Optimizing Display Advertisements Based on Historic User Trails

Neha Gupta, Udayan Khurana, Tak Yeon Lee*
Department of Computer Science
University of Maryland
College Park, MD 20742
neha,udayan,tylee@cs.umd.edu

Sandeep Nawathe
Tumri Inc.
San Mateo, CA
snawathe@tumri.com

ABSTRACT

Effective online display advertising requires a dynamic selection of the advertisement to be displayed when a web page is fetched. As the goal of displaying advertisement is to engage the users and obtain clicks, the advertisement which has the highest probability of click should be displayed. In this paper we address the problem of finding the most suitable display advertisement option for a user given his/her current browsing session. Using this historical browsing session information, we mine the association of different advertisement views, engagements and clicks, and apply Bayesian models to find the likelihood of an advertisement to be clicked given a specific set of events that describe a user session. A major challenge in training the model for optimum precision is the sparsity of click events, hence we propose the use of advertisement engagement as a success event like clicks to train the model more effectively. Our technique significantly outperforms the baseline technique of using prior probabilities for selecting advertisements.

1. INTRODUCTION

Online display advertising is a rapidly growing activity on the internet [1, 3, 5, 4, 2]. When a user requests a web page through his browser, the web server sends the webpage requested which may contain multiple advertisements. Earlier the advertisements were statically assigned to the webpage and same advertisements were displayed for extended periods of time. A newer trend has been the use of dynamic decision making about which advertisement to include. The process of attaching advertisements to the webpage has multiple agencies involved. In this business, the owner of the web server is usually referred to as the publisher, the company whose advertisement is to be placed is called the advertiser and the media agencies play the role of the mediator between the advertiser and the publisher. In the dynamic environment of the internet, an advertisement has to be selected and rendered within a few microseconds but the decision can be highly dynamic changing in real time. The crucial aspect of advertising is the reaction of the user which may be to view it, engage with it by doing a mouse over, or click and go to the webpage (landing page). The decision making process about selecting the advertisements aims to increase the engagement and click rate. The click rates are very low (< 0.01) and can be considered as statistically *rare events*.

The primary decision to select an advertisement from a set provided by an advertiser for a campaign rests with the media agent. The information about user actions is usually available in real time and can be taken into account in the real time decision process which has to be very efficient so that it can be completed within few microseconds. However, understanding the impact of an advertisement on a user is not easy. Studies[15][16] show that not only clicks, but advertisement views also have an impact on overall sales for a brand. In the framework proposed here, we use historic user trails - consisting of advertisements and actions taken on these, to predict the advertisements which have maximum probability of being clicked. Let us introduce the terminologies used in the rest of this paper:

- **Campaign:** A set of advertisements based on a single theme or promoting a new range of products such as Fathers Day campaign, Thanksgiving sale, etc.
- **Creative:** The different display components of a display advertisement like background image, color, slogans, banner, etc. are called the creatives. For example, in Fig. 1 different components of the pictures are different creatives.
- **Recipe**: Each dynamic advertisement is called a recipe. A recipe is a combination of different creatives. Two examples of recipes are given in Fig. 1.
- **Adpod** : An Adpod template is simply a banner advertisement wireframe carved out into key sub-components such as brand logo, call to action, featured product, price, background image, etc. by a knowledge worker. The AdPod template is the *shelf* or *container* that

*All authors contributed equally to the research.

