http://dx.doi.org/10.15830/amj.2014.16.3.23

To Bid or Not to Bid?
- Keyword Selection in Paid Search Advertising

Yingying Ma*
Luping Sun**

The selection of keywords for bidding is a critical component of paid search advertising. When the number of possible keywords is enormous, it becomes difficult to choose the best keywords for advertising and then subsequently to assess their effect. To this end, we propose an ultrahigh dimensional keyword selection approach that not only reduces the dimension for selections, but also generates the top listed keywords for profits. An empirical analysis using a unique panel dataset from a large online clothes retailer that advertises on the largest search engine in China (i.e., Baidu) is presented to illustrate the usefulness of our approach.

Key words: Keyword Selection, Online E-commerce, Paid Search Advertising

Ⅰ. Introduction

With the development of the Chinese economy, internet has played a more and more important role in people's life. According to China Internet Network Information Center (CNNIC) Statistics, by the end of June, 2010, the number of Internet users in China had reached 420 million, which accounts for 31.8% of the whole population. Facing such a large populous, Internet is gradually surpassing traditional medias like TV and newspapers and become a key channel connecting companies and consumers. This gives rise to the fast growth of online advertising business. More recently, Internet search engine companies (e.g., Google, Yahoo and Baidu) have revolutionized not only the use of the Internet by individuals but also the way businesses advertise to consumers (Mehta et al., 2007). According to "CHINA SEARCH ENGENES -STATISTICS& FACTS", the number of search

* Guanghua School of Management Peking University(mayingying@pku.edu)
** Central University of Finance and Economics

engine users in China reached 451.1 million in 2012 and this number rose up to 564 million in China by the end of the year. It follows that 80 percent of the internet users used search engines. As of 2011, revenue of search engines stood at 18.9 billion RMB, which marked an increase of 71 percent from the previous year. Advertising revenue recorded a year-on-year growth of 67 percent to 17.2 billion RMB and provided the main source of income of search engines. Baidu was the leading search engine in China ranked by revenue, unique visitors, and page viewings as of 2012.

Along with the rapidly increasing number of search engine users, search engines have become one of the most important ways for consumers to obtain information, which make paid search advertising a "rising star" among all forms of advertising. Paid search advertising is so popular now that it is gaining ground as the largest source of revenues for search engines (Ghose and Yang, 2009). According to the "US Online Advertising Forecast" from 2013 to 2017, overall online ad spending will grow from $42.3 billion to $61.4 billion.

In paid search advertising, companies purchase specific keywords and create an advertisement that will be displayed alongside organic (non-sponsored) web search results when a consumer searches for those keywords. This form of online advertising is superior to the others and is especially appealing for companies for the following reasons. First of all, one thing that distinguishes paid search advertising from other forms of online advertising, such as online banner advertisements and pop-ups, is that instead of flooding consumers with unwanted ads, this particular type of advertising is based on customers' own queries and can directly reach consumers who are interested in the product or service or certain brands. These consumers are more likely to be interested in our products, brands or service. Therefore, this kind of advertising is considered to be more effective and less disturbing.

Aside from its special ability to reach the target consumers, another advantage of paid search advertising is that it is more cost effective. Instead of paying a set amount of fee for an advertisement and regardless of whether it has been noticed by our target consumers or not, a company using the paid search advertising, bids for specific keywords, and only be charged when a consumer clicks on its ads. The company spending is based on the number of times consumers click on its ads. Moreover, the company can also specify the maximum daily spending on each keyword so as to well control the budget. The behavior of what consumer does after they have clicked our website and entered in it can also be tracked, so it is able to check the effectiveness and profitability of its ads.

This paper focuses on the advertisers, who must find effective keywords and solve a complex profit optimization problem. With the development of various techniques to expand

keywords, it is no longer a problem for companies nowadays to find relevant keywords. But facing the enormous possible keywords generated in paid search advertising, selecting the keywords on which to bid is another important issue.

