Summer 6-15-2016

PREDICTING ONLINE USER BEHAVIOR BASED ON REAL-TIME ADVERTISING DATA

Martin Stange
Leuphana University, martin.stange@uni.leuphana.de

Burkhardt Funk
Leuphana University, funk@uni.leuphana.de

PREDICTING ONLINE USER BEHAVIOR BASED ON REAL-TIME ADVERTISING DATA

Research

Stange, Martin, Leuphana University, Lueneburg, Germany, martin.stange@uni.leuphana.de

Funk, Burkhardt, Leuphana University, Lueneburg, Germany, funk@uni.leuphana.de

Abstract

Generating economic value from big data is a challenge for many companies these days. On the Internet, a major source of big data is structured and unstructured data generated by users. Companies can use this data to better understand patterns of user behavior and to improve marketing decisions. In this paper, we focus on data generated in real-time advertising where billions of advertising slots are sold by auction. The auctions are triggered by user activity on websites that use this form of advertising to sell their advertising slots. During an auction, so-called bid requests are sent to advertisers who bid for the advertising slots. We develop a model that uses bid requests to predict whether a user will visit a certain website during his or her user journey. These predictions can be used by advertisers to derive user interests early in the sales funnel and, thus, to increase profits from branding campaigns. By iteratively applying a Bayesian multinomial logistic model to data from a case study, we show how to constantly improve the predictive accuracy of the model. We calculate the economic value of our model and show that it can be beneficial for advertisers in the context of cross-channel advertising.

Keywords: Online User Behavior, Real-Time Advertising, Iterative Bayesian Multinomial Logistic Model.

1 Introduction

As more and more data is generated by customers, sensors, or governments, business intelligence and analytics become increasingly important. Many practitioners and researchers have been focusing on this topic and have developed methods to measure the impact of big data in recent years (Chen et al., 2012). In this paper, we focus on the impact of a fairly new source of big data generated during the real-time advertising process on the Internet.

In real-time advertising, advertising slots on a website are sold by auction in the 200 milliseconds after a website is called by a user. In the first few milliseconds after the call, website context information (such as content, language, quality) and an anonymous user id is sent to a so-called ad exchange. The ad exchange, which is a marketplace for advertising slots, sends out so-called bid requests to advertisers and their service providers, who employ bidding agents, which instantaneously select the advertising media which best fits to the current user's interests. In addition, these agents determine the maximum price the advertiser is willing to pay for the ad impression at auction. This information is bundled to a so-called bid response and sent back to the ad exchange, which forwards the advertising media of the highest bidder to the publisher's website and charges the highest bidder the second highest price (second price auction). The whole process is completely invisible for the user because, as the website is completely loaded, the auction is already closed. As the process happens tens of thousands of times per second, it is a distinct source of big data (Stange and Funk, 2014). In this paper, we develop a model to gain economic value from the massive amount of data that is generated during this process.

Bidding agents in real-time advertising usually make use of information from recent customer activity (i.e., cookie data) on the advertiser's website to determine appropriate advertising media for the current user. Users often recognize this by being exposed to advertising material of products they were searching for recently (re-targeting). Of course, real-time advertising is not limited to this rather simplistic form of advertising. However, there is often not enough information about the current user and his or her interests available to make better decisions (Perlich et al., 2012). In this context, we propose a new approach to derive users' interests based on the stream of bid requests that were generated by their browsing activity, and show that users' interests can be accurately predicted by only using this data. In our approach, a user's interest in a certain product is assumed if he or she visited a website related to this product during his or her journey. The method enables advertisers to expose ads only to users that exhibit a certain probability to be interested in their products. At the same time, users who will most likely never be customers can be ignored. In our case study, we focus on the users' interests in certain TV programs.

We contribute to IS research by determining the impact of bid request data from an advertiser's perspective (Chen et al., 2012). To determine this impact, we calculate the economic value of a person-centered model that can be used to understand and predict users' behaviors on the Internet based on bid request data. We develop an iterative Bayesian model that enables us to update once trained parameters with new data according to Bayes' rule. This model can be used by researchers to develop and extend decision support systems on the Internet and to develop new business models for predictive analytics in the field (Veit et al., 2014). The approach is not intended to replace well established methods to target users in online marketing contexts. Instead, our method is supposed to be a conceptual extension of the landscape of methods and tools to target users with proper advertisements. In practice, our approach may be valuable for TV stations or their agencies to coordinate TV and online advertising campaigns, for instance (Joo et al., 2015; Stange, 2015). In this context, the proposed method could be used to expose ads online only to users who show a high probability for having watched a certain TV program recently. We apply the approach to bid request data from a major ad exchange and show how benefits from cross-channel advertising activities can be increased.

The remainder of this paper is structured as follows: First, we review recent literature on using tracking data to analyze user behavior. Second, we describe our modeling approach and show how to integrate the model into bidding agents. Third, we describe data collection and preparation. Fourth, we present the results of the analysis and calculate its economic value from an advertiser's perspective. Finally, we discuss the implications of our study.

2 Related Work

In our study, we make use of several results from IS and marketing research, which are going to be outlined in the following paragraphs.

For many companies in e-commerce, it is crucial to identify their customers' interests for products and services. It is clear that the more information companies have available about their customers, the better they can customize their products to the clients' individual needs. For this reason, researchers have analyzed the users' click and purchase behavior on the Internet to make better marketing decisions (Bucklin and Sismeiro, 2009) in order to achieve, for instance, an optimal fit between advertising materials and users or to offer customized products.

