

Large-scale Behavior Targeting of Users

Ravish Sinha

Yahoo! Inc.
9, M G road
Bangalore
India
ravish@yahoo-inc.com

Abstract

Behavior Targeting is an effort which enhances Yahoo! with the capability to serve the users with the right content and appropriate advertising.

1 About Yahoo!

Yahoo! Inc. provides Internet services to users and businesses through the Yahoo! Network; and a range of tools and marketing solutions for businesses globally. It reaches close to 100 million users everyday, who surf, search, register, subscribe and do various other activities on the Yahoo! network. Users are the greatest asset of Yahoo! and it strives to provide better user experience and engagement.

2 What is Behavior Targetting?

Behavior Targetting allows Yahoo! to use user behavior to optimize sponsored listing products and help in its efforts with personalization. In this talk we will discuss how Yahoo! mines and processes the terabytes of data it collects from user interactions on the network, to generate user specific profiles which help Yahoo! serve them with the right content. We will also discuss the intelligence built on this data to come up with various models which can generate useful profile scores for both registered and unregistered users. The focus of the talk will be to understand the research which helped decide on which models are best suited for generating user profiles.

3 Key Challenges

The key challenges that we faced when developing this system fall in two broad categories - Modeling challenges and System challenges

3.1 Modeling Challenges

The modeling challenges we faced were about ascertaining the importance of various aspects of different user activities.

- Narrowing down on the right data sources to be used for regression models.
- How much various activities contribute towards signifying user interest?
- How much does recency of an activity indicate user interests?
- How does frequency of certain activities indicate user interests?
- Which activity are we trying to optimize using the target function?
- How do we measure success?

3.2 System Challenges

Managing and processing the large volumes of data that Yahoo! collects is a significant system design challenge. We need to process a lot of data to ensure high accuracies and at the same time the processing time needs to be near real time.

- Data size: Processing more than a terabyte of data everyday.
- How much history should we leverage?
- System should scale linearly with size in data (searches on Yahoo! are increasing multiple folds).
- Processing all data and generating interest scores in realistic time.