Online advertising – can it be both pleasant and profitable?

BA paper

VU University Amsterdam
Faculty of Sciences
De Boelelaan 1081
1081 HV Amsterdam

Author: Asparuh Hristov

Supervisor: Ger Koole

February 6, 2012

Preface

At the end of the Business Mathematics and Informatics master program, prior to the internship, students have to write their research paper. The goal is to describe a problem in a business process and a solution for it that integrates mathematics and computer science. In this paper, I am going to present the process of delivering online advertisements and different optimization algorithms that can facilitate this procedure.

I would like to thank my supervisor prof. dr. Ger Koole for introducing me to the mathematics behind the online advertising and for his support during the process of writing this paper.

Summary

A big part of the Web browsing experience depends on the online advertising. In the last decade, this market has grown immensely. Despite this fact, there is still little scientific literature related to the online ad serving concept. Therefore, we will present a framework that captures this task of high importance and complexity.

Choosing the right ads for each user greatly influence the revenues for the advertiser and the domain holder as well. In most cases, showing an ad that matches the user's interests will contribute to a higher profit, because the probability that the user will interact with the ad is increased. In addition, displaying an appropriate ad to the visitor increases his satisfaction from browsing the web site. Therefore, in this paper we will introduce some optimization algorithms that can facilitate the decision of which ads to show to the particular visitor.

In order to determine the best user-ad match, one can use information about previous behavior of the user regarding the ad. However, the online advertising market is changing very rapidly and such information is rarely available. Therefore, the task of accurate estimation of whether the user will interact with the ad becomes very challenging. We will present methods that take advantage of the data regarding the ad itself, the visitor, and the web domain. Furthermore, we will identify certain problems that can arise even in a case of a rich ad-user history, and we will introduce possible solutions for them.

Contents

1. Introduction

Nowadays, the Web is economically driven by online advertising. According to the latest Internet Advertising Revenue Report from the Interactive Advertising Bureau (IAB), revenues in the U.S. for the third quarter of 2011 amount to an impressive figure of over $7.8 billion [1]. This accounts to an increase of over 22 percent in comparison to the same quarter in 2010. One of the main advantages of online advertising is that in Internet, in contrast to any other medium (e.g., radio, TV, print, etc.), ads can be dynamically inserted into a web page for each user separately. In addition, possible further interaction between the user and the advertisement happens instantly, which allows easier and more accurate measuring of the ROI. Realizing the potentials of this relatively new form of advertising, most of the companies are dramatically increasing the part of their budget devoted to online ads.

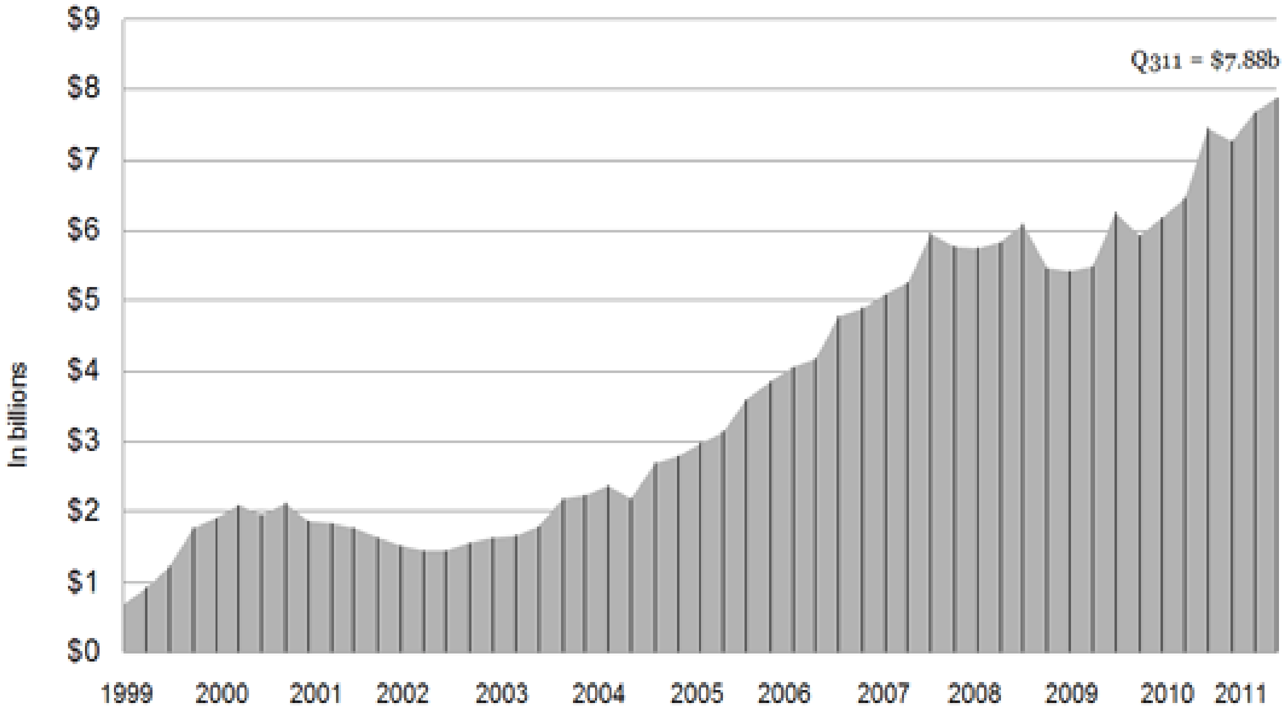

Figure 1. Internet advertising revenues in the U.S.
source: The Interactive Advertising Bureau

There are many types of online advertising, but two of them, namely “search-based ads” and “display ads”, contribute to nearly 75% of the total revenue [1]. Therefore, in this paper we will focus mainly on these forms of Internet advertising.