Chatterjee et al. (2003) developed a user journey model consisting of variables representing long term and short term advertising effects based on clickstream data that was generated in advertising campaigns. They conducted a hierarchical logistic regression to estimate the variables' effects on the users' click probability. With the results, the impact of individual advertising channels on the customers' click behavior can be extracted and thus, predictions can be made about the click probabilities of future users. This outcome can be used to increase the effects of display advertising campaigns.

However, users' click probability does not necessarily correlate with their probability to purchase a product or to register for a newsletter, i.e., the users' conversion probability (Lee et al., 2012; Pandey et al., 2011). In the context of real-time advertising, these conversion probabilities can be used to determine the size of bids. Many researchers (e.g., Perlich et al., 2012; Zhang et al., 2014) have developed methods to determine the most profitable bids from the perspective of a demand-side platform—a service provider who places bids on behalf of the advertisers. Approaches in the literature often focus on using real-time advertising for performance-oriented displaying of advertising materials, i.e., on selecting ads that will most likely lead to direct purchases (Chen and Berkhin, 2011), and on placing optimal bids for these ads (Zhang et al., 2014). In addition, published studies have focused on optimal selection of ads with respect to budget constraints, or recency and frequency capping (Yuan et al., 2013). Lee et al. (2012) showed how to use past performance data to effectively determine the right advertisement to be exposed to the right user on the right publisher website. The proposed models can be used to improve advertising effects or to reduce costs. However, performance-oriented ad selection aims to target users at a late stage in the sales funnel, where often sufficient information about a given user is available. By contrast, our approach aims to target users at an early stage in the sales funnel and focuses on predicting user interests only based on bid request data.

A growing number of advertisers use real-time advertising also for branding campaigns. In these campaigns, exposing the right ads to the right user is a challenge because in this early phase of the sales funnel only little information about users is available. In the context of branding campaigns, many authors have pointed out the importance of a cross-channel advertising strategy to increase users' awareness of certain products (e.g., Dinner et al., 2014; Duan and Zhang, 2014; Joo et al., 2015; Yang and Ghose, 2010). The idea behind this strategy is that advertising activities should not only be focused on individual advertising channels but should also consider synergies that can be observed when advertising activities on different channels are seamlessly coordinated. For instance, it might be very sensible to combine a TV advertising campaign with a complementary search engine advertising campaign or an e-mail advertising campaign instead of treating these advertising activities individually (Liaukonyte et al., 2015; Stange, 2015). In contrast to the effect of online advertising, however, the possibilities to measure the effect of offline advertising are limited (Kitts et al., 2014), and therefore it is often challenging to coordinate online and offline advertising campaigns effectively. We address this challenge by developing a method to increase benefits from awareness-related offline-online advertising campaigns using bid request data. We use the method to predict whether a user has watched a certain TV program recently.

Chen et al. (2012) demonstrated the increasing impact of big data analytics that can be observed in many industries. To identify the impact of bid request data on companies in e-commerce and marketing, we propose to measure its economic value as it is proposed by Nottorf and Funk (2013). In the context of online advertising, they determine the economic value of clickstream data by assigning (negative) costs to true and false predictions of a classifier that was trained using the data. We apply this method to a multinomial classifier that is trained using bid request data. Thereby, we show that the analysis of this kind of big data can be particularly beneficial for companies conducting awareness-oriented advertising campaigns.

3 Model Development

This section first discusses the general framework of the modeling approach used here. Second, it describes the proposed model in detail. Third, it presents a process that shows how the model can be used by bidding agents in real-time advertising.

3.1 Modeling Approach

The goal of our analysis is to calculate the probabilities that a user is going to visit certain websites during his or her journey. We interpret these probabilities as an indication of the user's interests. In our case study, we focus on the users' interest in a certain TV program. To identify whether a user is interested in a certain TV program or not, we use the websites of five different TV stations. For instance, we assume that a user who visits rtl.de is most likely interested in the TV program aired on RTL. Our method could, for instance, be used by TV stations or their agencies to enrich the TV advertising campaigns of their customers (i.e., the advertisers) with complementary online advertising campaigns. Of course, it is impossible to fully understand users' actual TV consumption behaviors only based on their online user journeys. However, since the goal is to demonstrate the possibilities of using bid request data to determine users' interests, we consider this to be only a minor limitation.

The modeling approach presented in this paper addresses the challenge of the high volume and velocity of bid request data. We handle the high velocity by iteratively applying a multinomial Bayesian logistic model, which uses prior knowledge about its parameters as follows: Initially, no information about the regression parameters is available. After the first iteration, the model returns parameter distributions based on the first batch of data. We extract the means and standard deviations of these distributions to then use these values as prior information for the subsequent run. Thus, the precision of the parameter distributions and the predictive accuracy of the model increase with the number of iterations.

We handle the high volume of bid request data by applying stratified sampling. As many analytical tasks in the context of e-commerce, such as the prediction of purchases or clicks, have to deal with rare events (i.e., conversions or clicks are very rare compared to the number of views), this approach can save computational costs. In our data set, over 99% of the overall data set contains user journeys that are irrelevant for our analysis. Stratified sampling makes the learning algorithm more efficient, since less data is required to estimate the parameters.

