

Highly Dimensional Problems in Computational Advertising

Andrei Broder

Yahoo! Research
broder@yahoo-inc.com

Abstract. The central problem of Computational Advertising is to find the “best match” between a given user in a given context and a suitable advertisement. The context could be a user entering a query in a search engine (“sponsored search”), a user reading a web page (“content match” and “display ads”), a user interacting with a portable device, and so on. The information about the user can vary from scarily detailed to practically nil. The number of potential advertisements might be in the billions. The number of contexts is unbound. Thus, depending on the definition of “best match” this problem leads to a variety of massive optimization and search problems, with complicated constraints. The solution to these problems provides the scientific and technical underpinnings of the online advertising industry, an industry estimated to surpass 28 billion dollars in US alone in 2011.

An essential aspect of this problem is predicting the impact of an ad on users’ behavior, whether immediate and easily quantifiable (e.g. clicking on ad or buying a product on line) or delayed and harder to measure (e.g. off-line buying or changes in brand perception). To this end, the three components of the problem – users, contexts, and ads – are represented as high dimensional objects and terabytes of data documenting the interactions among them are collected every day. Nevertheless, considering the representation difficulty, the dimensionality of the problem and the rarity of the events of interest, the prediction problem remains a huge challenge.

The goal of this talk is twofold: to present a short introduction to Computational Advertising and survey several high dimensional problems at the core of this emerging scientific discipline.