1.1 Search-based advertising

Most search engines include textual advertising next to their organic search results. The number of different advertisers willing to show their ads for a specific query is a lot larger than the available space and therefore it is essential for search engines to determine which ads to display. The order in which they are shown is also very important because the probability that a user clicks on the ad can decline with more than 90% according to the display position [2]. In this market, advertisers bid for keywords and according to their bid they might be given the opportunity to display their ad. The most common model is cost-per-click (CPC) billing, which means that the publisher (the search engine) is paid for each click on the corresponding ad. Therefore, it is essential to take account of not only the bid price but also the probability of a click.

1.2. Display ads

Display ads are shown on almost all Web pages. These Internet sites have an inventory (space) reserved for advertising. Each time a user accesses the page this inventory should be dynamically filled with display ads. This allows the publisher to present specific ads for each user separately.

In most cases advertisers pay the publishers according to determined CPC or CPA (cost-per-action). In the CPA model, the publisher is paid for each user that “converted”. Conversion may be defined in many different ways with regards to the advertiser’s goal but mainly refers to purchase of a product or filling in a form that commonly leads to a subscription. There is also a third model for payment: cost per impression (CPI or CPM which stands for cost-per-mille). An impression is generated when an ad is being displayed to a user (therefore, if the user has loaded a page with 3 ads on it, she will generate 3 impressions). Hence, the revenue of the publisher does not depend on a good matching between an ad and a user but solely on the number of times the ad is shown (regardless to which user and whether it was clicked). There is not much to be optimized in this case and therefore in the following we will consider only CPC and CPA based models.

2. Model

In this chapter, we will introduce the main parties involved in the online advertising and the interactions between them. As one will see, there are many different models with respect to the numerous types of online ads. Therefore, our goal is to describe a rather general framework capable of comprising most of the models used up to date. In some cases, it will be explicitly noted how the depicted model can be simplified in order to derive another one.

First, we will present the various parties related to the process of serving advertisements throughout the Web. After that, we will outline the actions that have to be taken by those parties in the process. Finally, two alternatives of bidding on publishers' inventory are described.

2.1 Parties involved in the model

In the following, we will describe a model encompassing the various activities with regards to online advertising. There are three main parties in the ad serving process:

- **User** – initialize the process by requesting a web page that has one (or more) zones reserved for advertising;
- **Publisher** – controls the content of the web page. The publisher has to determine the ads that are going to be shown to this specific user. The goal of the publisher is to gain maximum revenue from selling his online inventory reserved for ads (sometimes with constraints regarding user satisfaction);
- **Advertiser** – runs ad campaigns consisting of various ads. The goal of the advertiser is to get the attention of the user and build a relationship with her using the publisher's web page as the channel for placing advertisements.

A central issue in this market is how to match advertisers to publishers. Direct negotiations may work for large companies, but in most of the cases, there are intermediate parties dealing with this problem. Such intermediaries are:

- **Publisher network** – aggregates over publishers. The role of such network is to manage the ad space of many publishers;
- **Ad network** – aggregates over advertisers. Ad networks negotiate with publishers (or directly with publisher networks) in order to ensure appropriate matching of ads and web pages;
- **Ad agency** – a company dedicated to creating, planning and handling ad campaigns for its clients. For example, such agency may design the appearance of

a display ad or consult search-based advertisers about bidding prices for certain terms.

In practice, the services provided by those intermediaries may overlap. There are networks comprising both advertisers and publishers, ad agencies that also play the role of ad networks, etc. Therefore, one ad can take many different paths before being shown to the user. Some of these paths are depicted in figure 2.