3.2 Model Description

The dependent variable of our analysis is denoted as $K \in \{0,\dots,5\}$ and indicates whether a user opened the website of a specific TV station during the entire user journey or not. We regard this variable as a proxy variable that indicates users' interest in a certain TV station and their potential interest in certain products advertised on this station. We include the following target URLs: rtl.de ($K=1$), rtl2.de ($K=2$), vox.de ($K=3$), sat1.de ($K=4$), and prosieben.de ($K=5$). If a user did not open one of these websites, K is set to 0. The independent variables are all other URLs from which bid requests can be triggered. For instance, if a user i visits an URL U_j such as amazon.de and ebay.de during her or his user journey J_i, the set of independent variables for this user is 1 for both of these variables, and all other variables are set to 0. In Equation 1, X_{ij} is the j^{th} covariate from the design matrix X at row i, and U_j is the j^{th} entry from a list of URLs.

$$X_{ij} = \begin{cases} 1 & \text{if } U_j \in J_i \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

We use a Bayesian multinomial logistic regression model, which we estimate using JAGS (Plummer, 2003) based on Equation 2:

$$\begin{aligned} K_i &\sim \text{Multinomial}(p_i) \\ p(K_i = l|X_i) &= \frac{\exp(\alpha_l + X_i\beta_l)}{\sum_k \exp(\alpha_k + X_i\beta_k)} \\ \beta_{jk} &\sim \text{Normal}(b_{jk}, \sigma_{jk}) \\ \alpha_k &\sim \text{Normal}(a_k, s_k) \end{aligned} \tag{2}$$

In this equation, α_k represents the intercept for class k, and β_{jk} represents the j^{th} entry of the parameter vector β_k, i.e., the slope for the j^{th} URL in the list of parameters of class k. The values a_k and b_{jk} represent the prior knowledge of α_k and β_{jk}. The terms s_k and σ_{jk} represent the prior knowledge of the standard deviation of α_k and β_{jk}. The term $p(K_i = l|X_i)$ represents the probability that the row vector X_i is of class l. In each iteration, the prior values a_k, b_{jk}, σ_{jk}, and s_k are updated with the posterior means and standard deviations from the previous steps of the analysis. The initial values for α_k and β_{jk} are defined as 0, whereas σ_{jk} and s_k are set to 10.

In our case study, the number of users labeled with $K = 0$, i.e., users who never visited a website of a TV station, is high in comparison to the other classes. For this reason, stratified sampling for training the model has been recommended (King and Zeng, 2001). However, after parameter estimation, the sampling bias needs to be considered before using the results for prediction in accordance with Equation 3:

$$p(K_i = l|X_i) \quad = \quad \frac{\exp(\alpha_l + X_i\beta_l) \cdot \tau_l / \bar{y}_l}{\sum_k \exp(\alpha_k + X_i\beta_k) \cdot \tau_k / \bar{y}_k} \tag{3}$$

In this equation, τ_k represents the ratio of instances of class k in a random sample, and $\bar{y}_k$ represents the ratio of instances of class k in the training set (King and Zeng, 2001).

For each iteration, we only include variables of URLs that are contained in the data of the current training batch. The parameters of the other variables remain unchanged. This is possible due to the assumed statistical independence between the parameters and enables the model to include a large number of parameters.

3.3 Updating the Decision Engine

We propose to include the model in a real-time advertising decision support system as presented in Figure 1 and described in the following.

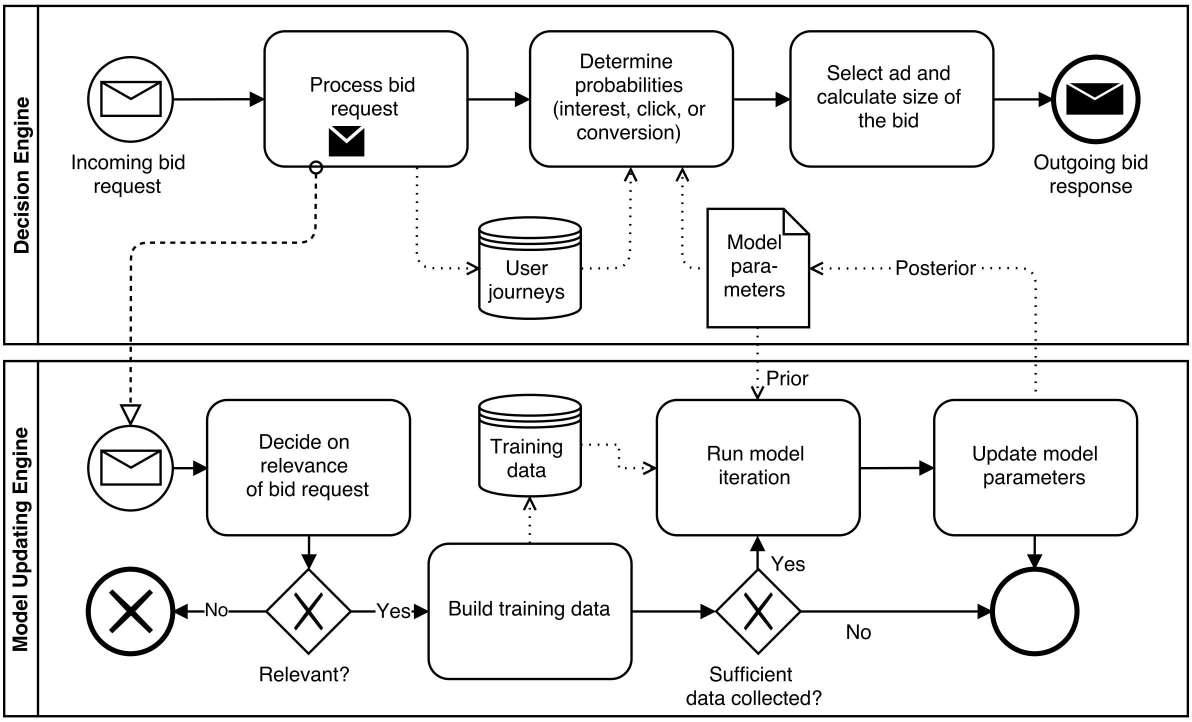

Figure 1. Bidding and model updating process. Each Bayesian model iteration uses prior information from the previous iteration.