The most powerful and useful keywords are able to yield a greater number of clicks and improve the pay-per-click (PPC) conversion rate. As a result, advertiser's (or seller's) profits increase. When the total number of possible keywords is enormous, however, the practice of identifying the most profitable keywords is not a trivial task. In practice, for a targeting ad, the company usually generates thousands of keywords. Usually, we can find that some of the keywords are widely searched; still, many of them generate very little traffic. Thus the activity data becomes very sparse; evaluating the performance of these keywords is difficult, especially when we have very limited observations. For instance, in this paper, our focal company bids on over 1648 keywords. We focus on estimating the relationship between each day's conversion and the advertised keywords. Problems we are expecting to meet include that: the daily traffic generated by different keywords is correlated. It is reasonable as customers usually search a series of keywords, rather than only search once before finding their satisfied results. Another difficulty lies in that if we want to identify the relationship between each day's conversion and each keyword and put some features of each keywords into a linear model with each day's conversion as response variable, we will find that the number of observations is far less than the predictor dimension, resulting in a "large p, small n" problem and cannot be resolved by the traditional regression methods. One usual approach is to delete some features of the sparsely searched keywords. However, it is frustrating in practice because such keywords may still have strong potential to be good advertising investments. In real practice, we can find that if we stop auction such keywords, usually there will be a huge drop in each day's total conversion volume. Therefore, these keywords may still have contributed to each day's conversion and deleting these keywords may be unreasonable. In this paper, we use a novel approach to identify how each advertised keyword helps to generate conversions, and this novel approach can resolve an ultra dimensional problem. We will introduce this approach carefully in the methodology part.

In particular, we are dedicated to three unresolved problems in prior work. First, since we have too many keywords, to well control the budget and carry out profit optimization tasks, it is important for the company to find a suitable approach to select those truly useful keywords. Traditional statistical methods are infeasible as the number of observations is much smaller than the predictor dimension. Second, given a conversion, we want to identify how search-related factors affect the products that are

eventually purchased. We also aim to examine how keyword-related factors such as the presence of retailer or brand information in the keyword, the position of the advertisement on the search engine results page, and the extent of latency are associated with consumers' purchase intention across products. Finally, we will also examine how firm profits change with either price discount or the keyword advertisement content.

The rest of this paper is organized as follows. Initially, we briefly introduce the paid search data and discuss our modeling approach. Then, our ultra high dimensional model, the data set, and the empirical results will further be presented. After that, we discuss the implications of our findings for the management of paid search ads. Finally, we conclude the paper and discuss future research of search engine marketing.

Ⅱ. Data and Variables

Our data contains daily information on paid search advertising from an e-commerce website in the clothing industry. This company has advertised on Baidu (The largest Search Engine in China). The data spans all keyword advertisements by the company during a period of 51 days; specifically the data lasts from November 15 in 2010 to January 5 in 2011. Most datasets used in prior work to investigate consumer behavior in online environments usually comprise of browsing behavior only and associated clickstream data. Our data have recorded each individual keyword's performance, such as conversions, clicks, impressions and many other important keyword features, which will be introduced in the empirical part.

When we began collecting the data, the e-commerce company we collaborated has just opened for about 3 months and is still in the initial stage of a business. Thus, the company not only used its own brand or generic keywords as advertisements, but also some well-known brand names that can be searched by advertisers in order to refer consumers to its Web site. Hence, we also want to identify which keyword specific characteristics are more effective for the company when it advertises on a search engine. For example, can the company-specific information (e.g., name, website) in the keyword helps to generate more conversion? Or brand-specific information is more effective? Fully familiar of this pattern will help the firm to generate more useful keywords and have more keyword management insight. Here, we have introduced two keyword-specific characteristics "Brand" and "company" as our dummy variables. To be precise, if the advertising company's name present in the keyword, and then we labeled the company-dummy as 1 and otherwise 0. Similarly, the brand-dummy takes value of 1 if a famous brand name is present in the keyword and 0 otherwise.