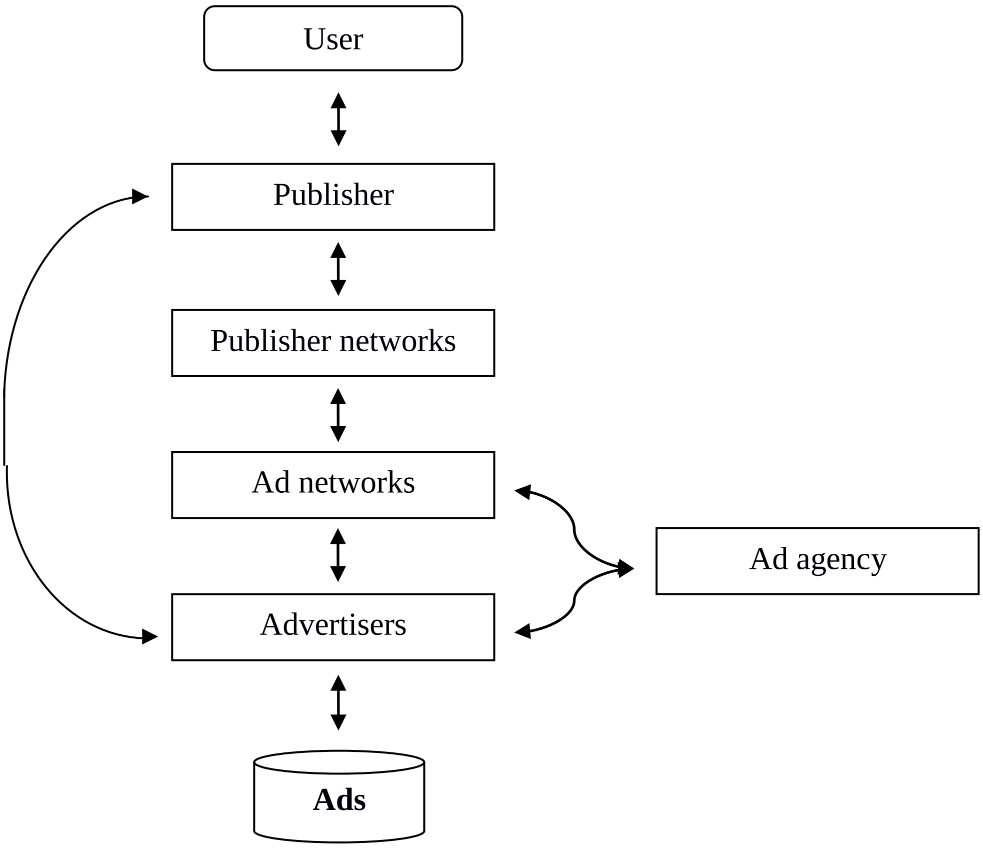

Figure 2. Ad paths

Note that all of the interactions (denoted with arrows) are two-sided because the whole process can be viewed as first “requesting” an ad among one of the paths and then “receiving” it back through the chain.

2.2 Ad exchanges

An emerging way to facilitate selling and buying online ads is through the so-called ad exchanges. Publishers and advertisers (or more generally ad networks) are brought together to a common marketplace in order to ensure elicit prices and to aggregate information [3]. Although there are many exchanges already established for trading various resources (stock, commodities, virtual credits, etc.), ad exchanges appeared recently. Most of the leading ad exchanges like RightMedia (owned by Yahoo) [4], AdECN (owned by Microsoft) [5], and DoubleClick (owned by Google) [6] were acquired in 2007. Such exchanges can be useful for both advertisers and publishers: advertisers have the opportunity to target more precisely their audience due to the larger inventory at the exchange, whereas publishers get liquidity and expect higher selling prices on marketplace with various ad networks and therefore higher competition between advertisers.

Ad exchanges are mainly trading display ads. In search-based advertising, the publisher is only one (the search engine), which results in one-sided marketplace. Furthermore, the ad serving process is simpler in this case and the publisher is supposed to determine which ad is most profitable to be shown on his own.

2.3 Ad serving process

In the following we will present the various stages of an ad serving process involving an ad exchange and *n* ad networks as intermediates:

1) User u requests web page w managed by publisher P. For the sake of simplicity, we will assume that w has to display exactly one ad to u.

2) Publisher P sends to the exchange E the following information: $(I_P(u), I_P(w), c_P)$, where $I_P(u)$ and $I_P(w)$ are details regarding u and w, and c_P is the minimum price P is willing to receive for this impression . We assume that $I_P(w)$ includes possible restrictions with regards to displaying an ad on w, as well as the dimensions of the inventory reserved for ads. Furthermore, we assume that $I_P(u)$ does not include information violating the privacy issues regarding u. In case of a search-based advertising, we suppose that $I_P(u)$ also contains the searched query typed by user u.

3) Exchange E sends to all ad networks $A_1, A_2, \dots, A_n$ the following information: $(I_E(u), I_E(w), c_E)$, where again $I_E(u)$ and $I_E(w)$ are details regarding u and w, and c_P is the minimum price E is willing to receive for this impression. Note that it is not necessary $I_E(u) = I_P(u)$, $I_E(w) = I_P(w)$ or $c_E = c_P$, but it is reasonable to assume that $c_E \geq c_P$.

4) Each ad network A_i sends to E its offer (ad_{A_i}, c_{A_i}), where ad_{A_i} is the ad which A_i wishes to display and c_{A_i} is the maximum price it is willing to pay. We will assume that ad_{A_i} complies with the restrictions opposed by $I_E(w)$ and $c_{A_i} \geq c_E$; otherwise, the offer is instantly rejected by E. Therefore, we can consider that each

ad network sends its offer to E, possibly with a bid $c_{A_i} = 0$ in case that it does not want to participate in the auction.