SIGIR 2011 Beijing, China

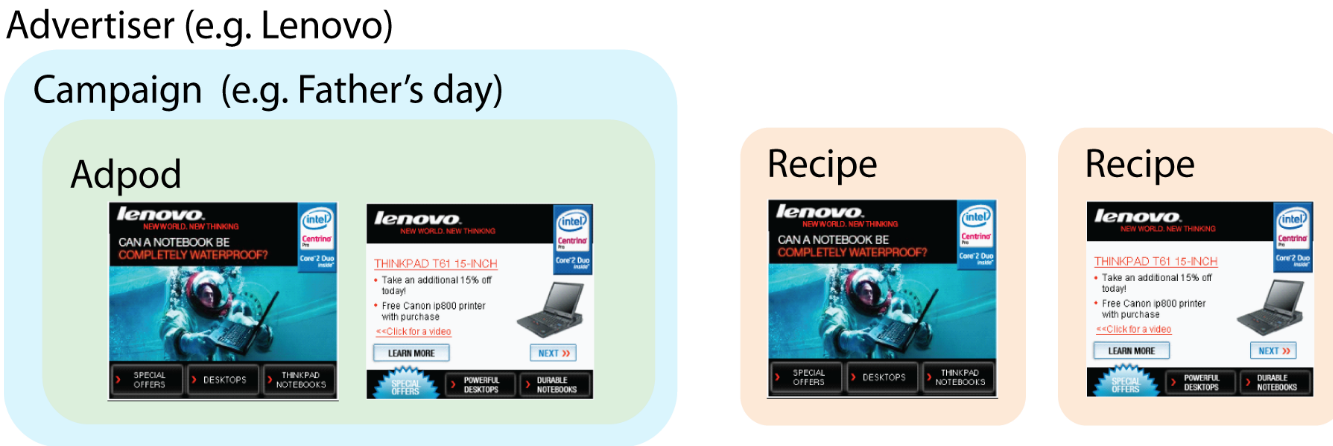

Figure 1: A typical example of a display advertisement and its hierarchical components

houses the various elements from the advertiser display components. A group of multiple recipes made up of a single template is called an adpod.

- **Impressions & Click Through Rate(CTR)**: The number of views a particular page or advertisement receives are called impressions. CTR is the number of clicks divided by the total impressions.
- **Engagement**: Engagement is a type of user interaction such as hovering mouse cursor over an ad, or interacting it in any way other than a click to redirect.
- **Session** is a set of actions done on one or more recipes for a given user in a fixed window of time.
- **Browsing Context** represents the state of the user with respect to advertisements that have been viewed/ engaged or clicked by him/her during a session.

There exists a hierarchy in the display advertising domain which is - each advertiser runs a number of campaigns and each campaign consists of a number of adpods. Each adpod consists of multiple recipes which are optimized in order to get the maximum performance of a campaign.

In this paper, we present a prediction model which uses past information on a campaign's usage data to predict the choice of the most suitable advertisement for any user in a given context. More precisely, we aim to identify the choice of the user by the display advertisements already seen, engaged or clicked by the user. And based upon that, we suggest an ad[1] that is most likely to be clicked, if shown next. In the following section, related work in the area is discussed, and the bayesian approach is discussed in the next section. The implementation and results are described in Section 4 & 5 and finally a conclusion is given.

2. RELATED WORK

Several solutions have been proposed for the problem of optimization in online display advertising, a comprehensive survey is out of scope for this paper. Optimization of advertisements can be done based on various factor such as a users' personal characteristics, preferences, and online behaviors as suggested in [19, 13].

Various Markov models have been proposed to capture the behavior of user sessions [11][14]. Different methods differ in the way they model links between events or whether they work towards the notion of a successful session or not. In online ad revenue maximization, different approaches have been tried, for example [6], is based on predicting the optimum time for an particular ad to be shown to a specific user. [12] works on creating decision rules to optimize the click through rate via optimum ad selection in a given situation and models like [9], or Google Ads use content information on the host page, like search query in order to give a more meaningful suggestion. More session based approaches that use propagation algorithms on graphs for user behavior modeling have been proposed in the recent past [10][7]. [8] specifically addresses the problem of optimum ad selection by traversing through aggregated user trails for the case of sponsored search. [17] suggests useful data mining techniques for frequent episodes in event sequences. In our work, we apply the model of user trails for the case of display advertisements for which the CTR is rather low. We also use other available user actions such as engagements for predicting the probability of a user clicking on an advertisement.

3. BAYESIAN MODELS

Bayesian Models have played a very important role in the area of Machine learning as they allow quantitative weighting of evidence supporting alternative hypothesis. Due to their probabilistic approach, Bayesian Models have been extensively used in inference algorithms [18]. This paper uses Bayesian Models on user action trails to select an advertisement to display such that the click probability is maximized. We believe that a user's past actions have a tremendous influence on his future behavior, hence we use past actions of a user to determine the probability of a user clicking on an advertisement. In this work, we build a prediction engine to suggest which ads are most likely to be clicked, given a set of ads a user has already acted on. According to our model,

[1]We use the word *ad* alternatively with *advertisement* in this paper.

