Recommendation for Advertising Messages on Mobile Devices

Chih-Chun Chan
Dept. of Electrical Engineering
National Taiwan University
Taipei, Taiwan
ccchan@arbor.ee.ntu.edu.tw

Yu-Chieh Lin
Research Ctr. for Information
Technology Innovation
Academia Sinica
Taipei, Taiwan
yuccalin@arbor.ee.ntu.edu.tw

Ming-Syan Chen
Research Ctr. for Information
Technology Innovation
Academia Sinica
Taipei, Taiwan
mschen@citi.sinica.edu.tw

ABSTRACT

Mobile devices, especially smart phones, have been popular in recent years. With users spending much time on mobile devices, service providers deliver advertising messages to mobile device users and look forward to increasing their revenue. However, delivery of proper advertising messages is challenging since strategies of advertising in TV, SMS, or website may not be applied to the banner-based advertising on mobile devices. In this work, we study how to properly recommend advertising messages for mobile device users. We propose a novel approach which simultaneously considers several important factors: user profile, apps used, and clicking history. We apply experiments on real-world mobile log data, and the results demonstrate the effectiveness of the proposed approach.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*Data Mining*

Keywords

clustering; mobile advertising; recommendation systems

1. INTRODUCTION

The rapid emergence of mobile devices and corresponding applications brings great impact into our daily lives. Advertising via mobile devices becomes an important approach of promotion. Other than the SMS (Short Message Service) advertising or the web-embedded advertising, we focus on the banner-based advertising, where an advertising message is displayed on the screen as a banner along with some app (Mobile Application).

The banner-based mobile advertising mechanism is as follows. Promoters generate the advertising messages. A mobile advertising corporation is paid by promoters to help them deliver the advertising messages to mobile device users. When an app is run on a mobile device, some advertising message is popped out as a banner. If the user is interested in the advertising message, the user might click on the banner to see further information.

To decide which advertising message should be delivered to the user, a naive approach is to deliver advertising message randomly. However, the advertising corporation would like to recommend the advertising messages that the user is most likely to click on. For recommendation systems available, content-based approaches are not applicable since the advertising corporation are not able to look into the content of users' behavior on mobile devices. Collaborative filtering approaches are not applicable either. Briefly, neighborhood-based collaborative filtering techniques [2] [3] [4] recommend items to a user based on other users sharing similar records with the target user. However, it is observed that the total of advertising messages is quite large, while most users have clicked on only few (or even none of) advertising messages. Thus, the collaborative filtering techniques may not perform well for recommending advertising messages due to the sparsity of the clicked advertisement records.

In this work, we propose to recommend advertising messages by simultaneously considering three factors: user profile, apps used, and clicking history.

2. METHODOLOGY

We propose the RAM (Recommendation for Advertising Messages) approach, including the training phase and the recommendation phase. In the training phase, the history mobile log data is used. We define a similarity metric based on the user profile, apps used, and clicking history. We partition the users into different groups based on the defined similarity metric, and we assign corresponding advertising messages to each group. In the recommendation phase, for each user running an app, we decide which group the user is most likely to belong to, and then we deliver the advertising messages corresponding to that group to the user. Details of the RAM approach are introduced as follows.

Based on the history mobile log data, for each user who has clicked any advertising message, we extract the information about the user as three vectors: PF, AP, and CH. The vectors represent the status of user profile, apps used, and clicking history (i.e., the advertising messages that are clicked by the user).[1] Then we define the similarity between

[1]Due to space constraint, the vectors are not described in detail in this paper.

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2577343.

users u and v as

$$Sim(u,v) = \pi_{PF} \times Cos_{PF}(u,v) + \pi_{AP} \times Cos_{AP}(u,v) + \pi_{CH} \times Cos_{CH}(u,v), \quad (1)$$

where $Cos(\cdot)$ represents the cosine similarity, π represents the weight, and $\pi_{PF} + \pi_{AP} + \pi_{CH} = 1.0$.

With the user similarity $Sim(u, v)$ defined, users with high similarities are partitioned into a group. We apply the DB-SCAN algorithm [1] to partition users into groups. The density-based clustering algorithm results in several groups, without requiring a pre-defined number of groups.