5) Exchange E determines the winner A_{i^*} and its price c_*. In most cases, this is done via second price auctions [10]. Note also that the relation $c_P \leq c_* \leq c_{A_{i^*}}$ must hold. Furthermore, E sends information $(ad_{A_{i^*}}, c_*)$ to P.

6) Publisher P displays $ad_{A_{i^*}}$ to the user u. Any interaction (click and/or conversion) of u regarding $ad_{A_{i^*}}$ is logged by P and sent to the exchange, which in turn sends it to all ad networks A_i or in some cases only to the winner A_{i^*}.

The above described ad serving process is also illustrated on figure 3. Note that this process can also be applied to search-based advertising by simply omitting the ad exchange E. In this case, as discussed in the previous section, the publisher has to determine the winner among all ad networks A_i.

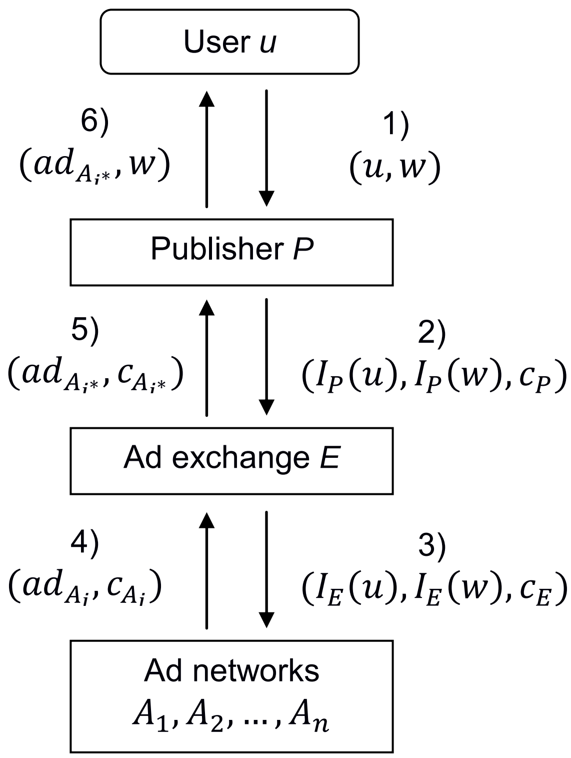

Figure 3. Ad serving process

As one has probably seen, the presented ad serving process allows ad networks to bid for each impression separately. The reason to model the process in such a way is shown in the next section. Furthermore, despite the fact that the impressions are traded between the publisher and the ad networks using CPM prices, the above described process does not limit usage of CPC or CPA prices between the advertisers and the ad networks.

2.4 Static versus Real-time bidding

The first generation of ad exchanges was relying on “static bidding”, which allows an advertiser to purchase a fixed number of impressions from a specific publisher for a fixed price. However, this makes it impossible to differentiate at the individual impression level. In contrast to the static bidding, real-time bidding (RTB) enables advertisers to bid on each impression separately instead of on an entire “bucket”. This means that one can value each individual opportunity in real time and decide whether or not to bid on it. Moreover, advertisers are able to offer higher prices for inventory and audience that they are interested in and lower prices for users and web pages that they do not value. Therefore, nowadays most ad networks buy online display ads via RTB (88% of them planned to do that in 2011 [7]. Among the first ad exchanges that started RTB support are DoubleClick [6], AdBrite [8] and OpenX [9]. According to Google, the inventory sold through real time bidding on their ad exchange DoubleClick [6] raised from 8% in the beginning of 2010 to the staggering 68% in May 2011 [7].
.

3. Model optimality

As discussed in chapter 1, most of the contracts between online advertisers and publishers are based on cost-per-click or cost-per-action model. Therefore, the publisher's revenue significantly depends on the degree of match between the displayed ad and the user. It is reasonable to believe that a better matching results in a higher probability of subsequent interaction between the ad and the visitor (such as clicking or conversion). Furthermore, showing to the user an ad that she is interested in greatly improves her satisfaction. On the other hand, displaying irrelevant ads will annoy the visitor and possibly even chase him away from the web page. Therefore, the task of determining which ads to be shown is essential for maximizing the publisher's revenue in a long-term perspective as well.

Advertisers seek to display their ad to an audience that is prone to undertake certain action with regards to the advertisement. In the marketing field, this is defined as targeting. In most cases, targeting contributes to a significant increase in the ROI.

In conclusion, the proper allocation of the ads being shown to each user leads to benefits for all three main parties: the user, the publisher, and the advertiser. The problem of finding the most relevant ad for a user can be formulated as determining the ad with the highest click through rate (CTR) for this specific user. The CTR for an ad is simply the probability that it will be clicked. For example, if an advertisement was viewed 1000 times and was clicked in 6 out of those 1000 times, we say that the CTR of this ad is 0.06.