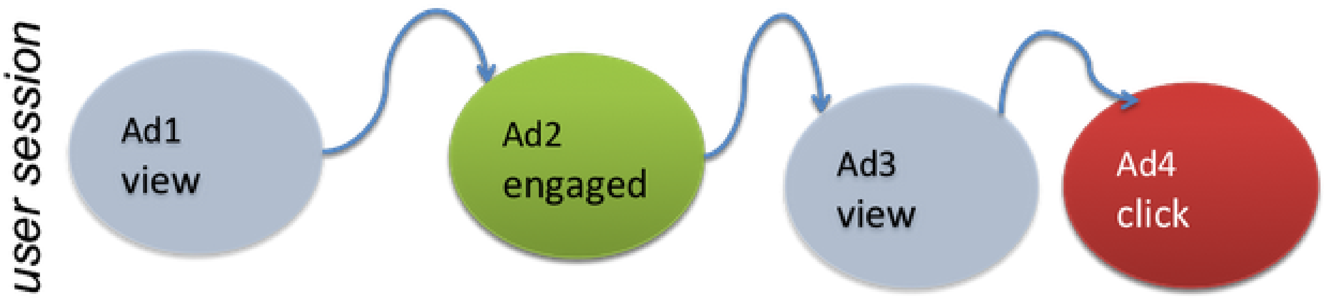

Figure 2: Typical User Session

$$P(ad_S, clicked|(ad_i, action_i), (ad_j, action_j), ..., (ad_n, action_n))$$
$$\propto P((ad_i, action_i)|(ad_S, clicked))$$
$$\times P((ad_j, action_j)|ad_S, clicked)...$$
$$\times P((ad_n, action_n)|ad_S, clicked) \times P(ad_S, clicked)$$

$$ad_{displayed} = \arg\max_S(P((ad_S, clicked)|(ad_i, action_i),$$
$$, (ad_j, action_j)..., (ad_n, action_n)))$$

If the users have already viewed or engaged in advertisements — $ad_i, ad_j, ad_k, ..., ad_n$, the above equation computes the advertisement ($ad_{displayed}$) for which the value of the probability of click is the maximum. The basic properties of this Bayesian model are:

- The model is based on aggregated user behavior based on their historic trails not on individual user behavior.
- The order in which the advertisements $ad_i, ad_j, ad_k, ...$ $, ad_n$ are shown is not considered in this simple Bayesian Model.

A simpler model would be to ignore the user actions and only consider the advertisements, this model is also tested in this paper. In spite of the independence assumptions, data sparsity in online advertising is still an issue, as the CTR is extremely low. Hence to use a richer training set, we consider the case of an engagement being a successful event like a click with lesser confidence since CTR is not the only measure of influence on the user. We experiment with the assumption of an engagement being partially effective as a click. As we will see in the results section, this significantly improves our results.

We also consider a more sophisticated model where users action on an advertisements in step i are considered dependent on step $i-1$.

$$P(ad_S, clicked|(ad_i, action_i), (ad_{i+1}, action_{i+1}), ..., (ad_n, action_n))$$
$$\propto P((ad_i, action_i), (ad_{i+1}, action_{i+1})|(ad_S, clicked))...$$
$$\times P((ad_{n-1}, action_{n-1}), (ad_n, action_n)|ad_S, clicked)$$
$$\times P(ad_S, clicked)$$

Since clicks are rare events, the information about the two step probabilities may not be available for all advertisements, for such cases we use a fallback model of using a single advertisement (1-ad). A typical user session can be pictorially represented as Fig. 2, it describes that the user viewed $Ad1$, then engaged with another $Ad2$, then viewed $Ad3$ before finally clicking $Ad4$.

4. IMPLEMENTATION

In the following subsections, we provide details of data processing, details of the experiment and the results obtained.

4.1 Data Processing

First, all the advertisements under a single campaign are selected and sessionized by a users's unique identity so that each data entry represents historical trail of a unique user. For the purposes of experiments done in this paper, we treated a user's activities as a single session for one day.

We choose the advertisements at two levels - adpod and recipe, represented by a unique AdpodID and RecipeID respectively, as they consist of the main hierarchies of advertisements in a campaign. Each adpod has a set of recipes which are mutually exclusive across different adpods. Three different types of actions - *clicks, engagements and views* are used in the model.