Ⅲ. Method Motivation

Normally, when the number of observations significantly exceeds the number of predictors, estimation/evaluation of the effects of individual predictors is straightforward. However, when the number of predictors reaches the number of observations, estimation of individual parameters becomes problematic as the model becomes saturated. The problem of a large number of predictors coupled with a relatively small number of observations is referred to as the "large p, small n" problem (West 2003). As mentioned in the last section, one of our main focuses in this paper is to identify all relevant predictors that help to explain the conversion of the keywords. In practice, advertisers always find that the number of useful keywords is very large, since many keywords can more or less generate some conversions. Therefore, the variable selection approach we used should allow more than n predictors to be selected. Here n represents the number of observations. Past literature has revealed that customers usually have submitted a series of search queries before they finally make a purchase. Thus, many keywords may be highly correlated. Since our model predictors contain each keyword's text formation information and search related characters, thus, it is reasonable to assume that the predictors are correlated.

After these careful thinking, we have finally decided our approach as follows. Since the predictors dimension is ultrahigh, we firstly use an ultrahigh dimensional correlation screening method, which is called Sure Independence Screening (SIS) to sort these keywords. SIS is based on correlation learning, and it filters out the variables that have weak correlation with the response. Sure screening, a property of SIS, means to keep all the important variables after the variable screening process. SIS not only screens out those insignificant predictors, but also gives the left predictors a specific rank to indicate their importance. More important predictors will be ranked in upper places. The SIS's theory properties ensure to keep all the significant predictors. But, in the same time, many insignificant predictors may still be left inside. We further use a modified BIC to keep the significant variables and drop the insignificant ones. At this time, the predictor dimension may still be high; we then use a high dimensional ridge regression to estimate the parameters. After carefully examining those estimates, we can clearly understand what kind of keywords can help to generate more conversions.

Ⅳ. Empirical Analysis

In our model, the dependent variable $Y = (Y_1, \ldots Y_T)' \in \Re^{T\times 1}$ is a vector with the t-th component Y_t representing the number of con-

versions of the t-th day.

$click_k = (click_{k1},...,click_{kT})' \in \Re^{T\times 1}$ has T components with $click_{kt}$ representing the number of clicks of the k-th keyword on the t-th day. $imp_k = (imp_{k1},...,imp_{kT})' \in \Re^{T\times 1}$ also has T components with imp_{kt} representing the number of impressions of the k-th keyword on the t-th day. Similarly, we define $Rank_k = (Rank_{k1},...,Rank_{kT})' \in \Re^{T\times 1}$ as the k-th keyword's average rank on each day. The t-th component of $Onsite_k = (Onsite_{k1},...,Onsite_{kT})' \in \Re^{T\times 1}$ represents the k-th keyword's "total time on site on the t-th day divided by "$click_{kt}$". We use this value to reduce the correlation between the variable "$click_k$" and "$Onsite_k$". If "$click_k$" equals 0, then we also have $Onsite_k = 0$.

$Pageview_k = (Pageview_{k1},...,Pageview_{kT})' \in \Re^{T\times 1}$ and its t-th component represents the number of average page views for the k-th keyword on the t-th day. Similarly with the variable "$Onsite_k$", we also use "the total number of page views divided by $click_{kt}$" to operationalize the variable "$Pageview_k$", which is also intends to reduce the correlation. Here, $\mathrm{I} = (\mathrm{I}_1,...\mathrm{I}_T)' \in \Re^{T\times 1}$ is a dummy vector, and if the t-th component of I equals 1, it represents this day happens to be a weekend. We use I to indicate whether there exists a weekday effect or not. We then define $Discount = (discount_1,...,discount_T)' \in \Re^{T\times 1}$, where $Discount$ is a dummy vector with $discount_t = 1$ representing that there is a discount on the t-th day and $discount_t = 0$ representing otherwise. We use this variable to identify the effectiveness of discount. Then we specify the model as follows.