3.1 Importance of estimating the CTR

We have seen that optimality for all three main parties can be easily achieved if one knows the actual CTR of a given user regarding each ad. However, in practice clicks are rare events and therefore accurate estimation is challenging. In addition, inaccurate estimation results in less profit. For example, consider again the model described in chapter 2 and let the pricing model between the ad network and the advertiser be CPC. Assume that an ad has CTR of 0.2 and CPC of €10. Suppose that the network has overestimated the CTR to 0.3 and bids €3. This will result in an expected loss of €1 for each impression. On the other hand, assume that the network has underestimated the CTR to 0.1 and bids €1. The expected profit in this case will be €1 for each impression. However, bidding less also means fewer auctions won and therefore smaller audience and fewer impressions. Therefore, ideally the ad network should bid €2 on the exchange.

We will also give an example of the importance of accurate CTR estimation in searched-based ads. Suppose that the pricing model is CPC, i.e., the search engine receives a fixed reward for each click on the corresponding ad. Therefore, the expected revenue for showing an ad is the product of its CTR and CPC. This means that an inaccurate CTR estimation will lead to an inaccurate profit estimation, which in turn will result in a wrong selection of the ads that have to be displayed.

3.2 Algorithms for estimating the CTR

In order to estimate the click through rate of a user regarding a specific ad, one can simply use the binomial maximum likelihood estimation (MLE) – the number of clicks divided by the number of impressions. However, in practice the CTR is relatively low. This leads to high variance in the estimate even for a moderate number of impressions. For example, if the true CTR is 1%, one has to display the ad 6700 times in order to be 90% confident that the estimate is within 2% of the true value. As discussed previously, an inaccurate estimation of the CTR results in a suboptimal selection of the displayed ads and therefore lost revenue and lower traffic for the better performing ads. Thus, one needs algorithm capable to predict more precisely the click probability even when there is scarce history of the behavior of various users regarding the given ad.

The conversion of a user is an even more rear event. With an average cost per action of €50 [11] an ad would require €5000 for observing one hundred actions. Once again, usage of the binomial MLE may lead to significant loses. Fortunately, most algorithms for estimation of the click probability can be easily adapted to efficiently predict the action probability in a CPA pricing model.

The number of different ads and users might be as large as hundreds of thousands. This makes it nearly impossible to gather a moderate number of impressions for each user-ad pair. Thus, most algorithms use some kind of clustering with regards to the users and/or the ads based on various similarities. After the segments are determined, one can estimate the CTR on the segment level, instead of on the individual user and ad level. The segmentation part of the algorithm is supposed to be executed offline due to running-time constraints. The decision of which ad to show should be done in less than 0.1ms [3].

The online advertising market is dynamically changing. Each day new advertisers and publishers enter the market, new ad campaigns are being launched and old ones are being modified. The interests and the behavior of each user change rapidly as well. This means that the offline segmentation should be executed very often (e.g., every half an hour).

One additional challenge for the optimization algorithms is that an underestimation of the CTR may lead to a constant underrating of the corresponding ad. This means that the ad might never be given the opportunity to be shown and therefore the wrong estimation would never be corrected. A possible solution to this problem is to display ads randomly with regards to some part of the impressions. For example, one can decide to show the ad determined by the algorithm in 80% of the cases and a random one in the other 20%. Note that the results from the random part might also be used as a baseline for the model. However, displaying ads at random leads to lower revenue in comparison to an ideally optimized approach. Therefore, there is a trade-off between gaining information and realizing the maximum possible revenue [12].

3.3 Privacy Issues

Without any doubt, identifying the interests and needs of a particular user can greatly assist the decision of which ad to show. However, the gathering of such data raises important issues concerning users' privacy [13], [14]. A possible way to tackle this problem is by allowing users to opt-out of having their private data shared with third-parties. This is exactly the case of storing the so-called "cookies" [15]. Most major web sites uses cookies, but users are able to disable this feature from the internet browser. In the following, we will discuss only algorithms which do not use any data that may violate the privacy issues.

4. Estimating CTR for new ads

Estimating the click-through-rate for ads is a crucial task in maximizing the revenue and improving the user satisfaction. One can greatly benefit from calculating an estimate of the CTR based on the ad's history if this ad is already shown moderate number of times. However, lots of new ads are being created every day in the online advertising market. Therefore, one should be able to estimate the CTR of an ad a priori. The most obvious way to calculate an a priori estimate is to simply take the average over all ads in the system. This technique might be easy and straightforward to implement but it is far from optimal. Our goal is to estimate the performance of an ad based on the CTR of ads similar to the one in question rather than on all ads. The reason for this is our belief that similar ads experience similar CTR.

4.1 Logistic regression

Logistic regression is a generalized linear model [16], [17] used for binomial regression. It is a common technique in statistics and it is being used for predicting the probability of occurrence of an event. This is the reason for choosing exactly this method. In our case, the event is simply whether or not the impression led to a click. In order to predict the probability of a click, we will use the logistic function:

$$f(z) = \frac{1}{1+e^{z}} ,$$

where z is the input and $f(z)$ is the output that takes values between 0 and 1 (note that $f(z)$ gives us the probability of a particular outcome). The variable *z* is known as logit and is defined as:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k ,$$

where β_0 is the "intercept" and β_1 , β_2 ,…, β_k are the "regression coefficients" of $x_1, x_2, \dots, x_k$ correspondingly. In other words, *z* is determined by the values of the various features in the model (x_i is the value of the *i*-th feature) and the corresponding weights (β_i is the weight given to the *i*-th feature). The vector $\boldsymbol{x} = (x_1, x_2, \dots, x_k)$ is also known as the predictor vector. Those variables might be either numerical or categorical, but they have to be independent.

