Electronic Commerce Research and Applications

Keyword Targeting Optimization in Sponsored Search Advertising: Combining Selection and Matching

--Manuscript Draft--

Manuscript Number:	ECRA-D-22-00216
Article Type:	Research Paper
Keywords:	Keyword targeting; Keyword matching; Keyword selection; Sponsored search advertising; Stochastic programming

Abstract:	In sponsored search advertising (SSA), advertisers need to select keywords and determine matching types for selected keywords simultaneously (i.e., keyword targeting), in order to effectively reach the right population. This paper aims to address the keyword targeting problem, which is a challenging task because of the incomplete information and high uncertainty in SSA environments. First, we construct a data distribution estimation model for keyword performance indices (i.e., impression and click-through rate) over three keyword matching options, and apply the Markov Chain Monte Carlo method to compensate for unobserved indices. Second, we formulate a stochastic keyword targeting model combining keyword selection and keyword matching operations to maximize the expected profit under the chance constraint of budget. Then we develop a branch-and-bound solution algorithm incorporating a stochastic simulation process for our keyword targeting model. Furthermore, based on realworld datasets collected from field reports and logs of SSA campaigns, computational experiments are conducted to evaluate the performance of our keywords targeting strategy. Experimental results show that, (a) our keyword targeting strategy outperforms four baselines in terms of profit; (b) keyword targeting optimization shows its superiority as the budget increases, especially in situations with more keywords and keyword combinations; (c) the data distribution estimation effectively addresses the problem of incomplete information over keyword matching options and significantly promotes the performance of keyword targeting. These results offer critical insights into keyword management for SSA advertisers.

Highlights

- This paper aims to address the keyword targeting problem in SSA.
- We construct a data distribution estimation model for keyword performance indexes (i.e., impression and click-through rate) over three keyword matching options, and apply Markov Chain Monte Carlo computational methods to compensate for unobserved indexes.
- Based on the above estimation, we formulate a stochastic keyword targeting model combining keyword selection and keyword matching in an integrated way, and develop a branch-and-bound algorithm combined with the stochastic simulation process to solve our keyword targeting model.
- We conduct computational experiments to evaluate the performance of our keywords targeting strategy with real-world data set collected from field reports and logs of sponsored search advertising campaign.

Keyword Targeting Optimization in Sponsored Search Advertising: Combining Selection and Matching

Huiran Li[1] and Yanwu Yang[2]
[1]School of Business Administration and Customs Affair, Shanghai Customs College, Shanghai 201204, China
[2]School of Management, Huazhong University of Science and Technology, Wuhan 430074, China
{lihuiran.isec, yangyanwu.isec}@gmail.com

Abstract: In sponsored search advertising (SSA), advertisers need to select keywords and determine matching types for selected keywords simultaneously (i.e., keyword targeting), in order to effectively reach the right population. This paper aims to address the keyword targeting problem, which is a challenging task because of the incomplete information and high uncertainty in SSA environments. First, we construct a data distribution estimation model for keyword performance indices (i.e., impression and click-through rate) over three keyword matching options, and apply the Markov Chain Monte Carlo method to compensate for unobserved indices. Second, we formulate a stochastic keyword targeting model combining keyword selection and keyword matching operations to maximize the expected profit under the chance constraint of budget. Then we develop a branch-and-bound solution algorithm incorporating a stochastic simulation process for our keyword targeting model. Furthermore, based on realworld datasets collected from field reports and logs of SSA campaigns, computational experiments are conducted to evaluate the performance of our keywords targeting strategy. Experimental results show that, (a) our keyword targeting strategy outperforms four baselines in terms of profit; (b) keyword targeting optimization shows its superiority as the budget increases, especially in situations with more keywords and keyword combinations; (c) the data distribution estimation effectively addresses the problem of incomplete information over keyword matching options and significantly promotes the performance of keyword targeting. These results offer critical insights into keyword management for SSA advertisers.

Keywords: Keyword targeting, Keyword matching, Keyword selection, Sponsored search advertising, Stochastic programming

1. Introduction

Sponsored search advertising (SSA) has become one of the most indispensable digital media channels. In the United States, SSA spending is projected to reach $171,641 million in 2021 (Statista 2021; IAB 2021). SSA is a prosperous market with three types of players, i.e., advertisers, search users and search engines, where keywords serve as a bond tying all three together (Yang et al., 2019). In SSA, advertisers have to select an appropriate set of keywords and determine suitable matching types for selected keywords simultaneously. This process is called keyword targeting (Yang et al., 2019). Keyword targeting controls the aggressive and restrictive degree for consumers' searches triggering sponsored search auctions, and helps advertisers better fit their promoted products to a particular search engine (Kiritchenko and Jiline, 2008; Google Ads, 2021). Well-targeted keywords will guarantee that the right advertisements are delivered to the right consumers (Yang et al., 2017). Therefore, it is critical for advertisers to effectively make keyword targeting decisions for their SSA campaigns.

In the literature, plenty of research efforts have been invested in formulating keyword selection models and developing corresponding solution algorithms (Rusmevichientong and Williamson, 2006; Kiritchenko and Jiline, 2008; Zhang et al., 2014), analyzing branded and competitor's keywords (Desai et al., 2014), and examining the keyword performance in SSA markets (Lu and Yang, 2017). In another independent research stream, keyword matching has been studied extensively in recent years from various aspects, such as identifying high-quality match mapping keywords (Radlinski et al., 2008; Gupta et al., 2009; Grbovic et al., 2016), profiling advertising metrics over matching options (Ramaboa and Fish, 2018), and bidding optimization under the context of broad match (Singh and Roychowdhury, 2008; Even Dar et al., 2009; Amaldoss et al., 2016). Operationally, it's of necessity to address keyword selection and keyword matching problems in an integrated way (i.e., keyword targeting), in order to help advertisers effectively reach the targeted population. However, to the best of our knowledge, there is no study on keyword targeting for search advertising campaigns. This paper attempts to fill this crucial gap.

In SSA, advertisers have to face many challenges while making keyword targeting decisions. First, advertisers have no complete performance information over the three keyword matching types for each keyword. In practice, the keyword-level historical records only contain performance indices (e.g., impressions and click-through rate) for a certain matching type chosen in advertising campaigns (Google Ads, 2021). However, advertising performance indices are systematically different over the three keyword matching types (Ramaboa and Fish, 2018; Yang et al., 2021a). When making keyword selection decisions, advertisers have to take the uniform assumption about performance indices over the three keyword matching types, which certainly results in suboptimal solutions. Second, the SSA environment is highly uncertain (Yang et al., 2013; Li and Yang, 2020). In such an uncertain market, advertisers must make keyword targeting decisions prior to the realization of values for keyword performance indices (Amaldoss et al., 2016). In addition, advertisers face budget constraints when making keyword-related decisions (Yang et al., 2015). That is, advertisers need to select appropriate keywords and matching options for maximizing the SSA profit with a limited budget.

This research aims to address the keyword targeting problem in the SSA context. First, we construct a data distribution estimation model for keyword performance (i.e., impression and click-through rate). It is supposed that performance indices follow the multivariate normal distribution over the three types of matching options. The Markov Chain Monte Carlo method is applied to compensate for unobserved performance indices over the three matching types. Second, we formulate a stochastic keyword targeting model to maximize the expected profit under the chance constraint of budget. Moreover, we develop a branch-and-bound solution algorithm incorporating a stochastic simulation process for our keyword targeting model. Furthermore, using a realworld dataset collected from field reports and SSA campaign logs, a series of computational experiments are conducted to evaluate the performance of our keyword targeting strategy against four baselines.

Experimental results show that (a) our keyword targeting approach performs better than four baselines in terms of profit; (b) our keyword targeting approach increasingly shows its superiority as the budget increases, especially in situations with more keywords and keyword combinations; (c) the data distribution estimation effectively addresses the problem of incomplete information over keyword matching options and significantly promote the performance of keyword targeting. In keyword targeting, compared with the single matching option, mixed keyword matching option enriches keyword portfolios and increases the expected profit. Our strategy can help advertisers find more high-profit yet less-cost keywords with appropriate matching options by searching the keyword targeting space for the global optimum.

These results suggest critical managerial insights into keyword decisions for advertisers in SSA. First, keyword targeting is a vitally important advertising decision. Advertisers with a larger number of available keywords and keyword combinations should pay more attention on keyword targeting. Second, various factors (e.g., selected keywords, matching types, the control of uncertainty and the budget) have influences on the performance of keyword targeting. Advertisers need to comprehensively evaluate and utilize these factors to get an optimal keyword targeting decision. Third, in the optimal keyword targeting solutions, the exact and phrase matching types take up considerable proportions. This challenges prior studies (e.g., Ballard, 2013; Amaldoss et al., 2016) where broad match is the most popular option. This reminds advertisers to explore keyword targeting strategies with mixed matching options. Our findings could enhance the understanding of keyword research and targeting decisions in SSA.

The contributions of this study can be summarized as follows. From an academic perspective, to the best of our knowledge, this is the first research effort on keyword targeting problem in SSA. This study adds to the advertising optimization literature in SSA by exploring keyword targeting decisions in the framework of stochastic optimization under the chance constraint of budget. Moreover, we proposed a data distribution estimation approach using a fully Bayesian model to address the problem of incomplete information over keyword matching options. From the methodology perspective, we propose a stochastic keyword targeting model under the chance constraint of budget and develop a corresponding branch-and-bound algorithm solution. This provides a feasible way for advertisers to make keywords targeting decisions in practice. In addition, our research can be generalized to keyword decisions in other keyword-based advertising forms.

The remainder of this paper is structured as follows. Section 2 provides a brief literature review in the next section. In Section 3, we build a data distribution estimation model and keyword targeting stochastic model in SSA, and give solution algorithms for the models. In Section 4, we conduct computational experiments and report the comparison results. Finally, we conclude this research in Section 5.

2. Related Work

In SSA, plentiful research have been made to explore sponsored search mechanism design (Huang and Kauffman, 2011; Yang et al., 2020) and search user behavior analysis (Lo et al., 2014; Vragov et al., 2019; Lian et al. 2021), empirical study on performance indices (Yang et al., 2018; Schultz, 2020; Yang and Zhai, 2022), and advertising decisions including bidding optimization (Küçükaydin et al., 2020; Kim et al., 2021), budget optimization (Yang et al., 2012; Park, 2020;

Yang and Xiong, 2020; Avadhanula et al., 2021; Yang et al., 2021b; Balseiro et al. 2021), keyword optimization (Qiao et al., 2017; Nie et al., 2019; Scholz et al., 2019; Song et al., 2021; Zhang et al., 2021). This study focuses on one particular type of keyword decisions, i.e., keyword targeting, which draws from two research streams, namely keyword selection and keyword matching.

2.1 Keyword Selection

Keyword selection is the basis for the effectiveness of search advertising campaigns (Szymanski and Lininski 2018). As a result, researchers have invested efforts to address the keyword selection problem through semantic mapping (Kiritchenko and Jiline, 2008; Arroyo-Cañada and Gil-Lafuente, 2019; Nagpal and Petersen 2020) and optimization techniques (Rusmevichientong and Williamson, 2006; Zhang et al., 2014; Yang et al. 2019). Based on the feature selection paradigm, Kiritchenko and Jiline (2008) analyzed the past performance of individual words and phrases generated from users' queries and selected the most promising keywords extended with highly predictive (positive and negative) words to maximize the profit, and showed that their approach can obtain high-quality keywords and discovered more specific combinations of keywords. Arroyo-Cañada and Gil-Lafuente (2019) developed a TOPSIS-based method, which sorted the keywords according to their distance to the positive and negative ideal solutions, and its effectiveness was proved in the aspect of increasing brand awareness and traffic volumes. Nagpal and Petersen (2020) constructed a conceptual framework to identify profitable keywords, controlling the endogeneity of competition to measures keyword relevance.

Keyword selection decisions can also be taken as optimization problems. For example, Rusmevichientong and Williamson (2006) identified profitable sets of keywords by sorting keywords in the decreasing order of profit-to-cost ratio, and formulated the keyword selection as a multi-armed bandit problem, taking into account the uncertainty of the click-through rate. Under budget constraints, Zhang et al. (2014) took the keyword selection as a mixed integer programming problem with objectives of maximizing the profit and the relevance of selected keywords and minimizing the competitiveness of these keywords, and presented a sequential quadratic programming solver. Based on a real SSA log dataset, they showed that the proposed method can help advertisers and search engines increase revenue. With consideration of the entire lifecycle of SSA campaigns, Yang et al. (2019) developed a multilevel keyword optimization framework to handle various keyword decisions (e.g., keyword generation, keyword selection and keyword assignment), showed that the proposed framework could reach the optimum in a steady way.

