as ad click and query matrix of the users. The entries of the matrix are defined by

$$U_{ij}^{ad} = \textit{number of times user i click ad j}$$

$$U_{ij}^{query} = \textit{number of times user i queries query j}$$

A natural way to do user segmentation is to group users by U^{ad}. However, users' ad click is extremely sparse in the real world application since many users never click ads. Classical BT makes a tradeoff between sparseness and information quality by directly use U^{query} to represent users for user segmentation.

Although users' web-browsing behavior and preference over ads have strong correlation, they don't reflect the same user intent. If a user input a query "holiday", maybe she is planning her vocation, and we can recommend ad about travelling site, hotel information etc. On the other hand, if a user input a query "bing.com", she might only want to find the URL to the bing.com. This example tells us that different queries have different level of relatedness to users' preference over ads. Directly using U^{query} for user representation may limit the quality of user segmentation.

3. TRANSFER LEARNING FOR BT

In order to overcome the limitations of classical BT. We treat users' web-browsing behavior and ad preference as two domains. Transfer learning is a class of learning algorithms using knowledge from auxiliary domain to help tasks in target domain. We propose to use transfer learning to transfer knowledge from web-browsing behavior to help segment users in the ad click domain. A common idea in transfer learning is to find a better representation. Auxiliary domain features are mapped to new feature matrix U^{new} . The new representation bridges the two domains and enables auxiliary data to help target domain.

WWW 2010, April 26–30, 2010, Raleigh, North Carolina, USA.
ACM 978-1-60558-799-8/10/04.

3.1 Classical BT in Transfer Setting

Classical BT directly uses U^{query} to segment users. This can be viewed as a special case of transfer learning. Identity mapping I is used as feature mapping function, with the mapping procedure

$$U^{new} = U^{query} I$$

Although this method is straight forward, it does make use of the information from auxiliary domain and "transfers" knowledge by directly use the data from the auxiliary domain. Data from auxiliary domain help solve the sparseness problem in the target domain and make segmentation over all users possible.

3.2 Query Ad Translation

Translated Learning[2] is a method to "translate" the auxiliary domain features to target domain. It assumes target domain features are generated by auxiliary domain feature through a markov chain. It uses a transaction matrix to translate the auxiliary domain's features to target domain.

We propose to adopt the similar idea to "translate" user query behaviors to their ad preference domain. We define $R(a|q)$ as the relatedness between query q and ads a. For queries that are not so related with ads. $R(a|q)$ will be low for all a. For queries that strongly relate to ads, $R(a|q)$ will take high value for the a that relates to q. We define translation matrix T as $T_{ij} = R(a_j|q_i)$. Mapping procedure can be described by following equation.

$$U^{new} = U^{query} T$$

We estimate $R(a|q)$ by the counts over the log history by

$$R(a|q) = \frac{number\ of\ times\ a\ displayed\ in\ q}{number\ of\ times\ of\ q\ is\ used}$$

Intuitively, mapping queries to related ads have two advantages: (1) Reducing the effect of ad irrelevant queries. (2) Mapping the similar queries into same ads. We want to point out that our translation matrix is different from the original solution in Translated Learning [2]. The translation matrix R is not a markov transaction matrix since $\sum_a R(a|q)$ doesn't necessary equal 1. This is because not all queries are equally important to ads. For the ad irrelevant queries, $\sum_a R(a|q)$ will be extremely low.

4. EXPERIMENT

We use two weeks' search log of a commonly used commercial search engine for experiments. The dataset contains 524,760 users and 1,098,400 ads. The data is split by week. The first week's data is used to segment users. The second week's data is used to evaluate the quality of the segments.

Ad click through rate (CTR) is a common evaluation for BT. However, test data is also sparse and evaluation may be inaccurate for the ads with small number of views. We adopt similar idea in [1] to use a smoothed group ad CTR to reduce the effect of small sample size. The CTR of ad a and group g is

$$CTR_{smooth}(a \mid g) = \frac{\sum_{u \in g} I_{click}(u,a) + \alpha|g|}{\sum_{u \in g} I_{view}(u,a) + \beta|g|}$$

$I_{click}(u,a)$, $I_{view}(u,a)$ are indicators of user u clicks/views ads a. $|g|$ is the group size. α and β are smooth factors. We set α / β to be the global CTR of all the ads. We define accept set by

$$G_{accept}(\lambda) = \{ <g,a> \mid CTR_{smooth}(a|g) > \lambda \}$$

CTR of accept set is defined as the number of clicks over number of views in the set. The recall of the accept set is the number of clicks in the set over number of all the clicks. We set λ from 1 to 0 to get the CTR-recall curve to evaluate the quality of segments.

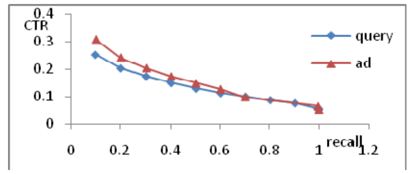

Figure 1. Smoothed CTR-Recall of segmentation by query and ad click over users who click ads.

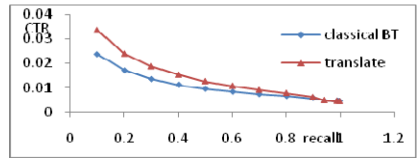

Figure 2. Smoothed CTR-Recall of the methods

Intuitively, we know query behavior and ad click are different. We try to show the difference empirically by experiment. We sample out users who click more than 2 ads. We segment these users by query and ad click separately. We compare the quality of segments in Figure 1. The result shows when ad click information is enough, ad click is better than query for segmentation, and confirms the difference between two domains.

We then compare the classical BT method and translated learning over all the users. Evaluation result is shown in Figure 2. From Figure 2, we can find that the translated learning performs better than classical BT. The results show that by putting BT in a transfer learning setting and use transfer learning method, we have a big room to improve the performance of BT.

5. CONCLUSION

In this paper, we formulate classical BT approach in a transfer learning view. We also adopted a novel transfer learning approach for BT in the same view. We show by experiments that we can improve the performance of BT by using transfer learning.

6. REFERENCES

[1] Ye Chen, Dmitry Pavlov, and John F. Canny. Large-scale behavioral targeting. In *KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 209–218, New York, NY, USA, 2009. ACM.

[2] Wenyuan Dai, Yuqiang Chen, Gui-Rong Xue, Qiang Yang, and Yong Yu. Translated learning: Transfer learning across different feature spaces. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems 21*, pages 353–360. 2009.

[3] Jun Yan, Ning Liu, Gang Wang, Wen Zhang, Yun Jiang, and Zheng Chen. How much can behavioral targeting help online advertising? In *WWW '09: Proceedings of the 18th international conference on World wide web*, pages 261–270, New York, NY, USA, 2009. ACM.