The process begins with the incoming bid request, which is analyzed and stored by the decision engine. Based on prior knowledge of the current user's interests, the decision engine selects the advertising material and the size of the bid and sends a bid response to the ad exchange. Simultaneously, the bid request is forwarded to the model updating engine, which first assesses the relevance of the bid request for the analysis. Depending on its relevance, the bid request is discarded or stored in a training database.

The Bayesian analysis is executed when the amount of data reaches a predetermined minimum sample size. Afterwards, the posterior information is sent to the decision engine, which then uses the updated parameters for prediction as new bid requests are processed. Since the model is trained with a stratified sample, the prediction algorithm must rescale the true probabilities of the individual classes in accordance with Equation 3.

4 Data Description and Preparation

Our data set was provided by a German cross-media online marketing agency and consists of bid requests from a major ad exchange in the form of URL query strings. A query string contains the URL of the website triggering the bid request, the anonymized ID of the current user, location information on the user, and the timestamp of the visit (Figure 2)[1].

The data set contains 3 TBytes of bid request data for a time period of 4 days. During this period, over 1.4 billion bid requests were triggered by over 35 million unique users. The data set contains 35,058,383 users who never visited the websites of the TV stations during their journeys, 275,167 users who visited rtl.de, 56,416 users who visited rtl2.de, 3,978 users who visited vox.de, 3,529 users who visited sat1.de, and 6,738 users who visited prosieben.de.

```
h=http%3A%2F%2Fwww.jetztspielen.de%2Fvda%2friendlyiframe_html_
40.2.1&t=1396894441.691&id=7358864011747200610&ip=93.84&s=DE&c
=Ludwigsburg&a=Mozilla%2F5.0+%28compatible%3B+MSIE+9.0%3B+Wind
ows+NT+6.0%3B+Trident%2F5.0%29

h=http%3A%2F%2Fwww.ebay.de%2F&t=1396894441.692&id=130762252527
275372&ip=88.65&s=DE&c=Munich&a=Mozilla%2F5.0+%28iPad%3B+CPU+O
S+7_1+like+Mac+OS+X%29+AppleWebKit%2F537.51.2+%28KHTML%2C+like
+Gecko%29+Version%2F7.0+Mobile%2F11D167+Safari%2F9537.53%2Cgzi
p%28gfe%29
```

Figure 2. Examples of bid requests from the raw data. The variable h represents the triggering URL, t the timestamp, id the ID of the current user. In our approach we do not include information on user agents (a) and geographical information (s, c). However, in a real-life situation, these variables may lead to more accurate predictions.

The process of data preparation is often not discussed in the literature, but we would argue that this process deserves critical attention because it may have implications for practitioners who intend to use our model. For this reason, we share our experience with the community and briefly describe how we transformed the data into user journeys. First, we stored the raw data (text files, each capturing one minute of traffic) in a Hadoop file system and then accessed it with Apache Spark. To reduce the number of features, we stripped the URL data after the top-level domain and aggregated the resulting list. We ranked the URLs by visit and encoded the URLs with their positions in the list. We removed all websites with fewer than 5 visits in the data set, resulting in over 500,000 base URL entries. Second, we grouped the data set by user ID and defined the dependent variable for each user by labeling the user with one of the six classes. Our sample does not contain users who visited more than one websites of a TV station, because we truncated all users journeys after the first visit of a TV station website to avoid a bias originating from bid requests that are directly related with the TV station website. Finally, we wrote each user journey into one line of

[1] Note, that it was not possible to draw conclusions about users' personal information from the data at any time.

an output file, starting with the class followed by a list of numbers separated with commas. These numbers represent the URLs a user visited during his or her journey. This process resulted in a text file of 4 GBytes.

5 Results

In this chapter, we first describe the results of the iterative parameter estimation. Second and third, we report the misclassification error of the model on a stratified holdout sample and on a random holdout sample. Fourth, we evaluate the model by calculating its economic value.

5.1 Parameter Estimation

We split the data into a training set (75% of the available data set) and a holdout set (25% of the available data set). To train the model iteratively, we split the training set into 75 separate batches, each containing 1,600 data records. Each batch contains 500 records of Class 0, 400 records of Class 1, 300 records of Class 2, 100 records of Class 3, 100 records of Class 4, and 200 records of Class 5. This ratio of classes is loosely based on the number of each classes' instances in the complete data set on logarithmic scale. We run the analysis with 15,000 parameters, i.e., we included the 15,000 most frequently visited websites into the model.

Each run of the MCMC sampler consists of 4 chains with 3,000 burn-in iterations and 3,000 sampling iterations. We keep every tenth record from the posterior sample to avoid auto-correlation of the Gibbs sampler. Each iteration took approximately 40 minutes on an Intel i7 4820K processor with 3.7 GHz. Figures 3 and 4 show the densities of four selected parameters for different iterations. Both figures indicate that the precisions of the parameters' posterior densities increase with the number of iterations. A comparison of Figure 3 and 4 shows that the less frequent a variable is included in the data, the less precise the estimation of its parameter, and the more iterations are required to obtain a desired parameter precision.

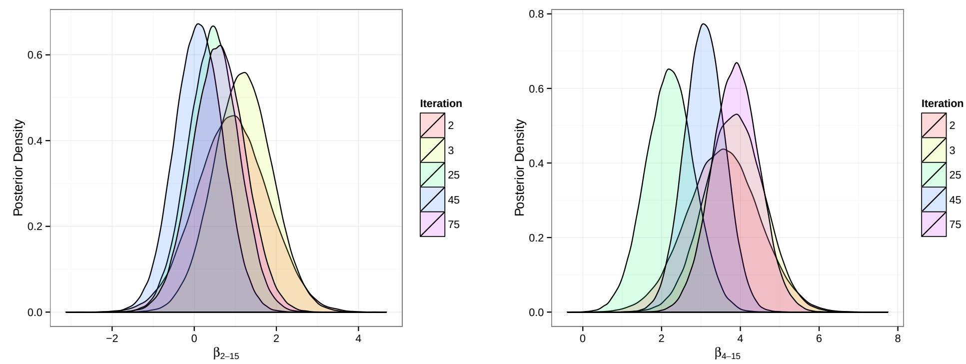

Figure 3. Posterior densities of $\beta_{2,15}$ and $\beta_{4,15}$.