Another research branch focused on empirical analysis for keyword selection, e.g., analyzing the competitor's keywords (Desai et al., 2014) and the SSA market (Lu and Yang, 2017). In order to understand the strategic benefits and costs of selecting keyword about an advertiser's own brand name and her competitor's brand name, Desai et al. (2014) modeled the effect of SSA depend on whether a competitor's advertisement is presented on the same results page, and found that selecting the keywords about her own brand name can preclude her competitor from buying the same keywords, while if both the advertiser and her competitor select her brand name, a prisoner's dilemma may be created to hurt both of their profit. In order to select keywords in a more strategic manner, Lu and Yang (2017) regarded each keyword as a market and developed a structural model to empirically investigate the spillover effects in an advertiser's keyword market entry decisions, i.e., the probability that an advertiser uses a keyword is affected by her competitor's keyword decisions, and demonstrated that the search engine revenue can be improved by around 5.7 percent with the keyword-specific competition information.

2.2 Keyword Matching

Our work also builds on the literature on keyword matching in SSA. Choosing a suitable keyword matching type for each keyword is essential to the success of advertising campaigns (Ghose and Yang, 2009; Li et al., 2016; Ramaboa and Fish, 2018). In general, there are three keyword matching types (i.e., broad match, phrase match and exact match) to choose in SSA, as shown in Table 1.

Table 1. Three Keyword Matching Types

Match Type	Matches to:	Example Keywords	Matches to:
Broad match[1]	searches that include misspellings, synonyms, related searches, and other relevant variations	internet advertising	Adwords
Phrase match	searches that match a phrase, or close variations of that phrase, which may include additional words before or after	internet advertising	internet advertising in China
Exact match	searches that match the exact term or are close variations of that exact term	internet advertising	internet advertising

[1] In Google Ads, there exists another type of matching option, "Broad match modifier", which is similar to broad match, except that the broad match modifier option only shows ads in searches that includes the words with a plus sign. We don't distinguish broad match and broad match modifier in this research.

In the field of keyword matching, most research focus on broad match. On one hand, broad match helps advertisers extend keywords that match user's intent expressed by the query (Radlinski et al., 2008). On the other hand, broad match helps search engine engage more competitors in its auctions through broader targeting, thus increasing its revenue in SSA (Levin and Milgrom, 2010). The existing literature in this area explored broad match from both direct and indirect perspectives in SSA. One research stream studied broad match directly by presenting effective broad match mapping mechanism, helping advertisers identify similar keywords and thus increasing their advertising reach and reducing their campaign management burden (Jones et al., 2006; Radlinski et al., 2008; Gupta et al., 2009; Grbovic et al. 2016). Some approaches require offline training, i.e., under the assumption of available human supervisions (Jones et al., 2006; Radlinski et al., 2008). Broad match models are trained relying on human-labeled relevance judgments using similarity characteristics to identify associated keywords. Other approaches don't assume the availability of human supervision. For example, Gupta et al. (2009) proposed a machine learning approach through identifying high-quality broad match mappings to estimate click-through rate, and utilized supervision in the form of implicit feedback from the advertisement click-through logs, replacing the need for expensive human labeling. Further, go beyond the supervised learning approaches, Grbovic et al. (2016) proposed a matching strategy based on the semantic embeddings to learn queries and ads from SSA datasets, which were learned from the data of user's search sessions, including search queries, clicked ads and search links, and contextual information, including dwell time and skipped ads, and showed that their approach had great performance in terms of relevance, coverage, and the growth of profit. Another research stream studied broad match indirectly by addressing optimization problems in SSA under the context of broad match (Singh and Roychowdhury, 2008; Even Dar et al., 2009; Mahdian and Wang, 2009; Amaldoss et al., 2016). Sponsored search auction allows advertisers to target a large amount of queries but only bid a few keywords under the support of broad match. Singh and Roychowdhury (2008) provided a framework to make definite statements about economic outcomes of broad match, and observed that if the quality of broad match is good, the auctioneer (i.e., search engine) can always improve her revenue by judiciously using broad match. Mahdian and Wang (2009) developed a clustering-based bidding language to place broad-matched bids to reduce the complexity of bidding and avoid the negative economic effect of broad match. Researchers at Google addressed the query dependence and implicit bid optimization under the broad match characterized in sponsored search auctions (Even Dar et al., 2009), and developed a linear programming model on the basis of a polynomial time algorithm which can obtain the optimal profit. Focusing on equilibrium analysis of broad match, Amaldoss et al. (2016) examined advertisers' strategies and profits under broad match using a game-theoretic model, and showed that the accuracy degree of the search engine would be increased up to the point where advertisers choose broad match, and the search engine gets their highest profits.

However, it is necessary to specify keyword matching types before conducting their empirical analysis and attribution strategies optimization in SSA (Ghose and Yang, 2009; Li et al., 2016; Jeziorski and Moorthy, 2018). Ramaboa and Fish (2018) analyzed the influence of various search advertising metrics (i.e., the length, click-through rate, cost-per-click, position, and quality score) over three matching options, and showed that as the keyword matching types become narrower, almost all the keyword performance indices increase, but there is no significant change in cost per click. Du et al. (2017) constructed a hierarchical Bayesian framework to empirically study the influence of keyword matching types on advertising performance, and demonstrated the importance of keyword selection in SSA. Yang et al. (2021a) empirically explored the relationship between matching types and advertising performance, and showed that exact match led to better performance than broad match.

2.3 Summary

Table 2 summarizes the related literature discussed above from aspects of techniques, whether it considers keyword selection, includes three matching types, and provides solutions.

Table 2. Keyword Targeting Related Research

Reference	Techniques	Keyword Selection	Keyword Matching Types	Solution
Rusmevichientong and Williamson (2006)	Multi-armed bandit	Yes	None	Yes
Kiritchenko and Jiline (2008)	Feature selection paradigm	Yes	None	Yes
Zhang et al. (2014)	Binary integer programming	Yes	None	Yes
Arroyo-Cañada and Gil-Lafuente (2019)	TOPSIS-based method	Yes	None	Yes
Yang et al. (2019)	Multilevel closed-form computational framework	Yes	None	Yes

Nagpal and Petersen (2020)	Conceptual model	Yes	None	Yes
Desai et al. (2014)	Analytical model	Yes	None	No
Lu and Yang (2017)	Structural model	Yes	None	No
Jones et al. (2006)	Retrieval models, Machine learning	No	Broad match	Yes
Radlinski et al. (2008)	Two-phase heuristic methodology	No	Broad match and exact match	Yes
Gupta et al. (2009)	Online learning algorithm	No	Broad match	Yes
Grbovic et al. (2016)	Semantic embeddings	No	Broad match	Yes
Singh and Roychowdhury (2008)	Polynomial time algorithm	No	Broad match	Yes
Mahdian and Wang (2009)	Approximation algorithms	No	Broad match	Yes
Even Dar et al. (2009)	Polynomial time algorithm	No	Broad match and exact match	Yes
Amaldoss et al. (2016)	Game-theoretic model	No	Broad match and exact match	No
Ramaboa and Fish (2018)	Descriptive statistics, Hypothesis testing	No	All three types	No
Du et al. (2017)	Hierarchical Bayesian model	Yes	All three types	No
Yang et al. (2021a)	Causal inference model	Yes	All three types	No
This study	MCMC, Stochastic Programming	Yes	All three types	Yes

As illustrated in Table 2, prior research explored either keyword selection or keyword matching separately, while ignoring the necessity to addressing keyword selection and keyword matching problems in an integrated manner. This study fills the research gap by providing an effective optimization strategy for keyword targeting, to select keywords and determines matching types simultaneously. To the best of our knowledge, this is the first study in this direction. Moreover, we contribute to the SSA literature by estimating data distribution for keyword performance indices (i.e., impression and click-through rate) over three keyword matching options. Specifically, we construct a data distribution estimation model, and apply the Markov Chain Monte Carlo method to compensate for unobserved performance indices. In addition, we contribute to the optimization literature by formulating a stochastic keyword targeting model combining keyword selection and keyword matching operations.

3. The Model and Solution

This section is split into two stages, i.e., data distribution estimation for keyword performance indices (i.e., impression and click-through rate) and keyword targeting stochastic optimization. In the first stage, we apply Markov Chain Monte Carlo computational methods (Chen et al., 2000) to make inferences about unobserved keyword performance indices over three matching options. In the second stage, based on the data estimation results, we construct a stochastic keyword targeting model combining keyword selection and keyword matching operations, and develop a branch-and-bound algorithm incorporating a stochastic simulation process to solve our keyword targeting model. The notations used in this paper are listed in Table 3.

Table 3. Notations

Notation	Definition
$d_{k,j,i}$	the impression of the keyword k in adgroup j under matching type i
$c_{k,j,i}$	the click-through rate (CTR) of the keyword k in adgroup j under matching type i
$v_{k,j}$	the value-per-click (VPC) of the keyword k in adgroup j
$p_{k,j}$	the cost-per-click (CPC) of the keyword k in adgroup j
$\theta_j^d(\theta_j^c)$	the 3-dimensional mean vector of the multivariate normal distribution for impression (or CTR) in adgroup j
$\Sigma_j^d(\Sigma_j^c)$	the 3×3 covariance matrix of the multivariate normal distribution for impression (or CTR) in adgroup j
$\mu_0^d(\mu_0^c)$	the prior mean of θ_j^d for impression (or CTR)
$\Lambda_0^d(\Lambda_0^c)$	the prior variance of θ_j^d for impression (or CTR)
$\nu_0^d(\nu_0^c)$	the prior scalar parameter of Σ_j^d for impression (or CTR)
$S_0^{d^{-1}}(S_0^{c^{-1}})$	the prior matrix parameter of Σ_j^d for impression (or CTR)
$\mathrm{x}_{k,j,i}$	the binary decision variable indicated whether the keyword k in adgroup j choosing matching type i
B	the SSA campaign budget for an advertiser
α	the prescribed probability of budget chance constraint
n_j	the number of keywords in adgroup j
m	the number of adgroups

3.1 Data Distribution Estimation for Keyword Performance Indices

In SSA, advertisers choose a matching type for each selected keyword, then keyword performance indices only under the chosen matching option can be observed. It results in a problem of incomplete information about performance indices over keyword matching options. Thus, when making keyword targeting decisions, advertisers first have to estimate keyword performance indices over the three matching types. The data distribution estimation scenario can be described as: given a set of keywords and performance indices (i.e., impression and click-through rate) for each keyword under a certain matching option, how to estimate unobserved performance indices for each keyword over the other two keyword matching options?

3.1.1 Data Distribution Estimation Model

We use a fully Bayesian model to jointly simulate distributions of keyword performance indices with unobserved data over all the three matching options (Carrigan et al., 2007). Different matching options lead to different advertising performance for keywords (Google Ads, 2021). This phenomenon was emphasized by Ramaboa and Fish (2018) stating that the traffic and click-through rate (CTR) change with matching options, although the cost-per-click (CPC) doesn't. Thus, we assume that keyword impression and CTR are significantly influenced by keyword matching option. Let $d_{k,j,i}$ denote the impression of the keyword k ($k = 1,2, \dots, n$) in adgroup j ($j = 1,2, \dots, m$) under matching option i in a search market, $i = 1,2,3$ indicate exact match, phrase match and broad match, respectively. Let $c_{k,j,i}$ denote the click-through rate (CTR) of the keyword k in adgroup j under matching option i. Since the cost-per-click (CPC) and value-per-click (VPC) are not significantly influenced by different matching options, we denote the cost-per-click (CPC) and value-per-click (VPC) of the keyword k in adgroup j as $p_{k,j}$ and $v_{k,j}$, respectively.

For keywords in the same ad group focus on the common promotional product (or service) are closely related (Google Ads, 2021), we can assume that the performance indices (i.e., impression or CTR) for keywords in the same adgroup over the three matching options follow the same multivariate normal distribution. Then, the keyword performance indices distribution estimation problem can be transferred into the adgroup performance indices distribution estimation problem. Specifically, given an advertising campaign with m adgroups where there are n_j keywords in adgroup j ($j = 1, \dots, m$), we denote the impressions for keyword k over the three matching options $i = 1,2,3$ in the same adgroup j as $d_{k,j} = (d_{k,j,1}, d_{k,j,2}, d_{k,j,3})^T$, $k = 1, \dots, n_j$.