Our goal is to identify a set of features that can be used as predictors for the CTR of an ad. Such predictors might be anything, such as: the number of words in the ad, whether or not a specific word is included in the title, whether it is a banner or a sponsored-search ad, etc. After specifying the set of features, one can learn the corresponding weights by executing the regression on the data that is gathered up to this point (e.g., using R [18]). We assume that there is sufficient history regarding older ads such that these weight coefficients can be estimated relatively accurate.

4.2 Term features

In the following, we will introduce features regarding the words contained in the title of an ad. Let $term$ be the title of the ad in question. Moreover, denote with $S(term)$ the set of all ads with title ad_{title} having at least one word in common with $term$. In addition, let $S_{mn}(term) \in S(term)$ contains those ads such that ad_{title} can be derived from $term$ by removing m words and adding another n (note that $|term| > m$, because $term$ and ad_{title} have common words). That is:

$$S_{mn}(term) = \left\{ ad \quad s.t. \quad \begin{matrix} |ad_{title} \cap term| > 0 \\ term - |ad_{title} \cap term| = m \\ ad_{title} - |ad_{title} \cap term| = n \end{matrix} \right\}$$

For example, if $term = \{buy, new, car\}$: an ad with $ad_{title} = \{rent,\ car\}$ will belong to $S_{21}(term)$; whereas an ad with $ad_{title} = \{buy,\ new,\ jeans,\ shirts,\ jackets\}$ will be an element in $S_{13}(term)$.

Given $S_{mn}(term)$ we can compute the mean CTR for this set :

$$CTR_{mn}(term) = \frac{1}{|S_{mn}(term)|} \sum_{x \in S_{mn}(term)} CTR_x \ ,$$

where CTR_x is the click-through-rate of x. Finally, we will define:

$$f_{mn}(ad) = \frac{\alpha\overline{CTR} + |S_{mn}(term)| * CTR_{mn}(term)}{\alpha + |S_{mn}(term)|} \ ,$$

where $\overline{CTR}$ is the mean CTR for all ads and α determines the strength of the prior, in terms of number of impressions. Based on the results in similar study [19], I believe that features like $f_{mn}(ad)$ and $|S_{mn}(term)|$ can contribute to significant reduction in the CTR estimation error.

4.3 Ad quality features

In the previous section, we have discussed features based on the title of the ad. However, in practice there are ads with similar or even identical titles performing completely different in terms of CTR. Factors such as appearance of the ad and reputation of the advertiser may play an essential role in deciding whether or not to click [20]. Therefore, one may try to capture the ad quality by defining various new features such as:

- Appearance of the ad: Is it a picture (.jpg) or a video (.gif)? Does it contain words such as “buy”, “purchase”, “join”, etc.? Does it contain prices? Are there special symbols or exclamation marks? Is it colorful or “black and white”?
- Reputation of the advertiser: Is this the first ad of the advertiser? How long is the displayed URL (this can be also seen as feature regarding the appearance of the ad, but here we assume that short domain names are more expensive and belong to “well-known” advertisers)?

The above stated features are just examples of possible predictors. Many of them were studied [19] and proven to additionally reduce the estimation error next to the term features.

4.4 Conclusions

Using logistic regression, one can greatly increase the accuracy of the CTR estimation for ads with scarce history. Possible predictor variables might include details about the words included in the ad or the ad quality. However, many of these ad features concern information, which may not be available. Obtaining this information might be challenging task, especially in case of display-ads. Therefore, the above described technique is mainly used for search-based ads, where the search-engine at least knows the bidding terms of each ad.

5. Estimating CTR based on user's features

In the previous chapter, we have seen how we can get a better estimation of the CTR using various features of the ad. In order to predict even more accurately the performance of an ad regarding a user, one has to take into account the behavior of the particular user. As already discussed, the assumption that we have information for each user-ad pair is unrealistic. Therefore, one has to estimate the probability of a click by the ad performance with regards to similar users.

In the following, we will describe how one can differentiate users based on their online history. The amount of data generated by each visitor makes it impossible to store and manage the whole available information in the long term. Moreover, users have multiple interests, which change very rapidly. A related paper [21] shows that one may benefit more if one takes into account only the short term, rather than the long term history. Therefore, we will present two algorithms devoted to capture the short term behavior. We will explicitly differentiate between the most recent information and the older data. The reason for this is our aim to identify the intention of the user to request the particular web page and her interests as well. One can combine the results from both methods by giving them different weights with regards to the overall decision.

5.1 Capturing the visitor's intention

We will present different features that can be used to identify how appropriate an ad is for the current interest of the user. Such feature can be the weather at this moment. For example, an ad for purchasing an umbrella may perform very differently depending on whether it is raining outside or not. Another feature might be the current time. Certain ads may experience their lowest CTR in the night, whereas others their highest.