5.2 Misclassification Error

After each iteration, we calculate the misclassification error of the model on a stratified holdout sample. The baseline of the misclassification error is established by making a random guess. This approach results in an error of 83.3% when, as in this case, an equal number of instances of the six classes is included in the holdout set. Figure 5 shows that the misclassification error decreases with the number of iterations, i.e., the more data is considered for training the model, the more likely it is that a data record is classified correctly (Stange and Funk, 2015).

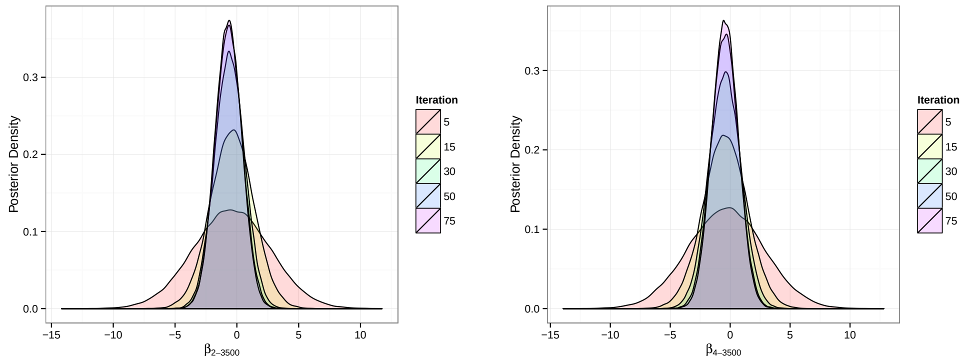

Figure 4. Posterior densities of $\beta_{2,3500}$ and $\beta_{4,3500}$.

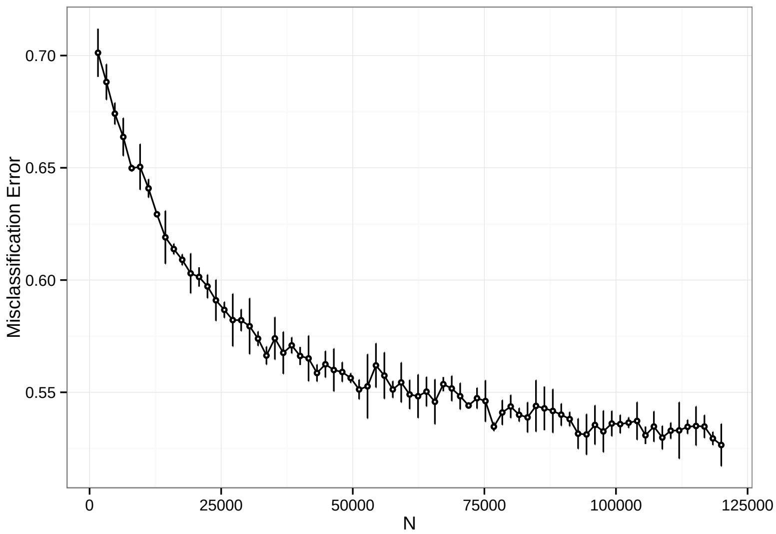

Figure 5. Misclassification error on the holdout sample depending on the number of training samples. The misclassification error seems to converge to a minimum as more and more iterations of the analysis are performed.

Figure 6 presents the confusion matrix describing the misclassification error of the model using the parameters obtained from the 75^{th} model iteration.

To determine how many user contacts have to be observed to achieve a desired misclassification error, we examine the relationship between the misclassification error and the observed user journey lengths. Figure 7 shows that the misclassification rate decreases as user journey length increases. This result is expected: The more data on users is available, the more accurate the prediction. The blue line in Figure 7 shows, for instance, that users who have been observed at least on 10 different websites are classified correctly in 65% of all cases (i.e., the misclassification error is approximately 35%). The red line shows that users who have been observed 8 to 10 times are classified correctly in more than 60% of all cases (misclassification error of 40%). The green line shows that users who have been observed less than 6 times will be classified correctly in 40% of all cases (misclassification error of 60%).

True Class \ Predicted Class	0	1	2	3	4	5
0	**1249**	532	421	224	75	113
1	709	**858**	612	237	84	114
2	574	524	**976**	226	113	201
3	494	454	441	**1133**	45	47
4	164	179	304	67	**1632**	268
5	198	201	363	70	204	**1578**

Figure 6. Confusion matrix of the prediction on the stratified holdout set after the last iteration of the analysis (misclassifcation error: 52.7%).

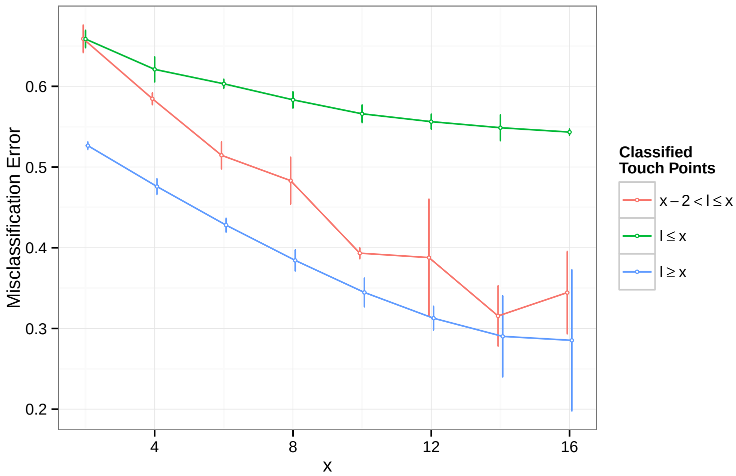

Figure 7. Misclassification error vs. user journey length. The more information on users is available, the more accurate the prediction of user classes.