Because the impression is non-negative $d_{k,j,i} \geq 0$, we use log function to transform the impression value in the interval $[0, +\infty]$ to the real line $[-\infty, +\infty]$, then the $d'_{k,j} = (d'_{k,j,1}, d'_{k,j,2}, d'_{k,j,3})^T = \left(log(d_{k,j,1}), log(d_{k,j,2}), log(d_{k,j,3})\right)^T$, thus $d'_{k,j} \sim MVN(\theta_j^d, \Sigma_j^d)$, where $MVN(\theta_j^d, \Sigma_j^d)$ is a multivariate normal distribution, where $\theta_j^d = (E\left[d'_{j,1}\right], E\left[d'_{j,2}\right], E[d'_{j,3}])^T$ is a 3-dimensional mean vector, and Σ_j^d is a 3×3 covariance matrix. Then the adgroup impression distribution estimation model is

$$d'_j = \{d'_{1,j}, d'_{2,j}, \dots, d'_{n_j,j}\} \sim MVN(\theta_j^d, \Sigma_j^d)$$
$$\theta_j^d = E\begin{bmatrix} d'_{j,1} \\ d'_{j,2} \\ d'_{j,3} \end{bmatrix}, \Sigma_j^d = E\begin{bmatrix} \Sigma_{j,11}^d & \Sigma_{j,12}^d & \Sigma_{j,13}^d \\ \Sigma_{j,21}^d & \Sigma_{j,22}^d & \Sigma_{j,23}^d \\ \Sigma_{j,31}^d & \Sigma_{j,32}^d & \Sigma_{j,33}^d \end{bmatrix} \tag{1}$$
$$j = 1,2, \dots, m.$$

In Bayesian statistics, the conjugate prior of the mean vector θ_j^d for the multivariate normal distribution (MVN) of impression over three matching options is another MVN, and the conjugate prior of the covariance matrix Σ_j^d for the distribution of impression over three matching options is an inverse-Wishart distribution,

$$\theta_j^d \sim MVN(\mu_0^d, \Lambda_0^d) \tag{2}$$

$$\Sigma_j^d \sim inverseWishart(\nu_0^d, {S_0^d}^{-1}) \tag{3}$$

where μ_0^d and Λ_0^d are the prior mean and variance of θ_j^d for impression, respectively; and ν_0^d is a scalar and ${S_0^d}^{-1}$ is a matrix.

Analogously, we denote the click-through rate (CTR) for keyword k over the three matching options $i = 1,2,3$ in the adgroup j ($j = 1, \dots, m$) as $c_{k,j} = (c_{k,j,1}, c_{k,j,2}, c_{k,j,3})^T$, $k = 1, \dots, n_j$. We use logit function (i.e., the inverse of the sigmoid function) to transform the CTR value in the interval $[0,1]$ to the real line $[-\infty, +\infty]$, then $c'_{k,j} = (c'_{k,j,1}, c'_{k,j,2}, c'_{k,j,3})^T = \left(logit(c_{k,j,1}), logit(c_{k,j,2}), logit(c_{k,j,3})\right)^T$ obey multivariate normal distribution $MVN(\theta_j^c, \Sigma_j^c)$, where $\theta_j^c = (E\left[c'_{j,1}\right], E\left[c'_{j,2}\right], E[c'_{j,3}])^T$ is 3-dimensional mean vector, and Σ_j^c is a 3×3 covariance matrix. Then the adgroup CTR distribution estimation model follows the same structure as the impression distribution estimation model,

$$c'_j = \{c'_{1,j}, c'_{2,j}, \dots, c'_{n_j,j}\} \sim MVN(\theta_j^c, \Sigma_j^c)$$

$$\theta_j^c = E\begin{bmatrix} c'_{j,1} \\ c'_{j,2} \\ c'_{j,3} \end{bmatrix}, \Sigma_j^c = E\begin{bmatrix} \Sigma_{j,11}^c & \Sigma_{j,12}^c & \Sigma_{j,13}^c \\ \Sigma_{j,21}^c & \Sigma_{j,22}^c & \Sigma_{j,23}^c \\ \Sigma_{j,31}^c & \Sigma_{j,32}^c & \Sigma_{j,33}^c \end{bmatrix} \tag{4}$$

$$j = 1,2, \ldots, m.$$

$$\theta_j^c \sim MVN(\mu_0^c, \Lambda_0^c) \tag{5}$$

$$\Sigma_j^c \sim inverseWishart(\nu_0^c, S_0^{c^{-1}}). \tag{6}$$

Based on the above impression and CTR distribution estimation model, we apply the Markov Chain Monte Carlo method (MCMC; e.g., Gamerman and Lopes, 2006) to make inferences about model parameters and data.

3.1.2 Gibbs Sampling

In the section 3.1.1, we assume that the performance indices (i.e., impression and CTR) for keywords in the same adgroup over the three matching options follow the same multivariate normal distribution, while these relationships between adgroups are independent. The processes of solving the keyword performance indices distribution estimation models through the Gibbs sampling are almost the same for impression and CTR. Therefore, we take the impression as the representative to illustrate the process of estimating the keyword performance index distributions over three matching options with incomplete data in this section. In the following, to simplify the expression, we use $d_j = \left\{d_{1,j}, d_{2,j}, \ldots, d_{n_j,j}\right\}$ to represent the variant of impression that has been transferred with log function[2]. For each adgroup j, the prior distribution of impression for the multivariate mean θ_j^d over three matching options is a MVN parameterized as

$$\begin{aligned} p\left(\theta_j^d\right) &= MVN\left(\mu_0^d, \Lambda_0^d\right) = (2\pi)^{-3/2}|\Lambda_0|^{-1/2}\exp\{-\frac{1}{2}\left(\theta_j^d - \mu_0^d\right)^T \Lambda_0^{d^{-1}}(\theta_j^d - \mu_0^d)\} \\ &= (2\pi)^{-3/2}|\Lambda_0|^{-1/2}\exp\{-\tfrac{1}{2}\theta_j^{d^T}\Lambda_0^{d^{-1}}\theta_j^d + \theta_j^{d^T}\Lambda_0^{d^{-1}}\mu_0^d - \tfrac{1}{2}\mu_0^{d^T}\Lambda_0^{d^{-1}}\mu_0^d\} \\ &\propto \exp\{-\tfrac{1}{2}\theta_j^{d^T}\Lambda_0^{d^{-1}}\theta_j^d + \theta_j^{d^T}\Lambda_0^{d^{-1}}\mu_0^d\} \\ &= \exp\{-\tfrac{1}{2}\theta_j^{d^T}A_0^d\theta_j^d + \theta_j^{d^T}b_0^d\}, \end{aligned} \tag{7}$$

[2] When estimating the impression, we used log function to transform the value of impression from interval $[0, +\infty]$ to the real line $[-\infty, +\infty]$; similarly, when estimating the click-through rate, we used logit function to transform the value of click-through rate value from interval [0,1] to the real line $[-\infty, +\infty]$. Then, in the modeling and experiments process, we would restore the value for estimated impression and CTR from $[-\infty, +\infty]$ to $[0, +\infty]$ and [0,1] using the inverse functions, separately.

where $A_0^d = \Lambda_0^{d^{-1}}$ and $b_0^d = \Lambda_0^{d^{-1}}\mu_0^d$. Conversely, Equation (7) says that if the random vector multivariate mean θ_j^d has a density on $\mathbb{R}^3$ that is proportional to $\exp\{-\frac{1}{2}\theta_j^{d^T}A_0^d\theta_j^d + \theta_j^{d^T}b_0^d\}$, then θ_j^d must have a MVN with covariance $A_0^{d^{-1}}$ and mean $A_0^{d^{-1}}b_0^d$.

In our estimation model, the impression of keywords in the j^{th} adgroup $\{d_{1,j},\dots,d_{n_j,j}\}$ are independent and identically distributed from the multivariate normal (θ_j^d, Σ_j^d), then similar calculations show that the joint density of the impression vector $d_{1,j},\dots,d_{n_j,j}$ is

$$p\left(d_{1,j},\dots,d_{n_j,j}\middle|\theta_j^d,\Sigma_j^d\right) = \prod_{k=1}^{n_j}(2\pi)^{-3/2}\,|\Sigma_j^d|^{-1/2}\exp\{-\frac{1}{2}\left(d_{k,j}-\theta_j^d\right)^T\Sigma_j^{d^{-1}}(d_{k,j}-\theta_j^d)\}$$

$$= (2\pi)^{-3n_j/2}|\Sigma_j^d|^{-n_j/2}\exp\{-\tfrac{1}{2}\textstyle\sum_{k=1}^{n_j}\left(d_{k,j}-\theta_j^d\right)^T\Sigma_j^{d^{-1}}(d_{k,j}-\theta_j^d)\}$$

$$\propto \exp\{-\tfrac{1}{2}\theta_j^{d^T}A_1^d\theta_j^d + \theta_j^{d^T}b_1^d\} \tag{8}$$

where $A_1^d = n\Sigma_j^{d^{-1}}$ and $b_1^d = n\Sigma_j^{d^{-1}}\bar{d}_j$ and $\bar{d}_j$ is the vector of impression variable-specific averages $\bar{d}_j = (\frac{1}{n_j}\sum_{k=1}^{n_j}d_{k,j,1}, \frac{1}{n_j}\sum_{k=1}^{n_j}d_{k,j,2}, \frac{1}{n_j}\sum_{k=1}^{n_j}d_{k,j,3})^T$. Combining Equations (7) and (8) gives

$$p(\theta_j^d|d_{1,j},\dots,d_{n_j,j},\Sigma_j^d) \propto \exp\{-\frac{1}{2}\theta_j^{d^T}A_0^d\theta_j^d + \theta_j^{d^T}b_0^d\}\times\exp\{-\frac{1}{2}\theta_j^{d^T}A_1^d\theta_j^d + \theta_j^{d^T}b_1^d\}$$

$$= \exp\{-\tfrac{1}{2}\theta_j^{d^T}A_n^d\theta_j^d + \theta_j^{d^T}b_n^d\}, \tag{9}$$

where $A_n^d = A_0^d + A_1^d = \Lambda_0^{d^{-1}} + n\Sigma_j^{d^{-1}}$ and $b_n^d = b_0^d + b_1^d = \Lambda_0^{d^{-1}}\mu_0^d + n\Sigma_j^{d^{-1}}\bar{d}_j$.

Equation (9) implies that the conditional distribution of the impression mean vector θ_j^d therefore must be a MVN with covariance $A_n^{d^{-1}}$ and mean $A_n^{d^{-1}}b_n^d$, so

$$Cov\left[\theta_j^d\middle|d_{1,j},\dots,d_{n_j,j},\Sigma_j^d\right] = \Lambda_n^d = \left(\Lambda_0^{d^{-1}} + n\Sigma_j^{d^{-1}}\right)^{-1} \tag{10}$$

$$E\left[\theta_j^d\middle|d_{1,j},\dots,d_{n_j,j},\Sigma_j^d\right] = \mu_n^d = \left(\Lambda_0^{d^{-1}} + n\Sigma_j^{d^{-1}}\right)^{-1}\left(\Lambda_0^{d^{-1}}\mu_0^d + n\Sigma_j^{d^{-1}}\bar{d}_j\right) \tag{11}$$

$$p\left(\theta_j^d\middle|d_{1,j},\dots,d_{n_j,j},\Sigma_j^d\right) = MVN(\mu_n^d,\Lambda_n^d) \tag{12}$$

The inverse-Wishart $(\nu_0^d, S_0^{d^{-1}})$ density of prior variance of the mean vector θ_j^d for impression over three matching options is given by

$$p(\Sigma_j^d) = \left[2^{\frac{3\nu_0^d}{2}}\pi^{\frac{3}{2}}\left|S_0^d\right|^{-\frac{\nu_0^d}{2}}\prod_{i=1}^3 \Gamma\left(\frac{[\nu_0^d+1-i]}{2}\right)\right]^{-1} \times \left|\Sigma_j^d\right|^{-(\nu_0^d+4)/2} \times \exp\{-\frac{1}{2}tr(S_0^d{\Sigma_j^d}^{-1})\}. \quad (13)$$

Combine the above prior distribution with the keyword impression distribution for $d_{1,j}, \dots, d_{n_j,j}$:

$$p\left(d_{1,j}, \dots, d_{n_j,j}\middle|\theta_j^d, \Sigma_j^d\right) = (2\pi)^{-\frac{3n_j}{2}}\left|\Sigma_j^d\right|^{-\frac{n_j}{2}}\exp\left\{-\frac{1}{2}\sum_{k=1}^{n_j}\left(d_{k,j}-\theta_j^d\right)^T{\Sigma_j^d}^{-1}\left(d_{k,j}-\theta_j^d\right)\right\}$$

$$= (2\pi)^{-3n_j/2}|\Sigma_j^d|^{-n_j/2}\exp\left\{-tr(S_\theta^d{\Sigma_j^d}^{-1})/2\right\}, \quad (14)$$

where $S_\theta^d = \sum_{k=1}^{n_j}\left(d_{k,j}-\theta_j^d\right)\left(d_{k,j}-\theta_j^d\right)^T$.