$$Y = \beta_0 + \beta_1 \mathrm{I} + \beta_2 Discount + \sum_{k=1}^{K} \alpha_k click_k + \sum_{k=1}^{K} \gamma_k imp_k + \sum_{k=1}^{K} \pi_k Onsite_k + \sum_{k=1}^{K} \psi_k Pageview_k + \varepsilon \quad (1)$$

$$\alpha_k = \alpha_{k0} + \alpha_{k1} S_{k1} + \alpha_{k2} S_{k2} + \alpha_{k3} imp_k + \alpha_{k4} Rank_k + \eta \quad (2)$$

$$\gamma_k = \gamma_{k0} + \gamma_{k1} S_{k1} + \gamma_{k2} S_{k2} + \gamma_{k3} Rank_k + \xi \quad (3)$$

ε, η, ξ are all random errors. S_{k1} and S_{k2} are dummy variables. If the k-th keyword contains the company's own information, then S_{k1}equals 1 and 0 otherwise.

Similarly, if the k-the keyword contains other famous brands' information, then S_{k2} equals 1 and 0 otherwise.

We then combine equations (1)-(3) together and use SIS (Fan &LV, 2012) to screen the variables. As we mentioned above, SIS is based on a correlation learning method, which filters out the variables that have no significant correlation with the response and give the relatively important variables a rank. After that, we use a modified BIC to select these truly important variables. Since there are candidate models with size larger than the observation number n, uniformly, we use high dimensional ridge regression to do parameter estimation and obtain the corresponding residual sum of squares, which will be used to calculate the BIC score.

Previous research has proved that using BIC to do variable selection, with probability tending to one, can keep all the important variables and it can dramatically narrow down the search for important predictors (Wang, 2009). After choosing the important variables (which form a much smaller sub model), we then apply the ridge regression to this sub model to make inferences on model parameters.

4.1 Estimation Results

Since the company has advertised 1627 keywords, which form a number of 6510 predictors. Based on our estimation results, we find that the estimated $\hat{\beta}_1$ and $\hat{\beta}_2$ are significant and positive, indicating the existence of weekday effect and that the promotions are effective to generate more purchases. We also notice that out of the huge number of predictors, only 1984 predictors have significant effect on the dependent variable (i.e., conversion) and their influences are quite different. Since the predictors are from certain features of the keywords, these significant predictors corresponds to 64.36% of the keywords, which means that about 35.64% of the keywords have no significant effect on conversion. From the estimated coefficients, we also find that about 40% of the1627 keywords have positive effect on conversion, as they have positive coefficients on impression, click, or some other features. Specifically, 69 keywords have positive values on α_{k0} and most of the other estimated coefficients (except for α_{k2}, α_{k3}, $\alpha_{k4}$and γ_{k3}) are also positive. This implies that when consumers search for the 69 keywords, the more customers click our website, the more likely they will purchase the company's products. The descriptive statistical analysis and estimation results for the 69 keywords are summarized in Table 1 and 2 respectively.

To further understand these keywords' statistical property, we have displayed the following keywords characters, like conversion, click, impression, price, rank, bounce, page-view, and time-on-site. "price" stands for the advertiser's allowed maximum cost for each click for one particular keyword. This variable can be used to identify the popularity of one particular

〈Table 1〉 Descriptive statistical analysis for the 69 star keywords

variable name	mean	Standard deviation	Max value
conversion	1.06	1.96	13
click	36.86	63.78	504
impression	1289.69	2489.11	12474
price	0.46	0.51	3
rank	3.08	2.69	24
bounce	11.23	22.48	231
pageview	1.50	1.04	3.62

