



Yoogli Certified Clicks White Paper

Executive Summary

Online marketers use a pay-per-click model to advertise their products to potential customers. Because every click costs them money, they are continually seeking ways of presenting the click opportunity to only the exact people who will purchase their product. Unfortunately, all current keyword-associated ad display techniques include non-responsive people in the target set. Also, people whose sole aim is to cost the advertiser money through click fraud can access these ads with simple keyword searches. Yoogli's sense matching technology enables marketers to dramatically increase the proportion of responsive people in their target audience, while making click fraud much more difficult.

Application – Certified Clicks

A big problem with the ad delivery models now being used in online marketing is that they are delivered to the potential customer via search engines by their association to key words being used in the user's search query. No matter what the user is actually looking for, if a keyword from an advertiser appears in a query, an ad will be presented for that word. Also, even where words are closely related in meaning or semantic domain, an ad will not be presented if the exact right word is not typed in. The result is an ad audience little more focused than a television audience, but with one difference – members of this audience can cost the advertiser money by clicking on the ad, even if interest is low or fraud is intended. Even where interest is genuine, the user must type in the correct word or they will never see the ad that might be of use to them.

Yoogli's certified clicks certify that the user who is clicking on an ad that is being delivered by the search engine actually has a past history of looking at similar domains that relate to the keywords tied to the ad being offered as well as the semantic space of the ad itself. With this history, or personalization profile, the user interaction rises to the level of a "certified click". With certified clicks it is possible to deliver keyword based ads, as well as semantically related ads. In this application, the user does not need to explicitly search for the words tied to the ad, and the ad won't come up just because the user types in a certain word – it is delivered based on an aggregate of past user behavior. The engine used to partition semantic space for advertisers is discussed in a separate white paper.

Semantic Matching

Every online advertisement has a destination URL. Each of these URLs can be semantically classified using a numerical vector that specifies membership in each of some number N of domains, and some number M of significant phrases which occur in the text of the URL and which are semantically aligned with the overall text of the URL. For this paper, a domain can be defined as a partition of some semantic space of interest, where the partition is defined using an enumeration of semantic entities that occur exclusively or primarily in that part of the overall space. A URL can be considered to belong strongly to that domain if it contains a significant number of the semantic entities that make up that domain. Conversely, a URL would be considered to belong weakly or not at all to a domain if it contained few or none of the semantic entities used to define that domain.

Similarly, a person can be assigned a membership vector to the same set of domains simply by examining their search and click history. Since every URL or even entire web site can be assigned a vector, assigning a person a similar vector is accomplished by simply combining the vectors from the sites they visit in some way. This combination can be done in many different ways depending on the application, including (but not limited to) averaging, averaging with hysteresis, weighted combination, or noise-adding averaging. This vector created for a person is called a “personalization vector”.

With a personalization vector and a vector for every web site, implementation of **certified clicks** becomes straightforward. The text of an ad, taken together with a URL that describes the referenced product, can be considered to be a single URL (for which a vector can be generated). The user’s personalization vector and an ad’s vector can be multiplied together using a dot product, a weighted dot product, or some other method, resulting in a match value M . The advertiser can select a matching threshold T , where if an ad has an M that exceeds the T , and the user clicks on the ad, a **certified click** is generated. This **certified click** may motivate greatly increased ad rates for the hosting site, or may be used to measure other things relating to marketing effectiveness.

Implementation

The system starts with zero knowledge about the user. When the user begins browsing, either through a participating site or by initializing the BHO, a new record is created in the database for that user. Every time the user clicks on a link, the contents of the returned page are used to update the picture of the user in the manner summarized above. The user may control the image being built up, or may even create multiple images of him or herself and select which image is presented as the current valid representation of the user’s state. These images, or “aspects”, represent the different roles that a given person might take on during the course of time. For example, a person might have a business aspect, wherein they are looking for things like printers, software, and computers, and a vacation aspect, where they are searching for good tours through Chile and Argentina. Once aspects are created by the user, browsing activity can be automatically assigned to the correct aspect simply by doing a global domain

count match between the current site and all available aspects. Similarly, the user's current aspect or role can be inferred from the sites being browsed.

Updating of aspects or core personalization attributes can happen each time a different site is visited, or on a more delayed basis (in the case where the user is running the BHO). Historical data may be maintained, however, it is not needed as the hysteresis maintains an implicit history. Sites that implement a certified clicks program need only query the central database to resolve the identity of a user's aspect and verify a sufficiently close match. The provided information can be additionally used to customize a coupon for the incoming user, based on ancillary interests the user might have.

Operations

Advertisers who want to use a certified clicks program need to go through a signup phase where they bid on keywords, associated microcontexts, and regions of semantic space. The Yoogli system analyzes the target URL and site for the ad. The result of the analysis is a semantic vector and a list of aligned keywords and key phrases. The advertiser can then elect to use the supplied semantic footprint, or modify it by adding and deleting keywords and phrases. Once that process is complete, the semantic auction engine is brought into play to give current bids on the remaining keywords (for the selected semantic region) as well as a bid on primacy in the selected semantic region for all keywords. The advertiser selects which pieces of the footprint are the most desirable for his particular application, outbids accordingly, and the process is complete. While this process is used for implementing a certified clicks program, it can also be used to enhance results for non-certified clicks. The only difference is that users do not need to qualify by having an aligned aspect in order for the ad to appear. Of course, rates for non-certified clicks are correspondingly lower, but the auction engine provides this two-tier structure at the time of the original bid.

Conclusion

Yoogli's certified clicks program provides a set of tools for advertisers who want to target their ads to users who exhibit certain types of behavior in certain domains. It requires users to "prove" they are interested in products which match, either loosely or exactly, the domains and keywords covered by the advertiser. Users are not even allowed to see ads until these conditions are satisfied. Also, users who exhibit proper semantic behavior do not need to type in specific keywords to see the relevant ads. As a result, the audience for an ad increases in quality, and click fraud becomes much more difficult.