Combining Equations (13) and (14), we have the conditional distribution of covariance matrix Σ_j^d for impression:

$$\begin{aligned}
&p\left(\Sigma_j^d\middle|d_{1,j}, \dots, d_{n_j,j}, \theta_j^d\right)\\
&\propto p(\Sigma_j^d) \times p\left(d_{1,j}, \dots, d_{n_j,j}\middle|\theta_j^d, \Sigma_j^d\right)\\
&\propto \left(\left|\Sigma_j^d\right|^{-(\nu_0^d+4)/2} \times \exp\{-tr(S_0^d{\Sigma_j^d}^{-1})/2\}\right) \times \left(|\Sigma_j^d|^{-n_j/2}\exp\left\{-tr(S_\theta^d{\Sigma_j^d}^{-1})/2\right\}\right)\\
&= \left|\Sigma_j^d\right|^{-(\nu_0^d+n_j+4)/2}\exp\{tr([S_0^d+S_\theta^d]{\Sigma_j^d}^{-1})/2\}. \qquad (15)
\end{aligned}$$

Thus, we have

$$\left\{\Sigma_j^d\middle|d_{1,j}, \dots, d_{n_j,j}, \theta_j^d\right\} \sim inverseWishart(\nu_0^d + n_j, [S_0^d + S_\theta^d]^{-1}). \quad (16)$$

To distinguish the observed and unobserved keyword performance indices, let $O_{k,j} = (O_{k,j,1}, O_{k,j,2}, O_{k,j,3})^T$ be a binary vector where $O_{k,j,i} = 1$ implies that within adgroup j, the impression $d_{k,j,i}$ of keyword k under matching option i is observed, whereas $O_{k,j,i} = 0$ implies that the impression $d_{k,j,i}$ is unobserved. In practice, for each keyword k, advertiser can only observe the keyword performance indices under a certain matching option. And there is no fixed rule for the matching choices. The random guarantees that the relationship between observed index variable $O_{k,j}$ and keyword impression $d_{k,j}$ are statistically independent and that the distribution of observed index variable $O_{k,j}$ does not depend on the parameters θ_j^d or Σ_j^d of the multivariate normal distribution for impression. The probability for the keyword performance indices from keyword k in adgroup j is

$$p\left(O_{k,j},\left\{d_{k,j,i}: O_{k,j,i}=1\right\}\middle|\theta_j^d,\Sigma_j^d\right)=p(O_{k,j})\times p(\{d_{k,j,i}: O_{k,j,i}=1\}|\theta_j^d,\Sigma_j^d)$$

$$=p(O_{k,j})\times\int\left\{p(d_{k,j,1},d_{k,j,2},d_{k,j,3})|\theta_j^d,\Sigma_j^d)\prod_{d_{k,j,i}:O_{k,j,i}=0}d(d_{k,j,i})\right\} \quad (17)$$

The probability for keyword performance indices from keyword k in adgroup j is $p(O_{k,j})$ multiplied by the marginal probability of the observed variables, after integrating out the unobserved variables. The integration is done using Gibbs sampling, making inference on θ_j^d and Σ_j^d, as well as to make predictions for the impression under unobserved keyword matching options.

Let d_j be the $k\times 3$ impression matrix of all the observed and unobserved potential keyword impression over three matching options in adgroup j. Let O_j be the $k\times 3$ observation index matrix in which $O_{k,j,i}=1$ if $d_{k,j,i}$ is observed and $O_{k,j,i}=0$ if $d_{k,j,i}$ is unobserved. Then, the matrix d_j can be assumed to be composed of two parts: one is the keyword impression under a certain matching type that we do observe $d_j^{obs}=\{d_{k,j,i}: o_{k,j,i}=1\}$; another is the keyword impression for other matching types that we do not observe $d_j^{unobs}=\{d_{k,j,i}: o_{k,j,i}=1\}$. To approximate the posterior distribution of unknown and unobserved quantities $p(\theta_j^d,\Sigma_j^d,d_j^{unobs}|d_j^{obs})$, we build a Gibbs sampling scheme as follows:

Given starting values of the multivariate normal distribution parameters for impression of the j^{th} adgroup $\{\Sigma_j^{d^{(0)}},\ d_j^{unobs^{(0)}}\}$, we generate $\{\theta_j^{d^{(s+1)}},\Sigma_j^{d^{(s+1)}},\ d_j^{unobs^{(s+1)}}\}$ from $\{\theta_j^{d^{(s)}},\Sigma_j^{d^{(s)}},d_j^{unobs^{(s)}}\}$ by

1). sampling $\theta_j^{d^{(s+1)}}$ from $p\left(\theta_j^d|d_j^{obs},\ d_j^{unobs^{(s)}},\Sigma_j^{d^{(s)}}\right)$;

2). sampling $\Sigma_j^{d^{(s+1)}}$ from $p\left(\Sigma_j^d|d_j^{obs},\ d_j^{unobs^{(s)}},\theta_j^{d^{(s+1)}}\right)$;

3). sampling $d_j^{unobs^{(s+1)}}$ from $p\left(d_j^{unobs}|d_j^{obs},\theta_j^{d^{(s+1)}},\Sigma_j^{d^{(s+1)}}\right)$.

In steps 1 and 2, the fixed observed impression for the j^{th} adgroup d_j^{obs} combines with the current unobserved impression value for the j^{th} adgroup of $d_j^{unobs^{(s)}}$ to shape a complete version of impression matrix $d_j^{(s)}$ with complete impression values for adgroup j. In order to obtain the full conditional distributions of the MVN parameters θ_j^d and Σ_j^d for impression, the k rows of the impression matrix of $d_j^{(s)}$ can then be plugged into Equations (12) and (16). In Step 3,

$$p(d_j^{unobs}|d_j^{obs},\theta_j^d,\Sigma_j^d)\propto p(d_j^{unobs},d_j^{obs}|\theta_j^d,\Sigma_j^d)$$

$$= \prod_{k=1}^{n_j} p(d_{k,j}^{unobs}, d_{k,j}^{obs} | \theta_j^d, \Sigma_j^d)$$

$$\propto \prod_{k=1}^{n_j} p(d_{k,j}^{unobs} | d_{k,j}^{obs}, \theta_j^d, \Sigma_j^d) \quad (18)$$

For each keyword k, we have to sample the unobserved keyword impression vector elements conditional on the observed keyword impression vector elements. This is made possible through the following MVN result: Let $d_j \sim MVN(\theta_j^d, \Sigma_j^d)$, let a be a subset of the variable indices $\{1,2,3\}$, and let b be the complement of a. According to inverses of partitioned matrices, we have:

$$\{d_{j_{[b]}} | d_{j_{[a]}}, \theta_j^d, \Sigma_j^d\} \sim MVN({\theta_j^d}_{b|a}, {\Sigma_j^d}_{b|a}), \text{ where}$$

$${\theta_j^d}_{b|a} = {\theta_j^d}_{[b]} + {\Sigma_j^d}_{[b,a]} \left({\Sigma_j^d}_{[a,a]}\right)^{-1} (d_{j_{[a]}} - {\theta_j^d}_{[a]}) \quad (19)$$

$${\Sigma_j^d}_{b|a} = {\Sigma_j^d}_{[b,b]} - {\Sigma_j^d}_{[b,a]} \left({\Sigma_j^d}_{[a,a]}\right)^{-1} {\Sigma_j^d}_{[a,b]}, \quad (20)$$

where ${\theta_j^d}_{[b]}$ refers to the elements of θ_j^d corresponding to the indices in b, and ${\Sigma_j^d}_{[a,b]}$ refers to the matrix made up of the elements that are in rows a and columns b of Σ_j^d.

3.2 Stochastic Optimization for Keyword Targeting

In SSA, advertisers have to make keyword targeting decisions, i.e., select keywords and determine matching types for selected keywords, before obtaining the values of keyword performance indices (e.g., impression and click-through-rate) over three keyword matching options. Based on the estimated distributions of keyword performance indices over the three matching options in the previous section, we construct a stochastic model for keyword targeting to maximize the expected profit in SSA. The keyword targeting decision scenario we consider in this research is: given a set of keywords, how can an advertiser select appropriate keywords and determine matching types for these selected keywords simultaneously?

3.2.1 The Objective of Keyword Targeting Optimization

The decision variable x_{kji} indicates whether the keyword k ($k = 1,2,\dots,n_j$) in adgroup j ($j = 1,2,\dots,m$) chooses matching type i ($i = 1,2,3$ corresponding to exact match, phrase match and broad match) or not, i.e.,

$$x_{k,j,i} = \begin{cases} 1, if\ the\ keyword\ k\ in\ adgroup\ j\ chooses\ matching\ type\ i; \\ 0, otherwise. \end{cases}$$

If keyword k in adgroup j is not chosen in any of the three matching types, we denote $x_{k,j,0} = 1$, then we have $\sum_{i=0}^{3} x_{k,j,i} = 1$. When $i = 0$, it means that keyword k is not selected by the advertiser. Let $d_{k,j,0} = 0$ and $c_{k,j,0} = 0$. Since the cost of the search advertising campaign is $\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3} x_{k,j,i}d_{k,j,i}c_{k,j,i}p_{k,j}$, we describe the profit in SSA campaign as $z(x_{k,j,i}) = \sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3} x_{k,j,i}d_{k,j,i}c_{k,j,i}(v_{k,j} - p_{k,j})$.

In this research, we use estimated adgroup data distributions to approximate the keyword performance indices distribution. We regard impression $d_{k,j,i}$ and CTR $c_{k,j,i}$ as random vectors capturing the uncertainty in SSA. Then, $z(x_{k,j,i})$ is a random variable. Therefore, keyword targeting decisions aim to maximize the expected profit generated from a SSA campaign, i.e., max $E\left[z(x_{k,j,i})\right] = E\left[\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3} x_{k,j,i}d_{k,j,i}c_{k,j,i}(v_{k,j} - p_{k,j})\right]$.

3.2.2 The Chance Constraint of Budget

It is naturally assumed that the available budget for advertiser is relatively less than a sufficient amount. In a SSA campaign, let $B > 0$ denote the advertising budget available for a SSA campaign, then we have $\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3} x_{k,j,i}d_{k,j,i}c_{k,j,i}p_{k,j} \le B$.

Due to the stochastic nature of $d_{k,j,i}$ and $c_{k,j,i}$, we can use the chance constraint of budget to control the cost for a SSA campaign. Specifically, the probability that the cost of a SSA campaign is less than the advertiser's available budget, is larger than or equal to a specific confidence level α, i.e., $P\left\{\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3} x_{k,j,i}d_{k,j,i}c_{k,j,i}p_{k,j} \le B\right\} \ge \alpha$.

3.2.3 The Stochastic Keyword Targeting Model

In summary, based on the data distribution estimation for keyword performance indices in section 3.1, we formulate the keyword targeting problem as a stochastic optimization model, given as:

$$
\begin{aligned}
\max \quad & E\left[\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3} x_{k,j,i}d_{k,j,i}c_{k,j,i}(v_{k,j} - p_{k,j})\right] \\
\text{s.t.} \quad & P\left\{\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3} x_{k,j,i}d_{k,j,i}c_{k,j,i}p_{k,j} \le B\right\} \ge \alpha \\
& \sum_{i=0}^{3} x_{k,j,i} = 1 \\
& \left(log(d_{k,j,1}), log(d_{k,j,2}), log(d_{k,j,3})\right)^T \sim MVN\left(\theta_j^d, \Sigma_j^d\right) \\
& \left(logit(c_{k,j,1}), logit(c_{k,j,2}), logit(c_{k,j,3})\right)^T \sim MVN\left(\theta_j^c, \Sigma_j^c\right)
\end{aligned}
\tag{21}
$$

$$\theta_j^d = E\begin{bmatrix} d'_{j,1} \\ d'_{j,2} \\ d'_{j,3} \end{bmatrix}, \Sigma_j^d = E\begin{bmatrix} \Sigma_{j,11}^d & \Sigma_{j,12}^d & \Sigma_{j,13}^d \\ \Sigma_{j,21}^d & \Sigma_{j,22}^d & \Sigma_{j,23}^d \\ \Sigma_{j,31}^d & \Sigma_{j,32}^d & \Sigma_{j,33}^d \end{bmatrix}$$

$$\theta_j^c = E\begin{bmatrix} c'_{j,1} \\ c'_{j,2} \\ c'_{j,3} \end{bmatrix}, \Sigma_j^c = E\begin{bmatrix} \Sigma_{j,11}^c & \Sigma_{j,12}^c & \Sigma_{j,13}^c \\ \Sigma_{j,21}^c & \Sigma_{j,22}^c & \Sigma_{j,23}^c \\ \Sigma_{j,31}^c & \Sigma_{j,32}^c & \Sigma_{j,33}^c \end{bmatrix}$$

$$x_{k,j,i} = 0/1, d_{k,j,0} = 0, c_{k,j,0} = 0, v_{k,j} \geq 0, p_{k,j} \geq 0$$

$$k = 1,2, \dots, n_j, i = 0,1,2,3, j = 1,2, \dots, m$$

In our stochastic keyword targeting model, the decision variable $x_{k,j,i}$ is 0-1 binary. The objective function is aimed at maximizing the expected profit of the SSA campaign. The first constraint is the chance constraint of advertiser's soft budget. The second constraint limits each keyword being selected for only one type of matching or not selected at all. In addition, we state the estimated data distributions and non-negative nature of the keyword performance indices.