〈Table 2〉 Coefficient estimation result for the 69 star keywords

Estimated Coefficient	The mean of estimated coefficient	Standard Deviation	Max Value	Percentage of Significance
β_0	0.08	0.00	0.08	100%
β_1	0.06	0.00	0.06	100%
β_2	0.70	0.00	0.70	100%
$\hat{\alpha}_{k0}$	4.41	0.35	12.82	100%
$\hat{\alpha}_{k1}$	1.39	0.31	4.52	69.57%
$\hat{\alpha}_{k2}$	-0.04	0.02	0.09	23.19%
$\hat{\alpha}_{k3}$	-0.01	0.29	0.02	42.03%
$\hat{\alpha}_{k4}$	-0.16	0.08	0.00	56.52%
$\hat{\gamma}_{k0}$	0.51	0.22	2.39	39.13%
$\hat{\gamma}_{k1}$	1.48	0.41	3.58	23.19%
$\hat{\gamma}_{k2}$	0.29	0.13	1.92	8.70%
$\hat{\gamma}_{k3}$	-0.14	0.27	0.00	15.94%
$\widehat{\pi_k}$	0.00	0.00	0.01	4.35%
$\widehat{\psi_k}$	0.01	0.03	0.12	8.70%

keyword. "pageview" measures the number of pages brought by one specific keyword in one day for each click. "rank" measures the keywords average position for each day. We record one click as a "bounce" when consumer clicks the website and close it less than 3 seconds. Thus, bounce can be used to measure how many clicks are useless.

As we can see from Table1, the average conversion for the 69 keywords equals 1.06, which means that averagely speaking, each keyword can generate one conversion for each day. Thus, these keywords are truly valuable, which can also be proved by their very high average pricing budget.

We use 4 measurement criteria to report the estimation results in Table 2. Specifically, "The mean of estimated coefficient" represents the mean estimation result for one type of predictors for the 69 keywords. Here, we use "Max" to report the largest estimator of one type of predictor for the 69 keywords. We further obtain the corresponding "Standard Deviation" for the 69 estimators. We use "Percentage of Significance" to record the significant percentage for each type of predictors. For example, if there are 36 significant $\hat{\alpha}_{k0}$ s, the corresponding "Percentage of Significance" value equals 52.17%.

The estimations for α_{k4} and γ_{k3} ($\hat{\alpha}_{k4}$ and $\hat{\gamma}_{k3}$) are either negative or zero for the 69 keywords, which implies that upper places can help these keywords to generate more conversions. The

mean of $\hat{\alpha}_{k1}$ is positive, which indicates that having the company's information in the keywords can help to generate more conversions. On the other hand, the mean of $\hat{\alpha}_{k2}$ is negative, which implies that containing other famous brands' information in the keywords may not directly help to generate conversions. We further notice that the percentage of significant estimators for both π_k and ψ_k are very low and the mean of $\hat{\pi_k}$ very close to 0. This suggests that when customers click the company's website through such keywords, how long they stay in the website have no significant relationship with their purchase behavior.

We further notice that there are 369 keywords have positive $\hat{\gamma}_{k0}$, but very few of these keywords have significant estimation for α_{k0}. Thus, we make influence that these keywords help to generate conversions mainly through the effect of impression. The descriptive statistical analysis for these 369 keywords is summarized in Table 3. As we can see from Table 3, the mean impressions of these keywords are very large, but the mean of conversions are very close to 0. In real practice, we try to stop auctioning for such keywords, but later we will find a significant shrinkage in each day's total conversion volume. This phenomenon and the coefficient estimation results together suggest that the 369 keywords might help to generate more conversions through an indirect way. Since these keywords have considerable number of impression, this indicates that many consumers are searching for such keywords. These consumers may not search for our products, as we observe the mean of clicks for these keywords is not very high. However, when customers search for these keywords and this company's ads show up on the search engine, the company's brand or products will leave an impression on the customers. If customers cannot find suitable brand/ products via the website they click, they may search again and this gives this company's website a second chance to be visited. For example, when customer search for "Nike shoes" and this company's ads about shoes appear on the search engine.