In case of search-based advertising, one has to serve ads each time a user searches something with specific query. It is reasonable to believe that this query represents very accurately what the user is willing to see. On the other hand, "display ads" are shown on various web sites that in most cases are visited by users with similar intentions. Therefore, in both situations one can use the currently viewed web page as a predictor of the user's interest.

Once a set of visitor's features is identified, the ad performance may be estimate in a way similar to the one described in chapter 4. Again, the corresponding weights of each predictor may be determined by executing logistic regression on a statistical software program [18].

5.2 Capturing previous interests

In this section, we will describe how one can use the browsing history of a user in order to capture her interests. Suppose that there is available information for *m* users and *n* web pages. This data can be stored in a matrix $D \in R^{mxn}$, where entry $d_{i,j}$ corresponds to user *i* and site *j*. The values in D can be assigned in various ways, for example:

- $d_{i,j} = \begin{cases} 1 - if\ i\ has\ visited\ j \\ 0 - otherwise \end{cases}$
- $d_{i,j} = \#\ times\ i\ visited\ j$
- $d_{i,j} = (\log(\#\ times\ i\ visited\ j) + 1) * \log \frac{1}{\#\ users\ visited\ j}$ (see ref. [21])

Using the matrix *D*, one can identify different groups of visitors with a segmentation algorithm (e.g., k-means [22], CLUTO [23], MinHash [24], etc.). After that, one can estimate the CTR of an ad with regards to each of the groups separately. In order for this prediction to be relatively accurate, we assume that there are enough users (and thus information) in each group.

Finally, we will introduce how this segmentation can facilitate the decision of which ad to show to a particular visitor. Knowing the previously viewed web sites for the user, one can easily determine her “nearest” segment, or in other words – the group of users with most similar interest to those of the visitor. Now, one may use directly the ad CTR for this group specifically. Another way to take advantage of this information is to define the user segments as a feature for the logistic regression. Therefore, the latter method is easier to combine with the previously discussed techniques.

6. Frequency caps

In the previous two chapters, we have presented various algorithms that facilitate the estimation of the CTR in case of a scarce history. However, even when there are enough observations of the actions of a specific user regarding an ad, one should avoid taking simply the average over all impressions of this user. In the following, we will discuss a technique which can greatly improve the overall CTR and increase the revenue of the publisher. Moreover, it contributes to higher user satisfaction.

Studies show that the click through rate of most advertisements is highest on the first impression [25]. It is reasonable to believe that if one user is not interested in a particular ad, she will not click it (or convert), despite being shown the same ad over and over again. In addition, if the impressions are traded on CPI basis, e.g., via an ad exchange, one is not losing only the opportunity to display an appropriate ad to the user, but also money for buying worthless impressions. On the other hand, showing the ad only few times will not create significant impact and will also result in lost revenue. There is a trade-off between efficiency and volume, and the marketers have to decide in what extent to impose the frequency cap. Too low a cap will result in a loss of potential clicks, whereas too high a cap will lead to over-saturation and a waste of money that is spent for acquiring inventory. In conclusion, the optimal frequency level is highly dependent on the advertiser's goal.

6.1 Frequency limits in the CPC model

The greatest impact of introducing frequency caps will be on the users that generate hundreds of impressions on an ad without having even one click on it. Showing the same ad to an inappropriate user so many times will reflect negatively on the precision of the CTR estimation. For example, imagine the extreme case when an ad was shown one time to 4 users and all of them have clicked on it, but also displayed 1000 times to a user without receiving any clicks from her. This will result in a CTR estimation of $4/1004 \approx 0.004$. Now suppose that there is a frequency cap set to 10. The first advantage is that the user with 1000 impressions will be given the opportunity to explore 100 different ads, instead of just the one in question. This will give more information regarding the other 99 ads and may also lead to potential clicks. The second advantage is that the CTR estimation becomes $4/14 \approx 0.285$, which is a much more precise indicator of the performance of the particular ad.

6.2 Frequency limits in the CPA model

One can also greatly benefit from using frequency caps in the CPA pricing model. The precise formulation of what is considered as an action depends on the advertiser, but in most cases it is associated with purchase of a product or some kind of subscription. In both situations there is possibility that conversion of a particular user can happen only once (e.g., subscribing with own email, opening a bank account, buying something “unique”, etc.). Therefore, if the user has already converted, there is no need to show him the ad again. In fact, as discussed in the previous section, generating such impressions is not only pointless, but will also lead to lost revenue. In addition, if there is no frequency capping on this level, the algorithm will promote the corresponding ad to be displayed to the user even more often, driven by the good match that led to the conversion.

Another possible application of limiting the frequency of certain impressions in the CPA model is concerning the click frequency. Normally, clicks are indicating interest, and thus the ad is shown over and over again until the user finally converts. This is also known as retargeting, and it is commonly used in the advertising market. However, there are certain cases when this may be misleading and may result in profit losses. For example, there are online “bots”, especially programmed to generate clicks for a specific ad (sometimes “clicking” thousands of times per day). Also a lot of clicks are unintentional due to a tricky ad positioning on the web page. In such cases, introducing frequency caps may improve the “reliability” of the data and lead to better performance of the system as a whole.