To assign the corresponding advertising messages to each group, we propose NCP (Normalized Clicking Proportion) to describe the relationship between a user group G_i and an advertising message AD_j. We define

$$NCP(G_i, AD_j) = \frac{C(G_i, AD_j)/\sum_{\forall y} C(G_i, AD_y)}{\sum_{\forall x} C(G_x, AD_j)/\sum_{\forall x}\sum_{\forall y} C(G_x, AD_y)},$$

where $C(G_i, AD_j)$ represents the sum of the count that each user in group G_i has clicked on advertising message AD_j. If the $NCP(G_i, AD_j)$ value is larger than one, users in group G_i is more interested in advertising message AD_j than all groups' average. This implies that advertising message AD_j could be a good choice for users in group G_i.

After the training phase is performed, we learn several groups of users and the corresponding advertising messages assigned to each group. In the recommendation phase, when a user runs an app, we deliver the proper advertising message by applying the following process. We first decide which group the user is most likely to belong to by extracting information as vectors and calculating the similarity as Eq. 1 between the user and each group. Then we choose one of the corresponding advertising messages assigned to that group and deliver it to the user. It is worthy noting that for a user without any clicked advertisement record, the weight π_{CH} in Eq. 1 is distributed to π_{PF} and π_{AP} to solve the cold-start problem.

3. EXPERIMENTS

The experimental dataset is provided by WAYSTORM, a mobile advertising corporation in Taiwan. The mobile log dataset contains two parts: the impression data and the click data. The impression data records the status when an advertising message (i.e., ad) is displayed to a user, including the device ID, app ID, ad ID, country code, latitude, longitude, and time. The click data records the status when a user click on an advertising message, including the device ID, app ID, ad ID, and time.

The dataset is collected from October to December in the year of 2012. It is observed that the span of a specific advertising message is not long. Thus, we would like to take recent mobile log data into consideration in the training phase. In the experiments, we divide the three-month mobile log data into eleven non-overlapped datasets, where each dataset contains the records in about a week. For each dataset, we use the first few days for the training phase, and we use the last few days for testing the recommendation phase.

Fig. 1 shows the CTR (Click Through Rate) after three approaches (i.e., Baseline, NBCF, and the proposed RAM) are applied to six of the eleven datasets.[2] The CTR is a general metric for evaluating the performance of a recommendation approach, which is the ratio of the number of clicked advertising messages over the number of total advertising messages displayed to users. The *Baseline* is the naive approach to randomly deliver advertising messages, which has been adopted by the advertising corporation. The *NBCF* approach represents the Neighborhood-Based Collaborative Filtering, which recommends the advertising messages that are clicked by other users sharing similar clicked advertisement records with the target user. According to Fig. 1, the proposed RAM approach outperforms both the Baseline and the NBCF. While the NBCF suffers from the sparsity of clicked advertisement records, the proposed RAM approach could recommend proper advertising messages even to the users without any clicked advertisement record by considering the user profile and the apps used.

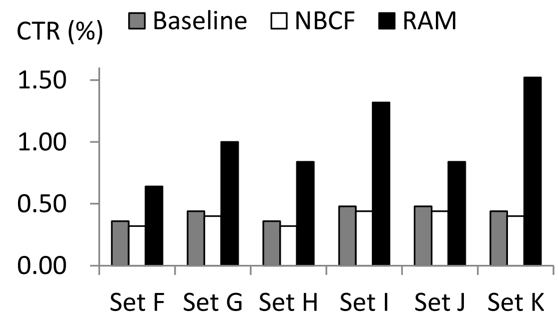

Figure 1: Results: Click Through Rate.

[2]Another five of the eleven datasets are used for setting the values of weights in the similarity metric.

4. CONCLUSION

In this paper, we propose a new approach, RAM, to recommend advertising messages in the banner-based mobile advertising environment. We define the similarity of users by simultaneously considering user profile, apps used, and clicking history. In the training phase, we cluster the users into groups using DBSCAN, a density-based clustering algorithm, and we assign advertising messages to each group based on the NCP (Normalized Clicking Proportion) proposed. In the recommendation phase, we decide which group a user belongs to, and the advertising messages corresponding to the group are recommended to the user. The experiments are performed on a real mobile log dataset, and the results demonstrate the effectiveness of the proposed approach.

5. REFERENCES

[1] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proc. of KDD*, 1996.

[2] Y. Koren. Collaborative filtering with temporal dynamics. In *Proc. of KDD*, 2009.

[3] B. Sarwa, G. Karypis J. Konstan. and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proc. of WWW*, 2001.

[4] Xiaoyuan Su and Taghi M. Khoshgoftaar. A survey of collaborative filtering techniques. In *Advances in Artificial Intelligence*, 2009.