3.2.4 The BBKMS Algorithm

The stochastic keyword targeting model can be solved by branch-and-bound algorithm combined with stochastic simulation. The branch-and-bound consists of a systematic enumeration of feasible solutions which helps reduce the computational efforts and find the optimal keyword targeting result (Kosuch and Lisser 2010). In particular, given a set of keywords and fixed budget, based on the estimated distributions of keyword performance indices over the three matching options in the section 3.1, an optimum solution should adaptively select appropriate keywords and determine their matching types to maximize the expected profit in uncertain SSA market. In SSA, advertisers assign most keywords into only one adgroup, while few keywords into more than one adgroups. Since the same keyword in different adgroups could have different keyword matching options, which lead to different keyword performances. Thus, we regard the identical keyword in different adgroups as different keywords. As every keyword can only choose one matching type, we treat the keyword-matching combinations as the basic unit and develop a branch-and-bound algorithm, called BBKMS, i.e., branch-and-bound algorithm for keyword matching and selection, to solve our stochastic keyword targeting model. In the following, we first describe the stochastic simulation process in the branch-and-bound. Then, we explain how we calculate the upper bound for the BBKMS. Finally, we present the BBKMS algorithm.

First, the stochastic simulation process is used to confirm whether the chance constraint is satisfied in the process of branch-and-bound, i.e., the probability that the current cost for selected keywords have exceeded campaign budget is under the acceptable threshold value. Specifically, in adgroup j, once a keyword k choosing matching type i is selected, we sample the random variables from their estimated distributions and check if the chance constraint of budget $P\left\{\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3}x_{k,j,i}d_{k,j,i}c_{k,j,i}p_{k,j} \le B\right\} \ge \alpha$ is satisfied in the stochastic simulation process, i.e., the probability that the total cost is less than the given budget is above a certain level. It is shown in lines 4 to 8 of the pseudocode for the algorithm BBKMS.

Next, to obtain the upper bound, i.e., SUP, for the BBKMS, we relax $x_{k,j,i}$ from a 0-1 binary variable to a continuous variable in the interval of [0,1] within the stochastic keyword targeting model (21). Under the continuous relaxation, the chance constraint of budget $P\left\{\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3}x_{k,j,i}d_{k,j,i}c_{k,j,i}p_{k,j} \le B\right\} \ge \alpha$ defines a convex set if function $\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3}x_{k,j,i}d_{k,j,i}c_{k,j,i}p_{k,j}$ is quasi-convex and keyword cost $s_{k,j,i} = d_{k,j,i}c_{k,j,i}p_{k,j}$ has a log-concave density (Prekopa, 1995). The first condition is well satisfied since the left-hand side of the inequation $\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3}x_{k,j,i}d_{k,j,i}c_{k,j,i}p_{k,j}$ is quasi-convex for its linear formulation. As for the second condition, according to Cholette et al. (2012), the number of clicks for keyword is binomial distributed with parameters$(d_{k,j,i}, c_{k,j,i})$, which can be accurately approximated by the normal distribution provided that $d_{k,j,i} \cdot c_{k,j,i} \ge 10$ and $d_{k,j,i} \cdot (1 - c_{k,j,i}) \ge 10$. In general, we could assume that the keyword impression $d_{k,j,i}$ and click-through rate $c_{k,j,i}$ is reasonably well satisfied to these two inequalities. Since the keyword cost is equal to the product of the number of clicks (the random variable) and the average cost per click (constant), it also obeys the normal distribution. Thus, the second condition can be proved. Consequently, after the continuous relaxation of the stochastic keyword targeting model, the chance constraint of budget $P\left\{\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3}x_{k,j,i}d_{k,j,i}c_{k,j,i}p_{k,j} \le B\right\} \ge \alpha$ defines a convex set under the normally distributed keyword cost.

Then, the continuous relaxed stochastic keywords targeting model (21) can be reformulated as an equivalent, deterministic second-order-cone-programming (SOCP) problem (Lobo et al. 1998). Through sampling from the estimations of impression and CTR, we can get the parameters of normally distributed keyword cost $s_{k,j,i} \sim N({\mu^{(s)}}_{k,j,i}, {\sigma^{(s)}}^2_{k,j,i})$. For the budget B is a fixed constant, inequality constraints $\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3}x_{k,j,i}s_{k,j,i} \le B$ is equal to

$$\frac{\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3}x_{k,j,i}s_{k,j,i}-B-\left(\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3}x_{k,j,i}E[s_{k,j,i}]-B\right)}{\sqrt{\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3}x_{k,j,i}^{2}Var[s_{k,j,i}]}} \leq -\frac{\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3}x_{k,j,i}E[s_{k,j,i}]-B}{\sqrt{\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3}x_{k,j,i}^{2}Var[s_{k,j,i}]}}, \tag{22}$$

where the left side of the inequality represents a standard normal variant. Thus, the chance constraint of budget $P\left\{\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3}x_{k,j,i}d_{k,j,i}c_{k,j,i}p_{k,j} \leq B\right\} \geq \alpha$ is equal to

$$P\left\{\eta \leq -\frac{\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3}x_{k,j,i}E[s_{k,j,i}]-B}{\sqrt{\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3}x_{k,j,i}^{2}Var[s_{k,j,i}]}}\right\} \geq \alpha, \tag{23}$$

where η follows a standard normal distribution. Then, the chance constraint of budget can be reformulated as

$$\phi^{-1}(\alpha) \leq -\frac{\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3}x_{k,j,i}E[s_{k,j,i}]-B}{\sqrt{\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3}x_{k,j,i}^{2}Var[s_{k,j,i}]}}$$

$$\Rightarrow \sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3}x_{k,j,i}E[s_{k,j,i}] + \phi^{-1}(\alpha)\sqrt{\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3}x_{k,j,i}^{2}Var[s_{k,j,i}]} \leq B,$$

$$\Rightarrow \sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3}x_{k,j,i}\,{\mu^{(s)}}_{k,j,i} + \phi^{-1}(\alpha)\sqrt{\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3}x_{k,j,i}^{2}{\sigma^{(s)}}^{2}_{k,j,i}} \leq B. \tag{24}$$

Consequently, through relaxing $x_{k,j,i}$ from a 0-1 binary variable to a continuous variable in the interval of [0,1] and reconstructing the chance constraint of budget into a deterministic inequality constraint within the stochastic keyword targeting model (21), we obtain a convex optimization model which can be solved by the interior point method (Wächter and Biegler 2006). Its solution can be used as upper bound (SUP) in the branch-and-bound process.

In general, we develop a branch-and-bound algorithm, i.e., BBKMS, to solve our stochastic keyword targeting model. The algorithm systematically enumerates the candidate optimal keyword targeting solutions, where the set of candidate solutions is thought of as forming a rooted tree with the full set at the root. It explores branches of the tree, i.e., subsets of the keyword targeting solution set. Before enumerating the candidate keyword targeting solutions of a branch, the branch is checked against upper (SUP) and lower (INF) estimated bounds on the optimum, and is discarded if it cannot produce a better keyword targeting solution than the existing best one found so far by the algorithm BBKMS. In the exploring process, to deal with the chance constraint of budget, we use stochastic simulation to check whether the chance constraint is still satisfied when selecting the next keyword. The overall framework and more details of the BBKMS algorithm are given below.

Algorithm（BBKMS）

Input:

$\{i|i=1,2,3\}$ – three types of keyword matching (i.e., exact match, phrase match and broad match)

$\{j|j=1,2,\dots,m\}$ –adgroup set

$\{k|k=1,2,\dots,n_j\}$ – shared keyword set in adgroup *j*

B – the soft budget constraint for the SSA campaign

α– the acceptable probability for the chance constraint of SSA campaign budget

θ_j^d, Σ_j^d – the estimated mean vector and covariance matrix for the impression of keyword in adgroup *j* over three matching options

θ_j^c, Σ_j^c– the estimated mean vector and covariance matrix for the CTR of keyword in adgroup *j* over three matching options

$v_{k,j}$ – the VPC of the keyword k in adgroup *j*

$p_{k,j}$ – the CPC of the keyword k in adgroup *j*

Output:

$x_{k,j,i}$ – the 0-1 binary decision variable for keyword targeting

Initialize the order of keywords by their decreasing observed expected profit, $x_{k,j,i}=0$, Keyword_Targeting_List = ∅.

1: **for** adgroup j from 1 to m

2: **for** keyword k from 1 to n

3: **for** matching type i from 1 to 3

4: $t'=0$ and $x_{k,j,i}=1$

5: extract $d_{k,j,i}$ and $c_{k,j,i}$ samples from $MVN(\theta_j^d,\Sigma_j^d)$ and $MVN(\theta_j^c,\Sigma_j^c)$

6: **if** $\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{i=0}^{3}x_{k,j,i}d_{k,j,i}c_{k,j,i}p_{k,j}\le B$ **then** we have $t'++$

7: **repeat** steps 5 and 6 for t times, and $\alpha'=t'/t$

8: **if** $\alpha'\ge\alpha$ and $\sum_{i=1}^{3}x_{k,j,i}\le 1$, **then** INF = max{the expected profit}, add the feasible solution to Keyword_Targeting_List, and SUP = ∞; else $x_{k,j,i}=0$

9: end for

10: end for

11: end for

12: **if** Keyword_Targeting_List = ∅, **then** go to step 16; **else** current_solution = solution in Keyword_Targeting_List with max{the expected profit}, go to step 13.

13: **if** SUP > INF for current_solution, **then** go to step 14; **else** delete the solution from Keyword_Targeting_List and go to step 12.

14: **if** there is no accepted keyword-matching left in the targeting solution that does not already have a plunged or rejected subset, **then** delete the solution from Keyword_Targeting_List, go to step 12; **else** following the ranking, choose the first accepted keyword-matching that does not already have a plunged or rejected subset and calculate SUP for the subset defined by rejecting this keyword-matching, go to step 15.

15: **if** SUP ≤ INF, **then** delete this subset, go to step 14; **else** plunge the subtree as described in step 1-11 and add the found branch together with SUP to Keyword_Targeting_List, **if** the expected profit of this solution > INF, **then** update INF, go to step 12.

16: **return** the result of keywords targeting $\boldsymbol{x}$.

4. Experimental Validation

4.1 Data Descriptions and Experimental Setup

We conduct a series of computational experiments to validate our keyword targeting strategy using a realworld dataset. The purpose of our experiments is two-fold. First, we aim to assess the advantage of our distribution estimation method in keyword targeting. Second, we evaluate the effectiveness of our keyword targeting strategy (i.e., BBKMS) by comparing with four baselines, in terms of the expected profit and the number of selected keywords.

Our experimental dataset is collected from field reports and SSA logs of daily advertising campaigns performance of an e-commerce firm selling gift cards and Christmas presents. The dataset records the daily performance at the keyword level from September 2011 to June 2017, including impressions, click-through rate (CTR), value-per-click (VPC), average cost-per-click (CPC) and the chosen matching types. The dataset contains 34 adgroups with 627 keywords. In Table 4, we give examples of keyword dataset with keyword ID, ad-group ID, matching types, impressions, CTR, VPC and CPC. Table 5 shows the summary statistics for the dataset, including the mean and standard deviation of keyword performance indices, and the proportion of the three matching types.