7. Conclusion

We have seen that all three main parties in the online ad serving process (the user, the publisher, and the advertiser) may benefit from a good match between the ad and the visitor. In most cases, it is reasonable to measure the goodness of this match with the click-through rate of the user for the ad. Therefore, we have presented various methods aiming to accomplish more accurate estimation of the CTR in comparison to the simple MLE.

In case of scarce ad-user data, one can make use of the logistic regression with a set of features regarding the ad, the user’s browsing history, and the contextual information about the ad request. In similar studies [19], it was shown that one can significantly decrease the estimation error by introducing predictors concerning the terms in the ad and the ad quality. Therefore, we have presented similar features, along with others capturing the user’s intention of viewing the web page, her interests, the domain, etc.

Finally, we have discussed a method which is able to limit the gathering of useless (or even misleading) information regarding an ad-user pair. However, such a frequency capping may also lead to a loss of potential clicks. The optimal frequency level highly depends on the advertiser’s goal and should be determined according to the specific model.

In conclusion, we have shown different optimization techniques that can be applied to the most of the online ad serving processes used up to date. The algorithms, which were discussed, are not only easy to understand and implement but also very efficient. Further improvement can be achieved by introducing more complicated methods, but this is out of the scope of the paper.

Bibliography

[1] Interactive Advertisement Bureau (IAB) Internet Advertisement Revenue Report (by PricewaterhouseCoopers), November 30, 2011

[2] Did-it, Enquiro, and Eyetools, “Eye Tracking Study”, http://www.enquiro.com/eye-tracking-pr.asp

[3] S. Muthukrishnan, “Ad Exchanges: Research Issues”, Plenary talk at WINE 2009, Google inc.

[4] RightMedia, http://www.rightmedia.com/

[5] AdECN, http://www.adecn.com/

[6] DoubleClick Ad Exchange (AdX) , http://www.doubleclick.com/products/advertisingexchange/

[7] “The Arrival of Real-Time Bidding and What it Means for Media Buyers”, Google Inc. 2011

[8] AdBrite, http://www.adbrite.com/

[9] OpenX, http://www.openx.com/

[10] B. Edelman, M. Ostrovsky, and M. Schwarz, “Internet advertising and the generalized second price auctions: Selling billions of dollars worth of keywords”, American Economic Review, 2007

[11] Jonathan Hochman, “The Historical Cost of Pay-Per-Click (PPC) Advertising” , 2010 , http://www.hochmanconsultants.com/articles/je-hochman-benchmark.shtml

[12] Assaf Zeevi Denis Saure, “Multi-armed Bandits: Applications to Online Advertising”, LNMB Conference 2012

[13] Leenes, R.E., “Do you know me?”, Tilburg University Legal Studies Working Paper, January 2008

[14] Tene, O., “What Google Knows: Privacy and Internet Search Engines” October 2007, Utah Law Review, Forthcoming. Available at SSRN: http://ssrn.com/abstract=1021490 or http://dx.doi.org/10.2139/ssrn.1021490

[15] HTTP Cookie, http://en.wikipedia.org/wiki/HTTP_cookie

[16] Logisitic Regression, http://en.wikipedia.org/wiki/Logistic_regression

[17] K. Bartz, V. Murthi, S. Sebastian, “Logistic Regression and Collaborative Filtering for Sponsored Search Term Recommendation”, In Proceedings of the Second Workshop on Sponsored Search Auctions, 2006

[18] Free software environment for statistical computing and graphics, R, http://www.r-project.org/

[19] M. Richardson, Ewa Dominowska, and R. Ragno, “Predicting Clicks: Estimating the Click-Through Rate for New Ads”, Microsoft 2007

[20] B. Jansen and M. Resnick, “Examining Searcher Perceptions of and Interactions with Sponsored Results”, In Proceedings of the Workshop on Sponsored Search Auctions, 2005.

[21] “How Much Can Behavioral Targeting help Online Advertising”, Microsoft 2009, WWW 2009 MADRID, Track: Internet Monetization / Session: Web Monetization

[22] T. Kanungo, D. Mount, N. Netanyahu, C. Piatko, R. Silverman and A. Wu. “An efficient K-means clustering algorithm: Analysis and implementation”, IEEE Transactions On Pattern Analysis and Machine Intelligence, VOL. 24, NO. 7, JULY 2002

[23] G. Karypis. CLUTO: a software package for clustering high-dimensional data sets. University of Minnesota, Dept. of Computer Science.

[24] P. Indyk and R. Motwani. “Approximate nearest neighbor: towards removing the curse of dimensionality”, Proceedings of the thirtieth annual ACM symposium on Theory of computing, Dallas, p. 604-613, May 24-26, 1998

[25] John Chandler-Pepelnjak and Young-Bean Song, “Optimal Frequency - The Impact of Frequency on Conversion Rates”, Atlas Digital Marketing Technologies, Microsoft Advertising Institute, 2004