4.2 Experimental Setup

The goal of the experiments done in this paper is to find the best configuration on the test dataset and generalize how each parameter effects performance. We chose four parameters of extracting features as described below:

- **ID of Advertisement** (RecipeID | AdPodID)
  Both IDs are available for each ad, however, the total number of RecipeID is three times larger than AdPodID. It may deteriorate prediction accuracy.
- **Feature space** (1-ad | 2-ads | 1&2-ad)
  From a sequence of ads, we can use either a single ad or multiple consecutive ads as a feature. Using 2-ads is more expressive than 1-ad, however, it can make training set sparser. To compensate it, fallback method, using 1-ad feature when 2-ads feature is not available, is applied.
- **Notion of Success** (Click event only | Click + Engage)
  When Click event is too rare to generate an accurate model, Engage (user's mouse activity on the ad.) can be interpreted as a measure of partial success. To be fair, different weights (Click:5, Engage:1) are assigned.
- **Action Code** (Use Action Code | Do not use)
  Each ad in the feature set has either View or Engage

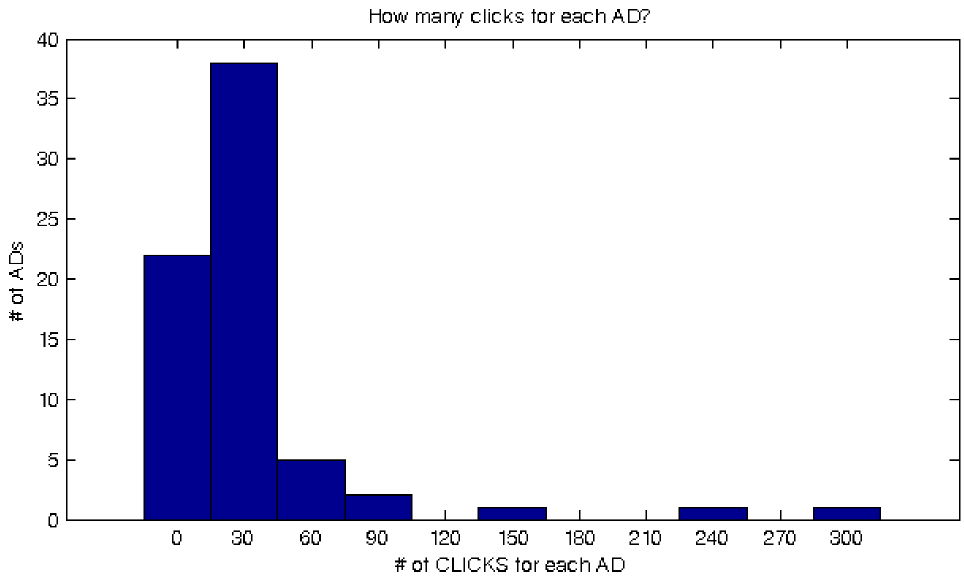

Figure 3: Distribution of Clicks For Each Advertisement

action code. Using action codes may provide more expressive model, however, it can make training set sparser.

When extracting features from session data, 4-folds technique is used in order to spare 25% of data points as testing set. For unseen events whose probabilities are zero, a small smoothing constant ($\lambda = 0.001$) for regularization. The baseline technique used to compare our work is that of choosing the advertisement whose prior probability of being clicked is the maximum regardless of the users historic trails.

$$ad_{displayed} = \arg\max_i (P(ad_i, clicked)) \quad (1)$$

5. RESULTS

To demonstrate the efficacy of our model, we use real data obtained from the logs of a dynamic display advertising company. In the logged data, the display ad server does not use historic user trail for predicting click probability, hence we avoid a bias in the training and test dataset. Since we cannot reveal any specific information about the advertisers and their campaigns, we call our data as *SampleDataset*. We use one day worth of log for a single campaign for an advertiser and use click probabilities as well as engagements as the measure of performance. The total number of impressions in the *SampleDataset* are 2.7 million, the number of recipes are 100 and number of adpods are around 30. Fig. 3 shows histogram of number of clicks vs. the number of ads and we see that most ads (at recipe level) are clicked less than 100 times and the number of clicks reasonably well-distributed.

Our prediction model ranks all the ads by predicted probabilities of being clicked under the given context of prior events in the session. To evaluate the accuracy of the ranked results, we used the position of the correct ad (which was actually clicked) in the list. We thus expressed the performance of each prediction model as a graph of cumulative accuracy over top-k predictions. For example, if k is set to 5, the performance of the model is a total occurrences of correct ads that appear in top 5 predictions. Fig. 5 illustrates k=5 is the sweet spot over various prediction thresholds. In both the results for recipes and adpods, it is clear that using only the prior probability is not a good idea. Although using Click+Engage as notion of success improves the performance a little (accuracy=0.3338, k=5), Baseline model still under-performs all the other tested models.