Table 4. Examples of the Keyword Dataset

Day	Keyword ID	Ad Group ID	Matching Type	Impression	CTR	VPC	CPC
2012/6/13	keyword-31	adgroup-13	broad	36	0.06	50	0.31
2012/6/20	keyword-31	adgroup-13	broad	18	0.17	18.3	0.25
2012/8/2	keyword-31	adgroup-13	broad	37	0.22	25	0.26
2012/9/3	keyword-31	adgroup-13	broad	49	0.16	25	0.29
2012/10/20	keyword-31	adgroup-13	broad	35	0.11	35	0.33
2013/3/16	keyword-31	adgroup-13	broad	12	0.16	24.5	0.16
……							
2016/3/8	keyword-402	adgroup-25	phrase	2	0.5	75	0.31
2016/11/12	keyword-402	adgroup-25	phrase	7	0.29	39.5	0.42
2016/11/30	keyword-402	adgroup-25	phrase	8	0.13	53.5	0.46
2016/12/14	keyword-402	adgroup-25	phrase	26	0.35	10	0.33
2016/12/17	keyword-402	adgroup-25	phrase	25	0.24	27.3	0.28
2016/12/21	keyword-402	adgroup-25	phrase	52	0.13	4	0.35
……							
2013/1/8	keyword-527	adgroup-9	exact	21	0.38	39.8	0.27
2013/5/20	keyword-527	adgroup-9	exact	21	0.14	6.67	0.12
2013/6/24	keyword-527	adgroup-9	exact	8	0.5	14.8	0.15
2013/9/10	keyword-527	adgroup-9	exact	11	0.09	100	0.15
2014/1/11	keyword-527	adgroup-9	exact	8	0.25	95	0.14
2015/1/10	keyword-527	adgroup-9	exact	8	0.38	64.5	0.17
……							

Table 5. Summary Statistics of the Dataset

Variable	Impression	CTR	VPC	CPC	Exact match	Phrase match	Broad Match
Mean	139.47	0.41	35.59	0.23	32.31%	15.98%	51.71%
SD	230.43	0.24	287.14	0.20			

In the stage of data distribution estimation, we estimate the parameters of keyword performance indices (i.e., impression and click-through rate) by using the statistical tool WinBUGS to implement Bayesian models with MCMC methodology (Spiegelhalter et al., 2003).

We run each process for 50,000 iterations. In the stage of keyword targeting stochastic optimization, we set the confidence level (i.e., α) for the chance constraint of advertising budget as 0.95. In this dataset, the total cost of all keywords in the SSA campaign is 1175.74. Thus, we gradually increase the computational campaign budget setting from 100 to 1000 in steps of 100 to verify the effectiveness of our keyword targeting strategy over different levels of budget.

4.2 Comparisons

We compare our keyword targeting strategy (BBKMS) with four baselines with respect to the expected profit, the profit-cost distribution and the number of selected keywords. The first baseline (BASE1-Greedy) selects keywords on the basis of the frequency of keywords used by the advertiser in past SSA campaign. Because there is limited research on keywords targeting optimization and no suitable state-of-the-art literature can be directly compared with ours, we developed two baseline strategies derived from literature in keyword selection for comparison. The second baseline (BASE2-PrefixOrder) is derived from (Rusmevichientong and Williamson 2006), which adaptively selects keywords based on a prefix ordering-sorting keywords in descending order of profit-to-cost ratio until the expected cost is close to the budget. The third baseline (BASE3-Competitiveness) is derived from (Zhang et al. 2014), which selects keywords with goals of maximizing advertiser's profit and minimizing the keyword competitiveness, under a budget control constraint. The index of "impression confidence based on competitiveness" in Zhang et al. is defined as $c = 1 - 1/(1 + e^{-\tau d})$ in the baseline, where $\tau > 0$ is a coefficient and d is the impression of keyword. The fourth baseline (BASE4-SelectNomatch) is a variant of our approach, which follows the stochastic keyword targeting model derived from Section 3.2 based on the original incomplete observed keyword matching data without the data distribution estimation process in Section 3.1. We design this baseline to validate the importance of our data distribution estimation.

Figure 1 shows the expected profit obtained by our approach (BBKMS) and four baselines at various budget levels. Figure 2 illustrates the profit and cost of keywords selected by the BBKMS and four baselines, respectively, with the budget constraint of 500.

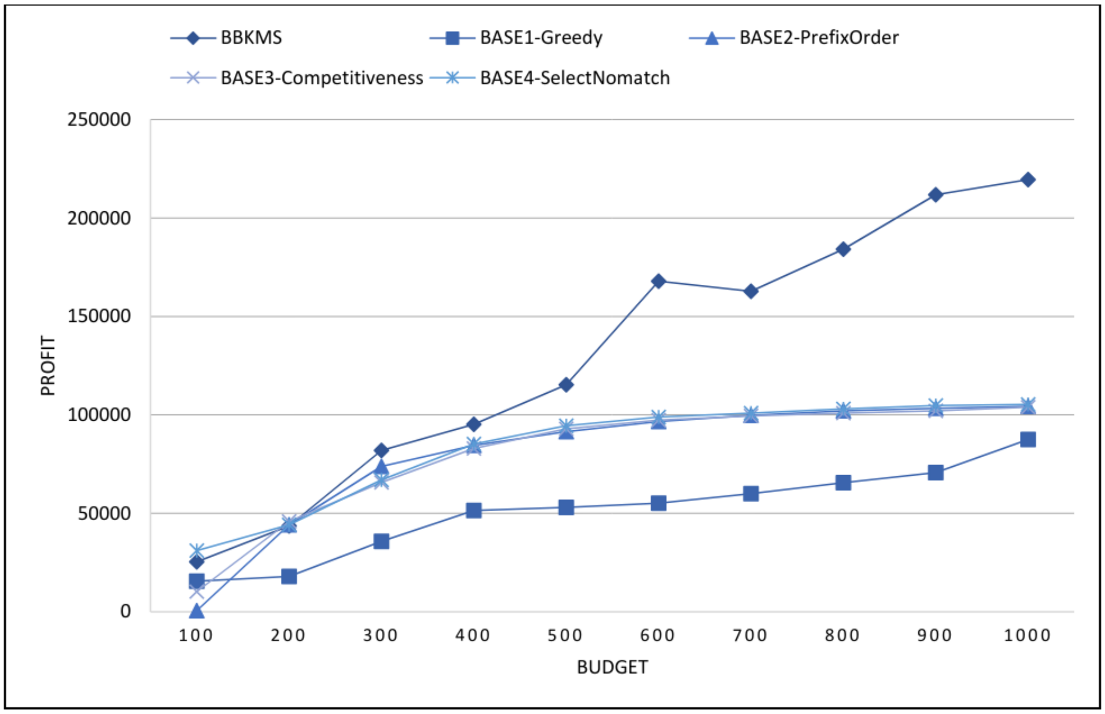
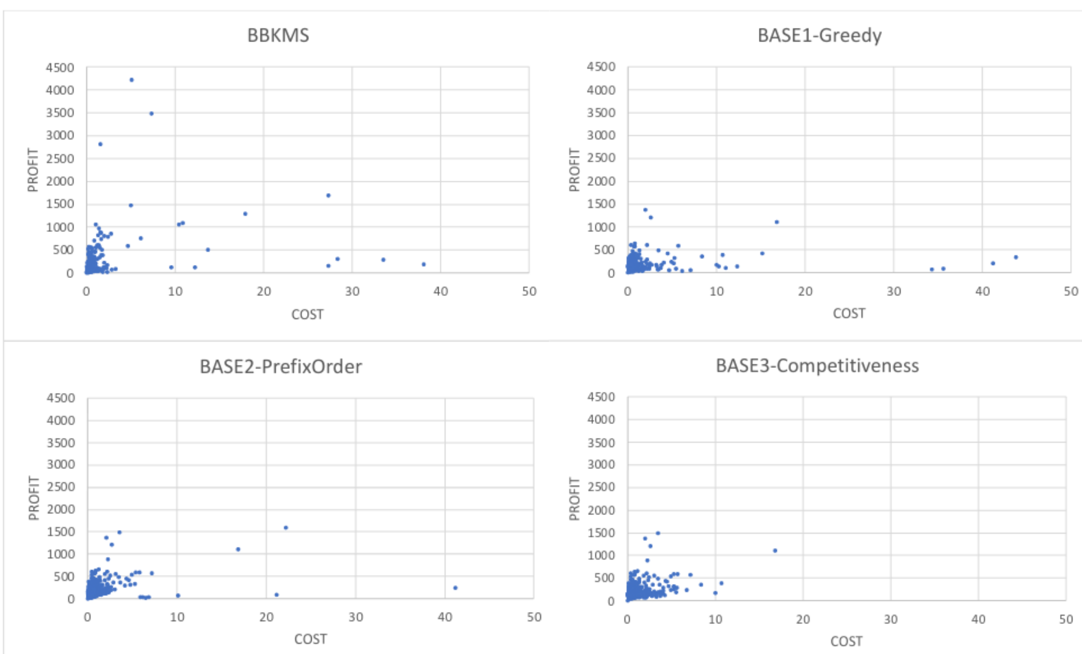


Figure 1. The Expected Profit Obtained by the BBKMS and Four Baselines



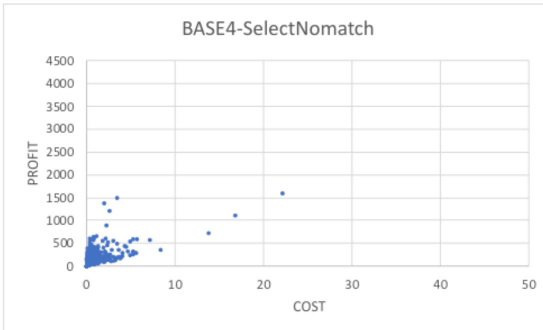


Figure 2. The Profit and Cost of Keywords Selected by the BBKMS and Four Baselines

From Figures 1 and 2, we observe the following:

(1) In general, our keyword targeting approach (BBKMS) outperforms the four baselines in terms of profit.

(2) All the baselines ignore both data distribution estimation and keyword matching optimization. They select keywords based on the original incomplete observed keyword matching data. Among them, BASE1-Greedy performs the worst. We can find the reason from Figure 2. When advertiser selects keywords based on her experience, she prefers choosing her familiar keywords that she used frequently in the past. However, some of these keywords might have higher costs and limited profits, but this phenomenon is difficult for advertiser herself to detect. This emphasizes the necessity to conduct keywords targeting optimization, which is more effective compared to advertiser's subjective judgment. Moreover, from Figure 1, we can observe that BASE4-SelectNomatch performs slightly better than all the other baselines. This result could be attributed to the superiority in using our stochastic optimization model which takes the uncertainty of SSA market into account, and the corresponding branch-and-bound algorithm combined with stochastic simulation can explore more possibilities in the keyword targeting solution space.

(3) When budget gradually grows to abundant, the decision environment becomes more complicated. In this case, compared with the four baseline approaches, the advantage of our keyword targeting approach (BBKMS) becomes more prominent in terms of profit growth. In another word, our keyword targeting approach (BBKMS) achieves highest profit. It's because, first, our approach includes data distribution estimation. Compared with the BASE4-SelectNomatch that uses the stochastic optimization model in section 3.2 as our keyword targeting approach (BBKMS), while bases on the original incomplete observed keyword performance indices, our keyword targeting approach (BBKMS) with the data distribution estimation gives

more specific and accurate estimation for keyword performance indices over three keyword matching options. We find more keyword-matching combinations for advertiser, which increases the possibility of profit growth in keyword targeting. In Figure 2, compared our approach (BBKMS), with the baselines that only consider keyword selection optimization, we can find that, through estimating the data distributions of keyword performance indices (i.e., impression and click-through rate) over three matching options, we can have a more complete understanding of keywords, i.e., it helps us find more potential high-profit yet less-cost keyword-matching combinations. This shows that our approach can help advertisers enrich their keyword targeting portfolios and increase profits. Second, after estimating data distributions for keyword performance indices over three matching options, we deal with the keyword selection and keyword matching problems in an integrated way (i.e., keyword targeting). In SSA, multiple keyword decisions are interdependent and jointly affect the advertising performance. The collaborative optimization, i.e., select keywords and determine matching types for selected keywords simultaneously, makes advertising decision more efficient.

The number of keywords selected by our keyword targeting approach (BBKMS) and four baseline approaches at various SSA campaign budget levels are shown in Figure 3.

