

IR Paradigms in Computational Advertising

Andrei Z. Broder
Google
1600 Amphitheatre Parkway
Mountain View, CA 94043
broder@acm.org

ABSTRACT

The central problem in the emerging discipline 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 streaming a movie, and so on. In some situations, it is desirable to solve the “dual” optimization problem: rather than find the best ad given a user in a context, the goal is to identify the “best audience”, i.e. the most receptive set of users and/or the most suitable contexts for a given advertising campaign. The information about the user can vary from scarily detailed to practically nil. The number of potential advertisements might be in the billions. Thus, depending on the definition of “best match” and “best audience” these problems lead to a variety of massive optimization problems, with complicated constraints, and challenging data representation and access issues.

In general, the direct problem is solved in two stages: first a rough filtering is used to identify a relatively small set of ads to be considered as potential matches, followed by a more sophisticated secondary ranking where economics considerations take center stage. Historically, the filtering has been conceived as a database selection problem, and was done using simple Boolean formulae, for instance, in sponsored search the filter could be “all ads that provide a specific bid for the present query string or a subset of it”. Similarly for the dual problem (audience definition) for, say, a sports car ad, the filter could be “all males in California, aged 40 or less”.

This “database approach” for the direct problem has been recently supplanted by an “IR approach” based on a similarity search between a carefully constructed query that captures the advertising opportunity and an annotated document corpus that represents the potential ads. Similarly, in the dual problem, the newer approach is to devise an efficient and effective representation of the users, then form a query that represents a prototypical ideal user, and finally find the users most similar to the prototype. The aim of this talk is to discuss the penetration of the IR paradigms in computational advertising and present some research challenges and opportunities in this area of enormous economic importance.

Categories and Subject Descriptors

H.3.5 [Information Storage and Retrieval]: Online Information Services—*Web based services*; H.4.3 [Communications Applications]: Information browsers; I.7.5 [Document and Text Processing]: Document Capture—*Document analysis*

General Terms

Algorithms, Design, Economics, Performance

Keywords

Computational advertising, Similarity search, Audience targeting