Chapter 68
Behavioral Targeting Online Advertising

Jun Yan
Microsoft Research Asia, China

Ning Liu
Microsoft Research Asia, China

Dou Shen
Microsoft Corporation, USA

Zheng Chen
Microsoft Research Asia, China

Teresa Mah
Microsoft Corporation, USA

Ying Li
Microsoft Corporation, USA

ABSTRACT

With the rapid growth of the online advertising market, Behavioral Targeting (BT), which delivers advertisements to users based on understanding of their needs through their behaviors, is attracting more attention. The amount of spend on behaviorally targeted ad spending in the US is projected to reach $4.4 billion in 2012 (Hallerman, 2008). BT is a complex technology, which involves data collection, data mining, audience segmentation, contextual page analysis, predictive modeling and so on. This chapter gives an overview of Behavioral Targeting by introducing the Behavioral Targeting business, followed by classic BT research challenges and solution proposals. We will also point out BT research challenges which are currently under-explored in both industry and academia.

INTRODUCTION

Nowadays, all media is starting to move online. The traditional offline mediums of print (newspapers, magazines), radio, TV are all transpiring into online counterparts such as news and magazine portals, customized radio stations, and on-demand television. Advertising, which is always bonded with media, is moving online accordingly. One of the major advantages of online advertising over traditional advertising is the transparency and tracking capability. Online advertising offers advertisers a unique opportunity to achieve higher ROI (Return on Investment) (Phillips, 2003) and effectiveness than with offline advertising. Specifically, due to the fact that online behavior is relatively easy to track, and a person receiving an advertising message is now identifiable, online advertising allows us to get closer to the promise of true one on one targeting. In any advertising

DOI: 10.4018/978-1-4666-2455-9.ch068

campaign, an advertiser has a target audience they want to reach. For instance, the audience in an advertiser's mind may be "Working Moms". If we try to reach these audiences through an offline medium like TV, the only choice is to present the ads on TV shows that are known to attract Working Moms, or on TV shows that attract people of similar age and gender demographics as that of Working Moms, i.e., "25-45", "Female". Clearly, this type of approximate targeting (or no targeting at all) introduces waste in advertising effectiveness, as many other people could be watching the same TV shows who are not Working Moms. In the online space though, there is much more data available, such as users' individual browsing and querying behaviors (White, & Morris, 2007; Hölscher, & Strube, 2000). From these and other types of behavioral data, we can accurately derive a user's demographics, interests, and attitudes, which allow us to provide advertisers with the exact users they are looking for.

To do behaviorally targeted advertising online, there are several key challenges. First, we need to answer the question of how to represent users, which is used for computing relevance between users and ads. Second, after the users are well understood, how to group users with similar intents (or segment users) for advertisers to target against is also a key question to answer. Last but not least, we have to address the challenge of matching user groups with ads to optimize both relevance and revenue.

This chapter of the book provides a brief overview of the Behavioral Targeting online advertising business, presenting existing solutions to the aforementioned challenges. In terms of user representation, there are generally three types of user profiles - static user profiles such as age and gender of users, behavioral profiles such as the recent Web browsing behaviors of users, and semantic user profiles such as the user intent mined from user behaviors. Details about user representation research will be introduced in the following subsections.

In terms of methodologies for creating Behavioral Targeting user segments, the straightforward option is to create rules that define users in a particular segment. For example, to define a Gamers segment, we can choose the users who have visited some pre-defined gaming *websites* or issued *search queries* like 'xbox', 'ps3', or 'wii' at least *five* times in the *last two days*. One of the pros of this approach is that the rules are simple to understand and easy to explain to advertisers. However, these rules can be subjective depending on the person who creates the rules and how well she/he understands what users in the segment would be interested in. Such a methodology is also not optimized for performance on CTR (click-through rate) or conversion rate, which are the metrics DR (direct-response) advertisers use to measure campaign effectiveness. The quality of a segment defined in such a fashion is bounded by the breadth and depth of the data that publishers/ad networks have to populate segment rules or models (i.e., many publishers do not have access to a broad view of a user's behavior online). Therefore, we foresee that only a few major players will exist in the BT market, since only a few companies in the industry today have the capital and the resources to collect and maintain the amount of data needed to generate accurate and valuable user segments.

Another more advanced methodology commonly used in the marketplace today to create BT segments is look-alike modeling. A look-alike modeling approach uses classification or clustering techniques from data mining and machine learning to automatically optimize against a response variable. It considers both segmentation (group users into segments) and matching (match user segments to ads) simultaneously. Given a set of users who have responded (i.e., clicked or converted) to an online ad and all of their online behaviors, look-alike modeling can then automatically identify the behavioral attributes that distinguish respondents from non-respondents, and generate a model which can run regularly to select users belonging to the segment under consideration.