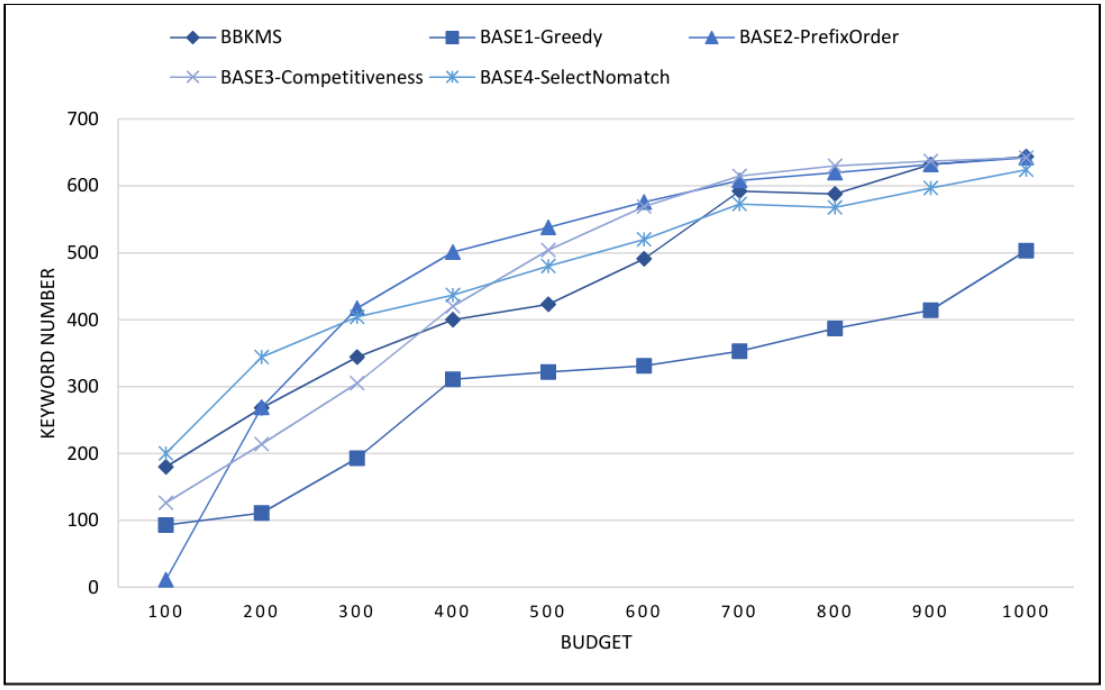


Figure 3. The Number of Keywords Selected by the BBKMS and Four Baselines

The following phenomenon can be observed in Figure 3:

(1) In general, the number of selected keywords for our keyword targeting approach (BBKMS) and the four baseline approaches show a similar increasing trend with the SSA campaign budget. However, there are some exceptions where the budget increases, but the number of selected keywords decreases. This is because, in SSA, there exists some high-profit and high-cost keywords (i.e., hot keywords). High-profit put advertiser into a fiercely competitive environment, and high-cost account for a large share of advertiser's budget. They might hurt advertiser's profit through running out of the budget too quickly and missing better opportunities to display her advertisings. Thus, when budget is limited, advertiser prefers selecting a large number of long tail keywords, i.e., low-profit and low-cost keywords in order to catch more advertising display opportunities and avoid the fierce competition from hot keywords. When the budget is enough, hot keywords will become a good choice for advertiser. And, in this case, advertiser replaces a large number of long-tail keywords with fewer hot keywords to obtains higher profit, and the number of selected keywords decreases.

(2) Most of the time, BASE2-PrefixOrder and BASE3-Competitiveness select more keywords than other approaches. This is mainly due to the nature of these two approaches. BASE2-PrefixOrder and BASE3-Competitiveness optimize the advertiser's expected profit by selecting high ROI keywords and less-competitive keywords, respectively. These keywords have one common feature, that is, relatively lower cost, thus their selected keyword numbers are larger than other approaches under the same budget.

(3) From Figures 1 to 3, we can see that, although our approach (BBKMS) selects fewer keywords than some baselines (e.g., BASE2-PrefixOrder and BASE3-Competitiveness), it achieves a higher profit. Thus, we can find that selecting more keywords into the advertising campaign will not necessarily bring advertiser higher profit. In our research, a more effective keyword targeting strategy can help advertiser obtain more profit than just selecting more keywords.

Figure 4 shows the proportions of keyword matching types over budget in our optimal keyword targeting solutions (i.e., BBKMS). According to the common sense of keyword targeting in SSA, broad match is the most popular option, followed by exact match, and phrase match is the least. Both the industry studies and academic research have used to reveal the popularity of broad match compared to exact match and phrase match. There are reports stated that 56% of clicks are through broad match keywords, compared to only 33% through exact match keywords and 11% through phrase match keywords on Google; similarly, on Bing, these numbers are 70%, 20% and 10%, respectively (Ballard, 2013; Amaldoss et al., 2016). However, contrary to this common belief, in Figure 4, we can find that in our experiments, the exact match and phrase match take up considerable proportions in the optimal keyword targeting solutions. This illustrates that although broad match has the advantages of increasing advertisers' reach, and facilitates SSA campaign management, the exact match and phrase match can't be ignored. All types of match play important roles in keyword targeting. Specifically, on one hand, broad match does maximize the potential to show an advertiser's ads on relevant searches reaching the broadest possible audience. So, it is a great way to drive lots of impressions and clicks. However, on the other hand, it might also exhaust a large amount of budget paying for irrelevant traffic that doesn't convert (WordStream, 2021). When changing the keyword matching option from broad match to phrase match (or even to exact match), it narrows down the targeted audience and thus limits the traffic, but the more targeted consumers are more likely to be interested in the promoted products (Klapdor et al., 2014). Compared to phrase match and exact match, broad match pushes advertising campaigns into a more intense competition through wider targeting (Levin and Milgrom, 2010). Thus, although broad match could lead to higher monetization of traffics for search engines (Gupta et al., 2009), it might harm advertisers' earnings (Amaldoss et al., 2016).

In summary, each matching option has its own advantages and disadvantages. Compared with single matching option, mixed keyword matching options help advertisers enrich keyword portfolios and obtains higher profit in keyword targeting, through reducing unwanted costs and keeping conversion rates on the relatively high level. It allows advertisers to balance the tradeoff between reaching the widely targeted audience and avoiding unnecessary spending on irrelevant clicks (Amaldoss et al., 2016).

Our keyword targeting approach chooses a suitable matching type for each keyword based on the historical keyword performance indices, rather than sets a one-size-fits-all option (e.g., setting all keywords as broad match). The strategy with mixed keyword matching options brings higher profit to advertisers. This result emphasizes the importance of data distribution estimation in keyword targeting optimization.

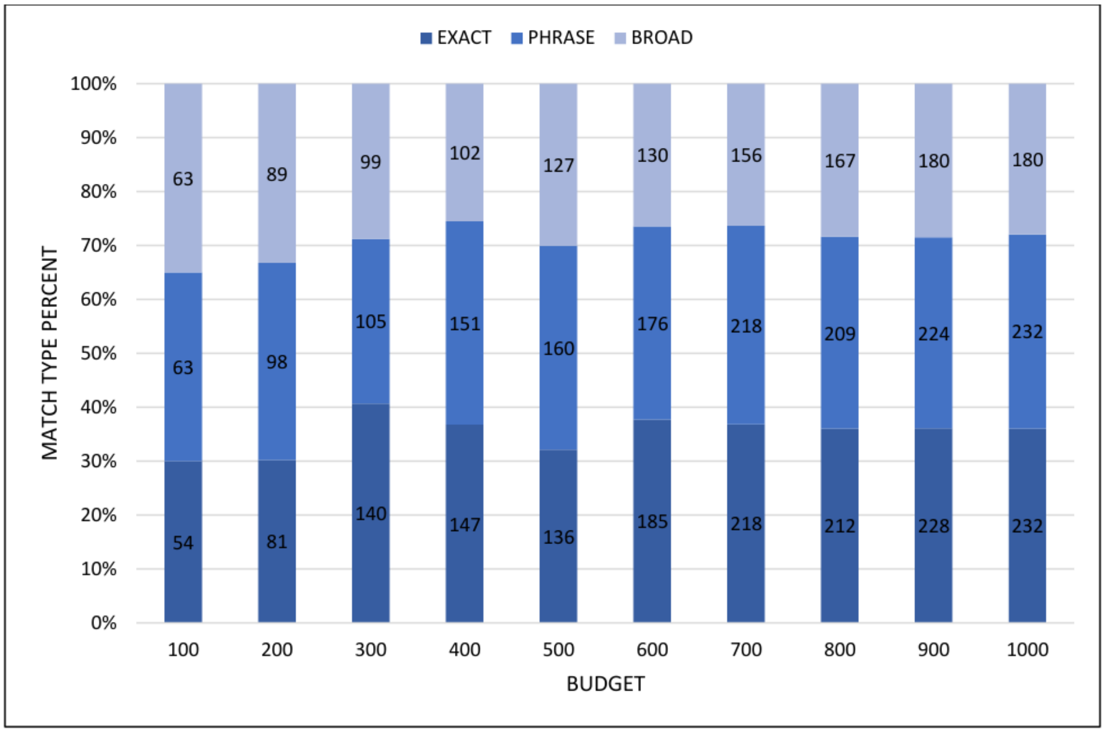


Figure 4. The Proportions of Matching Types of Selected Keywords

5. Conclusions

In this paper, we construct our data distribution estimation model for performance indices (i.e., impression and click-through rate), and apply Markov Chain Monte Carlo computational methods to compensate for unobserved indices over three matching options. Then based on the estimation results, we propose a stochastic keyword targeting model to maximize the expected profit under the chance constraint of budget, and develop a BBKMS solution to solve our keyword targeting model. Using the realworld dataset collected from field reports and SSA campaign logs, we conduct a series of experiments to validate the effectiveness of our keyword targeting strategy. Our experimental results illustrate that our keyword targeting approach outperforms the four baselines in terms of the expected profit.

5.1 Theoretical Implications

This research makes important scientific contributions in two aspects. First, to the best of our knowledge, this is the first research effort on keyword targeting problem in SSA. This study adds to the SSA literature by exploring keyword targeting decisions in the framework of stochastic optimization under the chance constraint of budget. Previous literatures have studied keyword selection (Rusmevichientong and Williamson, 2006; Kiritchenko and Jiline, 2008; Zhang et al., 2014; Desai et al., 2014; Lu and Yang, 2017) and keyword matching (Radlinski et al., 2008; Singh and Roychowdhury, 2008; Even Dar et al., 2009; Gupta et al., 2009; Grbovic et al. 2016; Amaldoss et al. 2016), separately. This study is the first research to conduct joint optimization for keyword selection and keyword matching, by constructing a stochastic keyword targeting model and developing a branch-and-bound solution. We further revealed that the joint keyword targeting optimization is important for profit growth in SSA.

Second, this research addresses the problem of unobserved keyword performance indices in SSA. Specifically, we first comb the relationship between keyword performance indices and keyword matching options on the basis of prior research related to keyword matching (Ramaboa and Fish, 2018; Google Ads, 2021), i.e., different matching type leads to different impression and click-through rate, while cost-per-click and value-per-click are not significantly influenced by different matching options for keywords. Then, taking into account the multi-layered structure of SSA (i.e., each ad group shares a set of keywords focusing on the same promotional product), this research develops a data distribution estimation model for unobserved keyword performance indices (i.e., impression and click-through rate) over three keyword matching options.

In addition, this research enhances the understanding of keyword decisions and is promising to be adapted to other keyword-based advertising forms (e.g., social media advertising and native advertising) (Yang and Gao, 2021).

5.2 Practical Implications

From a pragmatic perspective, findings from this study also shed light on important practical implications for advertisers in SSA. First of all, this research provides a feasible way for advertisers to make keywords targeting decisions in practice. Keyword targeting is a vital optimization decision that should not be ignored. Especially when the SSA market becomes more complicated, it is important to consider both keyword selection and keyword matching simultaneously. Specifically, with more available selected keywords and more possible keyword combinations, the positive effect of keyword targeting on the profit growth becomes more obvious with the increase of budget in SSA campaign. In this case, advertisers should pay more attention on keyword targeting. Second, the optimal keyword targeting solution is ultimately the result of a multifaceted trade-off. Various factors such as selected keywords, keyword matching options, the control of uncertainty and the budget, have influence on the performance of keyword targeting. In order to get an optimal keyword targeting decision, advertisers ought to comprehensively evaluate and utilize these factors in systematic. Third, we also find that our approach helps advertisers find more high-profit yet less-cost keywords through data distribution estimation. Given more specific and accurate keyword performance indices for different matching options, advertisers enrich their keyword portfolios and obtain higher expected profits by searching the whole keyword targeting space for the global optimum. Fourth, in the optimal keyword targeting solutions, the exact and phrase matching types take up considerable proportions. This challenges prior studies (e.g., Ballard, 2013; Amaldoss et al., 2016) where broad match is the most popular option. This reminds advertisers to explore keyword targeting strategies with mixed matching options. Since mixed keyword-matching portfolio contributes to the improvements in monetization, it also suggests advertisers to invest more effort and resources in improving data estimation for keyword performance indices and optimization strategy for keyword targeting.

