Optimal Bidding Strategy for Brand Advertising

Takanori Maehara[1,2]**, Atsuhiro Narita**[2]**, Jun Baba**[2]**, Takayuki Kawabata**[2]

[1] RIKEN Center for Advanced Intelligence Project

[2] CyberAgent Inc.

takanori.maehara@riken.jp, narita_atsuhiro@cyberagent.co.jp,
baba_jun@cyberagent.co.jp, kawabata_takayuki@cyberagent.co.jp

Abstract

Brand advertising is a type of advertising that aims at increasing the awareness of companies or products. This type of advertising is well studied in economic, marketing, and psychological literature; however, there are no studies in the area of computational advertising because the effect of such advertising is difficult to observe.

In this study, we consider a real-time biding strategy for brand advertising. Here, our objective to maximizes the total number of users who remember the advertisement, averaged over the time. For this objective, we first introduce a new objective function that captures the cognitive psychological properties of memory retention, and can be optimized efficiently in the online setting (i.e., it is a monotone submodular function). Then, we propose an algorithm for the bid optimization problem with the proposed objective function under the second price mechanism by reducing the problem to the online knapsack constrained monotone submodular maximization problem. We evaluated the proposed objective function and the algorithm in a real-world data collected from our system and a questionnaire survey.

We observed that our objective function is reasonable in real-world setting, and the proposed algorithm outperformed the baseline online algorithms.

1 Introduction

Background and Motivation *Real-time bidding* is the most successful framework for online advertising [Zeff and Aronson, 1999]. In this framework, each *impression* (a user's view of a website) is instantly sold in an auction, and each advertiser bids for the impression to display his advertisement to the user. An important advantage of this framework is the *personalization*: it can optimize advertisements for each user by selecting impressions. This reduces ineffective advertisements with their wasteful cost, and increases the effect of advertisements; thereby it increases the advertising revenue.

The fundamental optimization problem in real-time bidding is the *bid optimization problem*. The task of the bid optimiza-

tion problem is to find the optimal bid price for each impression to maximize the advertising effect under limited budget. Since we do not know the winnable price and the advertising effect in advance to the bidding, the problem is an optimization problem with uncertainty. This problem has been studied in various setting, e.g., [Zhou *et al.*, 2008; Mehta *et al.*, 2007; Chakrabarti and Kale, 2015; Buchbinder and Naor, 2009].

In this study, we consider a bid optimization problem for *brand advertising* [Hansen and Christensen, 2003]. Brand advertising aims at increasing the *awareness* of the companies or products, as opposed to the *direct response advertising*, which aims at selling specific products. As indicated by marketing studies, consumers' purchasing behaviors are influenced by their brand knowledge and familiarity [Chen and He, 2003; Park and Stoel, 2005]. Therefore, brand advertising is considered a useful tool to increase future revenues.

Brand advertising is studied in the area of economics [De Mooij and Hofstede, 2010], marketing [Randall, 2000; Zielske, 1959], and psychology [Aaker and Biel, 2013]. However, to the best of our knowledge, there are no existing studies in the area of online (computational) advertising. This may be because evaluating the effect of brand advertising is difficult since it requires conducting questionnaire surveys, or analyzing search engine queries and access logs; however, these cannot be obtained immediately and are difficult to quantify; therefore, they cannot be used in real-time bidding systems. On the other hand, brand advertising appears highly compatible with online advertising because it can be personalized for each user. For example, if a user is already exposed to an advertisement with a high frequency, it may be better to limit the user's level of exposure since the user may already recall the advertisement, and to show the advertisement to other users. Also, if a user is exposed to an advertisement with a low frequency, it may be good to show more advertisement to the user to make a strong retention about the advertisement.

To achieve large advertising effects, we have to balance the above two factors (i.e., broad users vs strong retentions). However, it is very difficult to find the optimal balance by hand. Therefore, we have to formulate user's memory retention by a mathematical model, and propose an online algorithm algorithm for brand advertising that automatically balance the factors.

Contributions In this study, we make the following theoretical contributions:

- We propose a new objective function to represent the brand advertising effect. This function represents the users' *forgetting curve* (*memory retention over time*), which is consistent with the existing studies in psychology [Ebbinghaus, 1913]. Moreover, our function is a *monotone submodular function*[1], which admits efficient optimization algorithms in various situations (Section 2).