5.3 Predictions based on a Random Sample

Stratified sampling is a feasible way to benchmark a model such as the one described here. However, in a real-life scenario, the distribution of user classes is greatly unbalanced. For this reason, we rescale the parameter estimations to predict user classes on the random sample in accordance with Equation 3.

We calculate the probabilities $p(K_i = l|X_i)$ to classify each record of a random holdout set. We obtain a rather small misclassification error of 2.5% (Figure 8). This value is so small, because of the high number of records of Class 0. Thus, classifying a record with $K = 0$ is nearly always correct. Even a model that yields $p(K = 0) = 1$ for each test record would nearly always classify correctly. For this reason, using the misclassification rate is not a meaningful means to evaluate the model. Instead, we calculate the economic value of applying the model, and, thereby determine the economic value of bid request data.

		Predicted Class					
		1	2	3	4	5	6
True Class	1	**584751**	6457	1528	58	496	1981
	2	3467	**129**	18	0	4	18
	3	719	13	**24**	1	0	5
	4	45	0	1	**1**	0	0
	5	69	2	1	0	**19**	0
	6	152	4	0	0	0	**42**

Figure 8. Confusion matrix obtained from applying the model to a random holdout set containing 600,000 records. The misclassification rate is 2.5%.

5.4 Economic Value of Bid Request Data

We determine the economic value of bid request data from the perspective of an advertiser who employs a bidding agent that places bids based on predictions made by our model. The bidding agent would not place a bid for predictions of Class 0, i.e., it would not answer to bid requests triggered by users who are unlikely to be interested in one of the TV programs. In case the prediction of Class 0 is correct, the behavior of the bidding agent produces no costs (true negative prediction). Otherwise, i.e., the current user is in fact interested in one of the TV programs (false negative prediction), the behavior of the bidding agent produces opportunity costs. These costs are determined by the lost contribution margin for not exposing a user to an ad, who would have clicked on the ad or at least have been attracted by it. For predictions of Class 1 through 5, the bidding agent would always place a bid. Consequently, it would produce costs that are equal to the costs of the ad impressions. In addition, it would generate benefits that can be derived by exposing an interested user to an ad that matches his or her interests (true positive prediction).

Based on the aforementioned scheme, we estimate the economic value of bid request data by applying our model to a random holdout sample. We assume the typical costs in the industry for ad impressions and typical benefits from user clicks on display ads at the time of writing. For false negative predictions, we assume costs ranging from 0.01 EUR through 0.40 EUR. We define a range of costs here, because the contribution margin may vary for different advertising scenarios. For false positive predictions, we assume a value of 0.001 EUR, i.e., the costs for an ad impression. For true positive predictions we assume a benefit equal to the contribution margin (i.e., 0.01 EUR through 0.40 EUR) minus the costs for the ad impression (0.001 EUR). A true negative prediction, i.e., no banner is shown to a user who would not have clicked, has no costs at all. For false predictions concerning the Classes 1 through 5, we also assume no costs because users belonging to these classes might, in general, be interested in products they see advertised on TV, but at another TV station. We assume that the costs for the ad impression is annulled by the benefit through the branding effect.

To obtain the best balance between true and false predictions and their costs and benefits we seek for the optimal cutoff value p_{cut} for $p(K > 0)$ that minimizes the costs of the classification by iteratively classifying the data from the holdout set with different values for p_{cut}. The cutoff value is used as follows: All user contacts X_i that show a probability $p(K = 0)$ less than p_{cut} are classified based on the probabilities of the other classes, i.e., $p(K \in \{1,\ldots,5\})$. All other user contacts X_i are classified as class $K=0$. Thus, when p_{cut} is very small, nearly all contacts are classified as one class between 1 and 5, and, as p_{cut} is increases, more and more contacts X_i are categorized as class $K=0$.

To calculate the benefits related to a given cutoff value, we define the maximum benefit per decision of a model B as the minimum of the costs for showing no impressions at all (C_{NI}) and the costs for showing impressions to all users (C_{AI}) minus the minimum costs of applying the model (C_M) using the optimal cutoff value (Equation 4):

$$B = \min(C_{NI}, C_{AI}) - \min(C_M) \tag{4}$$

Figure 9 presents the benefits per 1,000,000 decisions based on different benefit/cost ratios. The diagram indicates that using the model for classification is valuable in a relatively narrow range of benefit/cost ratios. On the left side of the ratio range (below a benefit/cost ratio of 40), the advertiser would never bid for a free ad slot, because the overall costs of impressions are higher than the sum of contribution margins. Above the ratio of 40/1, the advertiser is best advised to always use the model for classification. The highest benefit can be achieved when the benefit/cost ratio is at 75/1. As the ratio increases, the benefit that can be generated by applying the model decreases.

