Advertisement-Click Prediction Based on Mobile Big Data from HyXen AdLocus

Yu-Xiang Fei[1,*], Ji-Ying Chen[1], Shih-Hau Fang[1], Yu Tsao[2],

Jen-Wei Huang [3], Bo-Wei Liang [4]

[1]Department of Electrical Engineering and Innovation Center for Big Data and Digital Convergence, Yuan Ze University, Taoyuan, Taiwan.

[2]Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan.

[3]Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan.

[4]HyXen Technology Limitedliability Company, New Taipei City, Taiwan.

Abstract

The popularity of Internet has made advertisement marketing gone virtualized and location-based mobile advertising successful in recent years. Adlocus, an APP developed by HyXen Technology, is one good example to achieve this. This advertising software can tailor to the campaign needs and target users within a diameter of 1 km. However, the question is that is it possible to predict whether the user is willing to click on the advertisement. This paper adopts many ways to analyze how these relations influence in different kinds of mobile advertisement. A comprehensive performance comparison of different models is provided, and the analysis of different factors is also discussed, including click time, advertisement category, language, and mobile phone manufacturers.

Keywords: advertisement-click prediction, mobile devices, audience targeting.

1. Introduction

The popular network has recently increased. When watching advertisements on television, we feel a waste of time, then turning to other programs or doing other things so that advertisements are limited on television. However, in order to increasingly make an impression on users for advertisements, HyXen technology focused on location-based services to solve the problem. This company offered a service called AdLocus which uses positioning technology to provide mobile advertisement.

The commercial value of advertisements on the web depends on whether users click on the advertisements. Many issues, such as users' intent analysis and advertisement selection may affect the click probability of advertisements [1]. Hence, users who installed AdLocus will be provided a large number of advertisements of In-APP. Meanwhile, advertisements of Push which will sent notifications. Among these advertisements, such as sponsored search advertising, contextual advertising, display advertising, and real-time bidding auctions, have all relied heavily on the ability of learned models to predict ad click-through rates (CTR) accurately, quickly, and reliably [2].

The paper use collected Adlocus data to analyze user's behavior and predict how they hit Ads. The data, which has been collected through HyXen for the entire April-2015, containing over one hundred thousand samples per day. There are 21 data attributes, including advertisement category, type of connection, type of device, playing time, advertisers, carrier, user equipment, software version, language, city, and the like. We select seven out of these features to predict the click behaviors [3]. The machine learning algorithms used in this research include support vector machine, decision tree, and RUSBoost. This research can help advertising companies better perform audience targeting, connecting the advertisement with the right audience at the right time and the right place, and even stimulate interactions and have influence on them.

2. Results and Analysis

2.1. Database

There are two main different advertising patterns on AdLocus: 1. In-App (When open the app, Banner style Ads to appear on mobile phone) 2. Push (Send Ads to your phone's notice board). This experiment using the Push Ads, collected 30 days a month of data. Information items closer to six million, which contains a lot of text and character portions. The clicked rate is less than 5%.

2.2. Data Analysis

To predict the click-through rate and what factors most relevant, so analyzed 21 data attributes. The study picks out seven features of data attributes having great influence: advertisement category, type of device, playing time, os system of phone, phone used language, carrier, and city. This paper used the word embedding approach to transform the features to the numerical values. Fig. 1 shows the hourly CTR during 30 days. CTR is the whole day statistics per second. This figure shows that CTR is higher at 3:00 a.m. to 4:00 a.m., and the maximum CTR is 4.73 percent at 3:27. CTR is lower at 5:00 a.m. to 5:30 a.m. The total average is 2.4 percent.
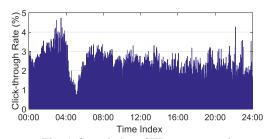

Fig. 1 Cumulative CTR per second

Fig. 2 shows CTR according to different categories of Ads. This figure shows that game/app shows the higher CTR, while travel shows the lowest CTR.
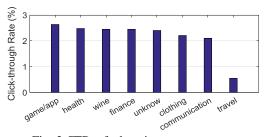

Fig. 2 CTR of advertisement category

Fig. 3 shows CTR according to different mobile phone language, where zh-tw represents traditional Chinese and zh-cn represents simplified Chinese. This figure shows that the highest CTR is unknown. This is because many phones are unknown language labels. The lowest CTR is simplified Chinese.
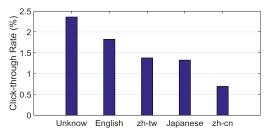

Fig. 3 CTR of mobile phone user's language

Fig. 4 shows CTR according to different mobile phone manufacturers. This figure shows that XiaoMi shows the higher CTR, while Apple shows the lowest CTR.
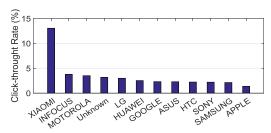

Fig. 4 CTR of mobile phone manufacturers

2.3. Prediction

This experiment adopted 4 types machine learning, including support vector machine (SVM), K-nearest neighbor (KNN), decision tree (DT), and RUSBoost. To predict the click based on training data (April-15 Wednesday) and testing data (April-1 Wednesday). Table 1 shows the results.

Table 1 Prediction accuracy

Algorithm	Accuracy (%)
DT(complex)	94.9
DT(medium)	95.0
KNN(weight)	95.6
KNN(fine)	94.2
SVM	96.4
RUSBoosted	73.5

The table shows SVM provides a higher accuracy rate. However, the CTR of the testing data only is 3.6%, making the baseline accuracy achieves 96.4%.

3. Conclusions

This paper adopts many ways to analyze how these relations influence in different kinds of mobile advertisement. A comprehensive performance comparison of different models is provided, and the analysis of different factors is also discussed, including click time, advertisement category, language, and mobile phone manufacturers. This study uses different methods to predict CTR. Results show that SVM shows the best accuracy.

Acknowledgement

References

[1] C. J. Wang and H. H. Chen, "Learning user behaviors for advertisements click prediction," Proceedings of the 34rd international ACM SIGIR conference on research and development in information retrieval Workshop on Internet Advertising, 2011.

[2] H. B. McMahan, et. al., "Ad click prediction: a view from the trenches," Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2013.

[3] X. He, et al., "Practical lessons from predicting clicks on ads at facebook," Proceedings of the Eighth International Workshop on Data Mining for Online Advertising, ACM, 2014.