If the customer couldn't find satisfied results

〈Table 3〉 Descriptive statistical analysis for the 369 bridge keywords

variable name	mean	Standard deviation	Max value
conversion	0.069	0.30	4
click	6.92	17.44	148
impression	325.95	1257.41	24320
price	0.30	0.44	2
rank	2.94	3.46	21.31
bounce	2.69	6.92	58
pageview	0.89	1.02	4.11

from his/her first click, they may search again and the company's shoes can gain an opportunity to be purchased. Therefore, such keywords indirectly help to generate more conversions and auctioning for these keywords is necessary.

We also notice that there are 78 keywords have no significant estimated coefficients and these keywords can more or less generate some clicks. Since we have to pay for the clicks generated by these keywords, a rational suggestion is to reduce the auctioning budget for such keywords. The descriptive statistical analysis for such keywords has been summarized in Table 4. Compared with Table 1 and 3, we find that keywords in Table 4 have the most average forward position, and the average bounce for such keywords is also very high. Since a bounce means that this click is useless but to ruin the budget. Therefore, reducing the click budget for such keywords is highly suggestible.

V. Conclusions and Discussions

While most of the previous research is conducted from the perspective of the Internet companies, this paper focused on the advertisers. Specifically, we proposed a new procedure to estimate an ultrahigh dimensional model. Through this procedure, we can get the coefficient estimation corresponding to each keyword. This will help us to identify how each keyword help to generate conversions. Specifically, we have identified three types of keywords, namely, the star keywords, bridge keywords and insignificant keywords. When we auction the star keywords, we should pay more attention to help these keywords to generate more clicks, as more clicks will greatly help these keywords to generate more conversions. For the bridge keywords, they help to generate more conversions through their impressions. The considerable number of impressions of these keywords can help to leave customers an impression on the

〈Table 4〉 Descriptive statistical analysis for the 78 insignificant keywords

variable name	mean	Standard deviation	Max value
conversion	0.00	0.01	1
click	0.64	4.95	421
impression	53.97	845.30	72031
price	0.17	0.22	2
rank	0.75	2.62	104
bounce	0.31	2.56	211
pageview	0.17	0.54	4.34

company's products or brand. Eventually, the customers will get familiar with this brand. The insignificant keywords have no significant effect on conversions, but they may still generate clicks, which will ruin the final profit. Therefore, we should reduce the auctioning budget for suck keywords.

〈Received May 29. 2014〉
〈Revised August 25. 2014〉
〈Accepted October 18. 2014〉

References

Chen, J. and Chen, Z. (2008), Extended Bayesian information criterion for model selection with large model spaces, Biometrika, Vol.95, Issue 3.

George, E. I. (2000), The variable selection problem, *Journal of the American Statistical Association*, Vol.95, Issue 452.

Mehta, A., Saberi, A., Vazirani, U. and Vazirani, V. (2007), Adwords and Generalized Online Matching, *Journal of the ACM*, Vol. 54, Issue 5.

Fan, J. and Lv, J. (2008), Sure independence screening for ultra-high dimensional feature space (with discussion), *Journal of the Royal Statistical Society*, Series B, Vol.70, Issue 5.

Ghose, A. and Yang, S. (2009), An empirical analysis of search engine advertising: sponsored search in electronic markets, *Management Science*, Vol. 55, Issue 10.

Tibshirani, R. (1996), Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society*, Series B, Vol. 58, Issue 1.

West, M. (2003), Bayesian factor regression models in the "large p, small n" paradigm: Bayesian Statistics, Proc. Seventh Valencia Internat. Meeting. Oxford University Press, New York, 723-732.

Wang, H., (2009), Forward regression for ultra-high dimensional variable screening, *Journal of the American Statistical Association*, Vol 104, Issue 488.

Zou, H. and T. Hastie. (2005), Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society*, Series B, Vol. 67, Issue 2.