5.3 Future Research

In the future, we seek to explore dynamic keyword targeting strategies for search advertising campaigns. Specifically, we continuously adjust keyword targeting using an online learning algorithm based on immediate market responses. Moreover, interactive relationships between keywords should be taken into account when making keyword decisions. Furthermore, joint optimization of various advertising decisions such as budget planning and keyword portfolios is interesting to explore in the SSA context.

References

Amaldoss, W., Jerath, K., & Sayedi, A. (2016). Keyword management costs and “broad match” in sponsored search advertising. *Marketing Science*, *35*(2), 259-274.

Arroyo-Cañada FJ. and Gil-Lafuente, J. (2019). A fuzzy asymmetric TOPSIS model for optimizing investment in online advertising campaigns. *Operational Research International Journal*, 19(3), 701-716.

Avadhanula, V., Colini Baldeschi, R., Leonardi, S., Sankararaman, K. A., & Schrijvers, O. (2021, April). Stochastic bandits for multi-platform budget optimization in online advertising. In *Proceedings of the Web Conference* 2021 (pp. 2805-2817).

Ballard, M. (2013). *RKG digital marketing report: Q2 2013*. Available at http://www.rimmkaufman.com/blog/rkg-digital-marketing -report-q2-2013-released/10072013.

Balseiro, S., Kim, A., Mahdian, M., & Mirrokni, V. (2021). Budget-Management Strategies in Repeated Auctions. *Operations Research*, 69(3), 859-876.

Carrigan, G., Barnett, A., Dobson, A., & Mishra, G. (2007). Compensating for missing data from longitudinal studies using WinBUGS. *Journal of Statistical Software*, *19*(7), 1-17.

Chen, M. H., Shao, Q. M., & Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Spinger-Verlag.

Cholette, S.; Özlük, Ö.; and Parlar, M. (2012). Optimal keyword bids in search-based advertising with stochastic advertisement positions. *Journal of Optimization Theory and Applications*, 152(1), 225-244.

Desai, P. S., Shin, W., & Staelin, R. (2014). The company that you keep: When to buy a competitor's keyword. *Marketing Science*, *33*(4), 485-508.

Du, X., Su, M., Zhang, X., & Zheng, X. (2017). Bidding for multiple keywords in sponsored search advertising: Keyword categories and match types. *Information Systems Research*, 28(4), 711-722.

Even Dar, E., Mirrokni, V. S., Muthukrishnan, S., Mansour, Y., & Nadav, U. (2009, April). Bid optimization for broad match ad auctions. In *Proceedings of the 18th International Conference on World Wide Web* (pp. 231-240).

Gamerman, D., & Lopes, H. F. (2006). *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference*. Boca Raton, FL: Chap- man & Hall/CRC.

Ghose, A., & Yang, S. (2009). An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management Science*, 55(10), 1605-1622.

Google Ads. *About keyword matching options*. Available at https://support.google.com/google-ads/answer/7478529?hl=en (accessed on November 15, 2021).

Google Ads. *How Ad Groups Work*. Available at https://support.google.com/google-ads/answer/2375404?hl=en (accessed on November 15, 2021).

Grbovic, M., Djuric, N., Radosavljevic, V., Silvestri, F., Baeza-Yates, R., Feng, A., Ordentlich, E., Yang, L. & Owens, G. (2016, July). Scalable semantic matching of queries to ads in sponsored search advertising. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 375-384).

Gupta, S., Bilenko, M., & Richardson, M. (2009, June). Catching the drift: learning broad matches from clickthrough data. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1165-1174).

Huang, H., & Kauffman, R. J. (2011). On the design of sponsored keyword advertising slot auctions: An analysis of a generalized second-price auction approach. *Electronic Commerce Research and Applications*, 10(2), 194-202.

Interactive Advertising Bureau. *2021 Marketplace Outlook Survey Results*. Available at https://www.iab.com/wp-content/uploads/2020/12/2021-IAB-Marketplace-Outlook-Dec-2020-FINAL.pdf (accessed on January 28, 2021).

Jeziorski, P., & Moorthy, S. (2018). Advertiser prominence effects in search advertising. *Management Science*, 64(3), 1365-1383.

Jones, R., Rey, B., Madani, O., & Greiner, W. (2006, May). Generating query substitutions. In *Proceedings of the 15th International Conference on World Wide Web* (pp. 387-396).

Kim, A. J., Jang, S., & Shin, H. S. (2021). How should retail advertisers manage multiple keywords in paid search advertising?. *Journal of Business Research*, 130, 539-551.

Kiritchenko, S., & Jiline, M. (2008, September). Keyword optimization in sponsored search via feature selection. In *New Challenges for Feature Selection in Data Mining and Knowledge Discovery* (pp. 122-134). Antwerp, Belgium: JMLR.

Klapdor, S., Anderl, E. M., von Wangenheim, F., & Schumann, J. H. (2014). Finding the right words: The influence of keyword characteristics on performance of paid search campaigns. *Journal of Interactive Marketing*, *28*(4), 285-301.

Kosuch, S., and Lisser, A. (2010). Upper bounds for the 0–1 stochastic knapsack problem and a B&B algorithm. *Annals of Operations Research*, 176(1), 77–93.

Küçükaydin, H., Selçuk, B., & Özlük, Ö. (2020). Optimal keyword bidding in search-based advertising with budget constraint and stochastic ad position. *Journal of the Operational Research Society*, 71(4), 566-578.

Levin, J., & Milgrom, P. (2010). Online advertising: Heterogeneity and conflation in market design. *American Economic Review*, 100(2), 603-607.

Li, H., & Yang, Y. (2020). Optimal keywords grouping in sponsored search advertising under uncertain environments. *International Journal of Electronic Commerce*, 24(1), 107-129.

Li, H., Kannan, P. K., Viswanathan, S., & Pani, A. (2016). Attribution strategies and return on keyword investment in paid search advertising. *Marketing Science*, 35(6), 831-848.

Lian, Y., Chen, Z., Pei, X., Li, S., Wang, Y., Qiu, Y., et al. (2021, April). Optimizing AD Pruning of Sponsored Search with Reinforcement Learning. In *Companion Proceedings of the Web Conference 2021* (pp. 123-127).

Lo, S. K., Hsieh, A. Y., & Chiu, Y. P. (2014). Keyword advertising is not what you think: Clicking and eye movement behaviors on keyword advertising. *Electronic Commerce Research and Applications*, 13(4), 221-228.

Lobo, M. S., Vandenberghe, L., Boyd, S. & Lebret, H. (1998). Applications of second-order cone programming. *Linear Algebra and Its Applications*, 284(1-3), 193-228.

Lu, S., & Yang, S. (2017). Investigating the spillover effect of keyword market entry in sponsored search advertising. *Marketing Science*, 36(6), 976-998.

Mahdian, M., & Wang, G. (2009, September). Clustering-based bidding languages for sponsored search. In European Symposium on Algorithms (pp. 167-178). Springer, Berlin, Heidelberg.

Nagpal, M., & Petersen, J. A. (2020). Keyword Selection Strategies in Search Engine Optimization: How Relevant is Relevance?. *Journal of Retailing*, 1-18.

Nie, H., Yang, Y., & Zeng, D. (2019). Keyword generation for sponsored search advertising: Balancing coverage and relevance. *IEEE Intelligent Systems*, 34(5), 14-24.

Park, H. (2020). MLP modeling for search advertising price prediction. *Journal of Ambient Intelligence and Humanized Computing*, 11(1), 411-417.

Qiao, D., Zhang, J., Wei, Q., & Chen, G. (2017). Finding competitive keywords from query logs to enhance search engine advertising. *Information & Management*, 54(4), 531-543.

Radlinski, F., Broder, A., Ciccolo, P., Gabrilovich, E., Josifovski, V., & Riedel, L. (2008, July). Optimizing relevance and revenue in ad search: a query substitution approach. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 403-410).

Ramaboa, K. K., & Fish, P. (2018). Keyword length and matching options as indicators of search intent in sponsored search. *Information Processing & Management*, *54*(2), 175-183.

Rusmevichientong, P., & Williamson, D. P. (2006, June). An adaptive algorithm for selecting profitable keywords for search-based advertising services. In *Proceedings of the 7th ACM Conference on Electronic Commerce* (pp. 260-269). Ann Arbor, Michigan. ACM.

Scholz, M., Brenner, C., & Hinz, O. (2019). AKEGIS: automatic keyword generation for sponsored search advertising in online retailing. *Decision Support Systems*, 119, 96-106.

Schultz, C. D. (2020). The impact of ad positioning in search engine advertising: a multifaceted decision problem. *Electronic Commerce Research*, 20(4), 945-968.

Singh, S. K., & Roychowdhury, V. P. (2008) To broad-match or not to broad-match: An auctioneer's dilemma. In *Proceedings of the fourth Workshop on Ad Auctions, Conference on Electronic Commerce (EC'08)* (pp. 1-33), Chicago, IL.

Song, Z., Chen, J., Zhou, H., & Li, L. (2021, March). Triangular Bidword Generation for Sponsored Search Auction. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining* (pp. 707-715).

Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2003). *WinBUGS user manual*.

Statista. *Search Advertising*. Available at https://www.statista.com/outlook/219/100/search-advertising/worldwide (accessed on January 28, 2021).

Szymanski, G., & Lininski, P. (2018, September). Model of the effectiveness of Google Adwords advertising activities. In *2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)* (Vol. 2, pp. 98-101). IEEE.

Vragov, R., Di Shang, R., Smith, V., & Porter, D. (2019). Let's play the search game: Strategic and behavioral properties of sponsored search auction mechanisms. *Electronic Commerce Research and Applications*, 33, 100809.

Wächter, A. & Biegler, L. T. (2006). On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming,* 106(1), 25-57.

WordStream. *Google Ads Match Types: How Do Keyword Match Types Work in Google?* Available at https://www.wordstream.com/keyword-match-types (accessed on January 28, 2021).

Yang, C., & Xiong, Y. (2020). Nonparametric advertising budget allocation with inventory constraint. *European Journal of Operational Research*, 285(2), 631-641.

Yang, S., Pancras, J., & Song, Y. A. (2021a). Broad or exact? Search Ad matching decisions with keyword specificity and position. *Decision Support Systems*, 143, 113491.

Yang, W., Xiao, B., & Wu, L. (2020). Learning and pricing models for repeated generalized second-price auction in search advertising. *European Journal of Operational Research*, 282(2), 696-711.

Yang, Y. and Zhai, P. (2022). Click-through rate prediction in online advertising: A literature review. *Information Processing & Management*, 59(2), 102853.

Yang, Y., & Gao, T. L. (2021). The path to people's responses to native advertising in social media: A perspective of self-presentational desire. Information & Management, 58(3), 103441.

Yang, Y., Feng, B., Salminen, J., & Jansen, B. J. (2021b). Optimal advertising for a generalized Vidale–Wolfe response model. *Electronic Commerce Research*, forthcoming.

Yang, Y., Jansen, B. J., Yang, Y., Guo, X., & Zeng, D. (2019). Keyword optimization in sponsored search advertising: A multilevel computational framework. *IEEE Intelligent Systems*, 34(1), 32-42.

Yang, Y., Li, X., Zeng, D., & Jansen, B. J., (2018). Aggregate Effects of Advertising Decisions: A Complex Systems Look at Search Engine Advertising via an Experimental Study, *Internet Research*, 28(4), 1079-1102.

Yang, Y., Yang, Y. C., Jansen, B. J. and Lalmas, M. (2017). Computational Advertising: A Paradigm Shift for Advertising and Marketing?. *IEEE Intelligent Systems*, 32(3), 3-6.

Yang, Y., Zeng, D., Yang, Y., and Zhang, J. (2015). Optimal budget allocation across search advertising markets. *INFORMS Journal on Computing*, 27(2), 285-300.

Yang, Y., Zhang, J., Qin, R., Li, J., Liu, B., and Liu, Z. (2013). Budget strategy in uncertain environments of search auctions: A preliminary investigation. *IEEE Transactions on Services Computing*, 6(2), 168-176.

Yang, Y., Zhang, J., Qin, R., Li, J., Wang, F. Y., & Qi, W. (2012). A budget optimization framework for search advertisements across markets. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 42(5), 1141-1151.

Zhang, J., Zhang, J., & Chen, G. (2021). A semantic transfer approach to keyword suggestion for search engine advertising. *Electronic Commerce Research*, forthcoming.

Zhang, Y., Zhang, W., Gao, B., Yuan, X., & Liu, T. Y. (2014). Bid keyword suggestion in sponsored search based on competitiveness and relevance. *Information Processing & Management*, 50(4), 508-523.