As shown in the left figure, while 1-ad and 1 & 2-ads configurations show almost the same performance(0.7∼0.9, k=5), 2-ads is not as good as (0.6 ∼ 0.7, k=5). These results suggests that using more expressive feature for a sparse training dataset gives deteriorated performance when compared to a simpler model. Using both Click and Engage as successful case is a clear winner(0.8∼0.9, k=5).

Another result using adpods is shown in the right graph in Fig. 5. The graph shows much better performance overall because the number of adpods is much smaller than the recipes, hence data is denser. In Table 1, the characteristic results for $k = 1, 5, 10, 20$ suggest that shadowed parameters (1-ad, 1&2-ads, Click+Engage, k=5) give out the best accuracy in the test dataset at the hierarchical level of recipes.

6. CONCLUSION

We proposed a display advertisement optimization system that delivers ads based on users past activities. Our Bayesian model is a very simple yet effective method to predict a list of ads ranked by probability of being clicked, however, the rarity of click events deteriorate accuracy. In order to overcome this issue, we applied four parameters (two level of ad hierarchies, feature spaces, notion of success, and action code) of feature extraction from data set. In the result all the tested models outperform the baseline model to considerable degrees which shows that the historical trails of user's session are helpful to improve CTR(Click Through Rate). More specifically, using engagements as partial success is proven to be a valid technique for training Bayesian model with insufficient positive actions. Our results also show that the top-k recommendation system seems to be a promising solution for advertisement selection. A sweet spot

Table 1: Characteristic Results for different values of k.

Recipe, ActionCode		k=1	k=5	k=10	k=20
1-ad	Click	0.5753	0.7641	0.8356	0.9224
	Click+Engage	0.6073	0.8463	0.8950	0.9285
1&2-ad	Click	0.5274	0.7568	0.8465	0.9240
	Click+Engage	0.5890	0.8219	0.8813	0.9269
2-ad	Click	0.4183	0.6458	0.7527	0.8595
	Click+Engage	0.4666	0.6748	0.7644	0.8419