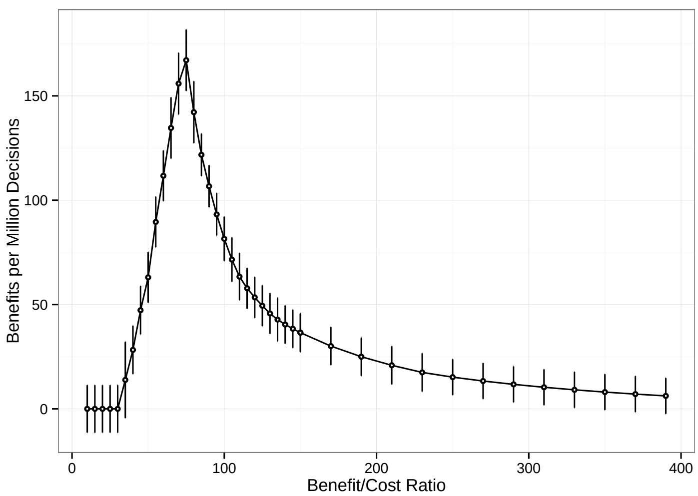

Figure 9. Max benefits vs. cost benefit ratio. The model is most beneficial when the benefit/cost ratio is 75/1, i.e., when the costs for a lost contribution margin are 75 times greater than the costs for an impression. The optimal cutoff value for this benefit/cost ratio is $p_{cut} = 0.007898$.

6 Conclusion

In this paper, we develop a model employing bid request data that can be used to derive users' interests in certain products or services. Our approach has implications for researchers and practitioners from IS and marketing (Spann et al., 2013).

6.1 Implications

The model developed here enables researchers and practitioners to make predictions about users' interests for arbitrary products or services by using corresponding websites as dependent variables. A company offering smart phone contracts, for instance, can use bid requests triggered by a mobile phone product website to identify users who are interested in a new smart phone very early, namely before they visit the website of the mobile phone company. In this way, our approach extends the landscape of methods and techniques to target users based on cookie data or third-party data. In marketing research, the method can be used to measure the impact of television ads on the users' online shopping behavior more accurately because it allows for identifying users who have watched a certain TV program recently (Stange, 2015). In addition, we contribute to IS by developing a framework to extend and improve existing decision support systems employed in e-commerce and marketing to target users based on cookie data. Furthermore, we provide a method to measure the impact of bid request data that is based on the valuation of the analysis conducted in our case study.

In practice, the method can be used to increase profits from cross-channel advertising campaigns. In particular, awareness-oriented campaigns can benefit from the proposed model. These campaigns aim to reach a broad audience early in the sales funnel where relatively little is known about a given user's interests. Thus, using bid request data streams in addition to cookie data and other third-party data is a promising possibility to improve effects from branding campaigns. The integration into existing decision engines employed by ad exchanges or demand side platforms is relatively easy because of the iterative approach of the analysis. They could use the proposed method to extend their portfolio of targeting services. Aside from cross-channel advertising, the probability of a user being interested in certain content can be used by e-commerce companies to customize their products and services.

6.2 Limitations

Although the approach proposed here suggests successful possibilities for using bid request data to predict user behavior, it also has primarily five limitations: First, our bid request data was not filtered for any data from background processes such as ad servers, which are not necessarily directly related to the user journey. In addition, the data contains websites in different languages. These websites should eventually be excluded from the data because is very likely that a user is not interested in a German TV program if he or she has only visited French or Polish websites during his or her journey. In addition, we expect a bias in our training data because it is likely that not all (sub) websites of the TV stations we focused on use real-time advertising to sell their advertising inventory. Hence, users who exclusively visited such (sub) websites are falsely labeled with Class 0 in our training data. Second, the Gibbs sampler used here to estimate the parameter values during the iterative analysis is relatively slow. It takes about 40 minutes to run 4 chains simultaneously with 6.000 sampling iterations on an Intel i7 4820K processor. This duration might be unsuitable for some applications, especially when the number of relevant bid requests is much higher than in our case study. However, because of the statistical independence of the parameters, it is possible to parallelize sampling of parameters on GPUs using programming languages such as CUDA-C. These improvements would reduce computational costs to a great extent. Another opportunity to speed up computation time is to employ variational Bayesian methods. These methods can be used to approximate parameters of simple models such as the one presented here. However, the complexity of models that can be estimated with these methods is limited. Third, we assume statistical independence between the URLs to make the model computationally tractable. However, website visits by a given group of users with the same interests are generally not statistically independent. In addition, using TV station websites as proxy variables to determine users' TV consumption behavior can only be an approximation because many users might tend to watch the program of a certain TV station but never visit its website. Fourth, the valuation of the model is simplified, because it assumes that every impression can be bought at the same price. In real-time advertising, however, the general approach is to pay individual prices for individual users. Thus, the calculated benefits based on true and false predictions is only an approximation of the true benefit that could be achieved in a real-life system. Fifth, our approach is not intended to replace current methods applied in behavioral targeting because only relying on bid request data streams could lead into a 'big data trap' (Lazer et al., 2014). Therefore, we recommend to use the proposed method as an extension to techniques and tools used in behavioral targeting. More conceptual research is required on how to integrate predictions from our model into existing bidding agents in RTA.

6.3 Outlook

Unlike typical bid request data from other major ad exchanges, our data set neither contains information on free advertising slots nor any contextual information. The model does not contain time-dependent variables that are commonly used in the user journey analyses, such as the number of contacts with a certain website or the time difference between contacts (Chatterjee et al., 2003). However, due to the flexible approach using Bayesian estimation that employs MCMC sampling, it is rather simple to extend the model by new variables and hierarchy levels. For instance, website languages or user locations could be used to determine users' interests depending on their location or spoken languages. In addition, text-mining techniques could be used to derive context information from websites, which could be used as additional independent variables. This kind of information could be very useful for predicting user behavior and would enable e-commerce companies to understand their customers' behavior even better. The high number of parameters could be reduced by applying regularized Bayesian regression that employs non-normal prior distributions (Kyung et al., 2010). However, further investigation is needed on how to use non-normally distributed posteriors as prior information in an iterative modeling approach such as the one presented here.

We present an approach to predict the probability for a user being interested in a certain product. It is clear that these predicted probabilities are not the only values that bidding agents require to place a bid. The integration of predictions from our model into real-life bidding agents, which also need to consider all other available information about customers, publisher websites, advertisers' budgets, and advertising materials is still a challenge (Lee et al., 2013).

In summary, we demonstrate that bid request data is a very promising source of big data on the Internet that is worth further investigation by researchers at the intersection of IS and marketing.

References

Bucklin, R., and Sismeiro, C. (2009). "Click here for Internet insight: Advances in clickstream data analysis in marketing," In: *Journal of Interactive Marketing* (23:1), pp. 35–48.

Chatterjee, P., Hoffman, D., and Novak, T. (2003). "Modeling the clickstream: Implications for web-based advertising efforts," In: *Marketing Science* (22:4), pp. 520–541.

Chen, H., Chiang, R. H. L., and Storey, V. C. (2012). "Business Intelligence and Analytics: From Big Data to Big Impact," In: *Management Information Systems Quarterly* (36:4), pp. 1165–1188.

Chen, Y., and Berkhin, P. (2011). "Real-time bidding algorithms for performance-based display ad allocation," In: *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1307–1315.

Dinner, I., van Heerde, H. J., and Neslin, S. A. (2014). "Driving Online and Offline Sales: The Cross-channel Effects of Traditional, Online Display, and Paid Search Advertising," In: *Journal of Marketing Research* (51:5), pp. 527–545.

Duan, W., and Zhang, J. (2014). "The Impact of Referral Channels in Online Customer Journey," In: *Proceedings of the 35th International Conference on Information Systems, Auckland, New Zealand*.

Joo, M., Wilbur, K. C., and Zhu, Y. (2015). "Effects of TV advertising on keyword search," In: *International Journal of Research in Marketing, In Press, available online: http://dxdoiorg/101016/jijresmar 201412005*.

King, G., and Zeng, L. (2001). "Logistic regression in rare events data," In: *Political analysis* (2:1), pp. 137–163.

Kitts, B., Bardaro, M., Au, D., Lee, A., Lee, S., Borchardt, J., Schwartz, C., Sobieski, J., and Wadsworth-Drake, J. (2014). "Can Television Advertising Impact Be Measured on the Web? Web Spike Response as a Possible Conversion Tracking System for Television," In: *Proceedings of 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, New York, NY, USA*.

Kyung, M., Gill, J., Malay, G., and Casella, G. (2010). "Penalized regression, standard errors, and Bayesian lassos," In: *Bayesian Analysis* (5:2), pp. 369–412.

Lazer, D., Kennedy R., King, G., and Vespignani, A. (2014). "The Parable of Google Flu: Traps in Big Data Analysis," In: *Science* (343), pp. 1203–1205.

Lee, K., Jalali, A., and Dasdan, A. (2013). "Real Time Bid Optimization with Smooth Budget Delivery in Online Advertising," In: *Proceedings of the Seventh International Workshop on Data Mining for Online Advertising, Chicago, IL, USA*.

Lee, K., Orten, B., Dasdan, A., and Li, W. (2012). "Estimating conversion rate in display advertising from past erformance data," In: *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, Beijing, China*, pp. 768–776.

Liaukonyte, J., Teixeira, T., and Wilbur, K. C. (2015). "Television Advertising and Online Shopping," In: *Marketing Science* (34:3), pp. 311–330.

Nottorf, F., and Funk, B. (2013). "The economic value of clickstream data from an advertiser's perspective," In: *Proceedings of the 21st European Conference of Information Systems, Utrecht, Belgium*.

Pandey, S., Aly, M., and Bagherjeiran, A. (2011). "Learning to target: what works for behavioral targeting," In: *Proceedings of the 20th ACM international conference on Information and knowledge management, Glasgow, United Kingdom*, pp. 1805–1814.

Perlich, C., Dalessandro, B., and Hook, R. (2012). "Bid optimizing and inventory scoring in targeted online advertising," In: *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, Beijing, China*, pp. 804–812.

Plummer, M. (2003). "JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling," In: *Proceedings of the 3rd International Workshop on Distributed Statistical Computing, Vienna*.

Spann, M., Hinz, O., and Ramachandran, V. (2013). "Business and Information Systems Engineering and Marketing," In: *Business & Information Systems Engineering* (55:3), pp. 119–120.

Stange, M. (2015). "The Impact of TV Ads on the Individual User's Purchasing Behavior," In: *Proceedings of the 35th International Conference on Information Systems, Fort Worth, Texas*.

Stange, M., and Funk, B. (2014). "Real-Time Advertising," In: *Business & Information Systems Engineering* (56:5), pp. 305–308.

Stange, M., and Funk, B. (2015). "How much tracking is necessary?," In: *Proceedings of the 23rd European Conference on Information Systems, Muenster, Germany*.

Veit, D., Clemons, E., Benlian, A., Buxmann, P., Hess, T., Kundisch, D., Leimeister, J. M., Loos, P., and Spann, M. (2014). "Business Models," In: *Business & Information Systems Engineering* (6:1), pp. 45–53.

Yang, S., and Ghose, A. (2010). "Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence?" In: *Marketing Science* (29:4), pp. 602–623.

Yuan, S., Wang, J., and Zhao, X. (2013). "Real-time bidding for online advertising: measurement and analysis," In: *Proceedings of the Seventh International Workshop on Data Mining for Online Advertising, Chicago, IL, USA*.

Zhang, W., Yuan, S., and Wang, J. (2014). "Optimal real-time bidding for display advertising," In: *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining New York, NY, USA*.