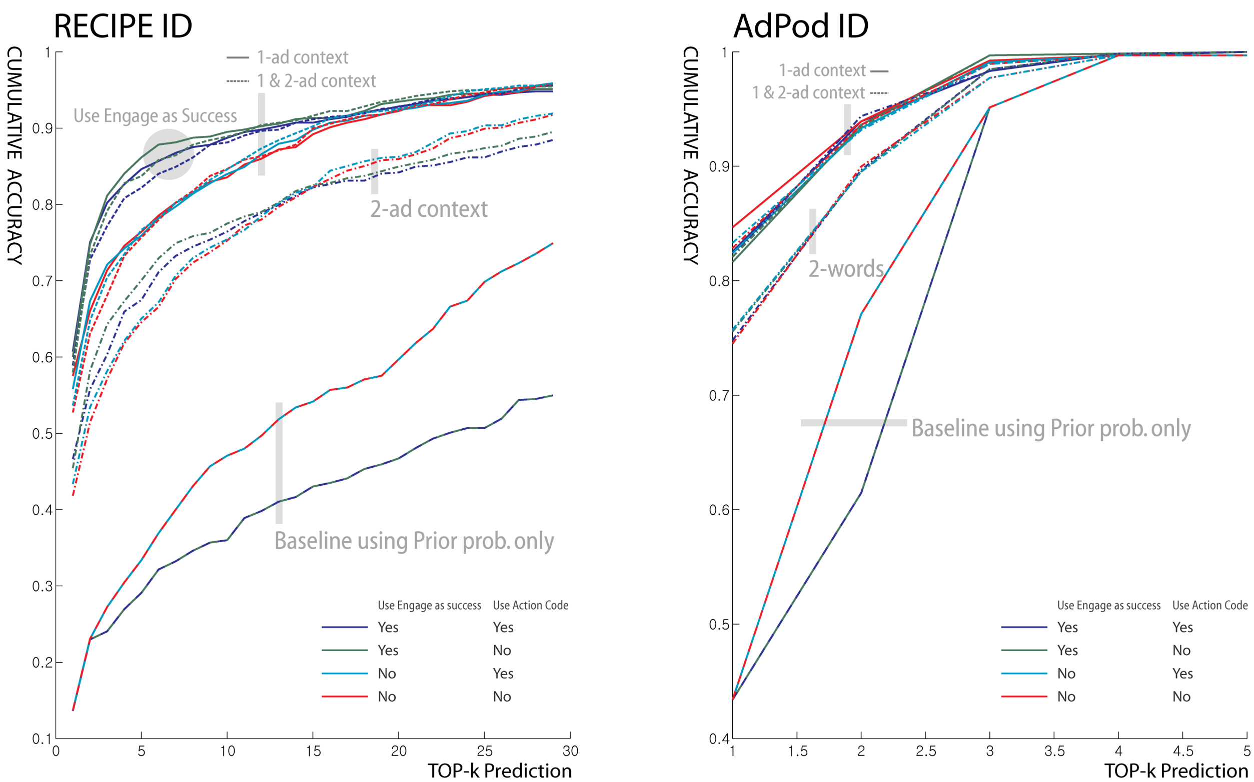

Figure 4: Cumulative Accuracy over Top-k predictions. left: RecipeID, right: AdPodID

at $k = 5$ suggests that the total set of advertisements can be narrowed to 5 when selecting advertisements to get good accuracy.

Applying this model on larger datasets using users demographic and geographic features, making long chain graphical models and online testing are all intriguing topics to be explored. The insights gained through this work are not only beneficial to an online display advertising system but can be applied in other domains dealing with sequential events of rare positive cases such as medical history inference and emergency prediction.

7. ACKNOWLEDGEMENTS

We are very thankful to Prof. Lise Getoor and Prof. Ashok Agrawala for their valuable insights during the course of this work. We are also grateful to Tumri Inc. for sharing their user logs for the purpose of this research.

8. REFERENCES

[1] http://advertising.microsoft.com/.

[2] http://advertising.yahoo.com/.

[3] http://www.google.com/ads/.

[4] http://www.teracent.com/.

[5] http://www.tumri.com/.

[6] N. Abe and A. Nakamura. Learning to optimally schedule internet banner advertisements. In *Proceedings of the Sixteenth International Conference on Machine Learning*, ICML '99, pages 12–21, San Francisco, CA, USA, 1999. Morgan Kaufmann Publishers Inc.

[7] A. Anagnostopoulos, L. Becchetti, C. Castillo, and A. Gionis. An optimization framework for query recommendation. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 161–170, 2010.

[8] N. Archak, V. S. Mirrokni, and S. Muthukrishnan. Mining advertiser-specific user behavior using adfactors. In *Proceedings of the 19th international conference on World wide web*, pages 31–40, 2010.

[9] A. Ashkan, C. L. Clarke, E. Agichtein, and Q. Guo. Estimating ad clickthrough rate through query intent analysis. In *Proceedings of the 2009 IEEE/WIC/ACM*

International Joint Conference on Web Intelligence and Intelligent Agent Technology-Volume 01, pages 222–229, 2009.

[10] P. Boldi, F. Bonchi, C. Castillo, D. Donato, A. Gionis, and S. Vigna. The query-flow graph: model and applications. In *Proceeding of the 17th ACM conference on Information and knowledge management*, pages 609–618, 2008.

[11] H. Cao, D. Jiang, J. Pei, E. Chen, and H. Li. Towards context-aware search by learning a very large variable length hidden markov model from search logs. In *Proceedings of the 18th international conference on World wide web*, pages 191– 200, 2009.

[12] K. Dembczynski, W. Kotlowski, and D. Weiss. Predicting ads click-through rate with decision rules. In *Proceedings of the 17th International Conference on World Wide Web (WWW 2008), Beijing, China*, 2008.

[13] T. Graepel, J. Q. Candela, T. Borchert, and R. Herbrich. Web-scale bayesian click-through rate prediction for sponsored search advertising in microsoft's bing search engine. In *ICML*, 2010.

[14] A. Hassan, R. Jones, and K. L. Klinkner. Beyond DCG: user behavior as a predictor of a successful search. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 221–230, 2010.

[15] K. L. Keller. Brand equity and integrated communication. *Integrated communication: Synergy of persuasive voices*, pages 103–132, 1996.

[16] R. Lewis and D. Reiley. Retail advertising works! 2008.

[17] H. Mannila, H. Toivonen, and A. I. Verkamo. Discovery of frequent episodes in event sequences. *Data Mining and Knowledge Discovery*, 1(3):259–289, 1997.

[18] T. M. Mitchell. *Machine Learning*, volume 1 of *Wiley Series in Probability and Mathematical Statistics*. John Wiley and Sons, Inc., international edition, 1997.

[19] J. Yan, N. Liu1, G. Wang1, W. Zhang, Y. Jiang, and Z. Chen. How much can behavioral targeting help online advertising? In *WWW*, 2009.