- We propose an algorithm for the bid optimization problem under the *second price mechanism*[2],which is commonly used in real-world systems. We first suppose that the winnable price of each impression is *known*. In this case, the problem is formulated as an online knapsack constrained submodular maximization problem, and we derive an algorithm for this case with a competitive ratio[3] of $(2 - L/U + O(\epsilon)) \log(Ue/L)$, where e is the base of the natural logarithm, ϵ is the capacity ratio, and U and L are the lower and upper bounds of marginal efficiency, respectively. Then, we convert the algorithm for the case that the winnable price of each impression is *unknown* by using the technique introduced in [Zhou *et al.*, 2008] (Section 3).

We also made the following experimental contributions:

- We conducted a questionnaire survey for the awareness of an advertisement. The result shows the submodularity of the forgetting curve in a real-world dataset; hence it supports the validity of the proposed objective function.

- We conducted a numerical experiment to evaluate the proposed algorithm on a real-world dataset collected from our real-time bidding system. We observed that the proposed algorithm has an issue in practical use that it exhausted the budget too early (Section 4).

2 Problem Formulation

2.1 General Framework

We consider an advertiser's optimization problem. Let $V = \{1, \ldots, n\}$ be a sequence of impressions that arrives in a one-by-one manner in an online context. Each impression $j \in V$ has arriving time t_j, target user u_j, and some features, e.g., ip address, browser, webpage, etc. When impression j arrives, the advertiser must determine the bid price $\tilde{c}_j \in \mathbb{R}$ for this impression in real-time under the condition that the total spent budget is at most his budget $B \in \mathbb{R}$. We assume the second price mechanism. Then, an advertiser wins the auction if his

bid price is greater than the highest bid $c_j \in \mathbb{R}$ made by the other advertisers, and his payment is c_j. If he loses the auction, he does not spend any money.

Let $f : 2^V \to \mathbb{R}$ be a function such that $f(X)$ represents the *advertising effect* when the advertiser presents the advertisements for impressions $X \subseteq V$. We only consider a non-stochastic advertising effect, e.g., an expected value of something. Then, the problem is mathematically formulated as follows:

$$\begin{aligned}
\text{maximize} \quad & f(X) \\
\text{subject to} \quad & c(X) \leq B, \\
& X = \{j \in V : \tilde{c}_j \geq c_j\}, \ \ \tilde{c}_j \in \mathbb{R}_+
\end{aligned} \quad (2.1)$$

where $c(X) = \sum_{j \in X} c_j$.

2.2 Objective Function

The primary consideration in the problem (2.1) is what objective function f should be used. For brand advertising, f should correspond to *the expected number of individuals who recall the advertisement*, integrated over time. By ignoring the interaction between users, we obtain the following formula:

$$f(X) = \int_{-\infty}^{\infty} g(X; t)dt = \int_{-\infty}^{\infty} \sum_{u:\text{users}} g_u(X_u; t)dt \quad (2.2)$$

where X_u denotes the subset of X associated with user u, and $g_u(X_u; t)$ denotes the probability that user u recalls the advertisement at time t when the advertisements are displayed for impressions $j \in X_u$.[4]

In cognitive psychology, the function of memory retention is referred to as the *forgetting curve*. [Ebbinghaus, 1913] observed in his seminal work that a forgetting curve has the following properties:

1. A forgetting curve decreases rapidly (rapid decreasing property).

2. The decline of a forgetting curve slows down due to repetition (repetition effect).

Now these are well-established facts in cognitive psychology [Baddeley, 1999]. However, a theoretical model that is unanimously agreed upon by several researchers does not exist [Averell and Heathcote, 2011]. Moreover, existing models in cognitive psychology [Greene, 1989; Rawson and Dunlosky, 2013] are not suited for use in mathematical and algorithmic analysis. Therefore, we introduce a new mathematical model of the forgetting curve, which captures the above two properties and admits an efficient algorithm.

Let $p_j : \mathbb{R} \to [0, 1]$ be an arbitrary rapidly decreasing function[5], which denotes the forgetting curve of a user when he is exposed to just one advertisement by impression j. Typically, the following Wickelgren's power law [Wickelgren, 1974; Wixted and Carpenter, 2007] can be employed:

$$p_j(t) = \begin{cases} 0 & t < t_j, \\ \lambda_j(1 + \beta_j(t - t_j))^{-\gamma_j}, & t \geq t_j, \end{cases} \quad (2.3)$$

---

[1]A set function $f : 2^V \to \mathbb{R}$ is said to be *monotone* if for any $X \subseteq Y$, $f(X) \leq f(Y)$. f is said to be *submodular* if for any $j \in V$ and $X \subseteq Y \subseteq V \setminus \{j\}$, $f(j|X) \geq f(j|Y)$ where $f(j|X) = f(X \cup \{j\}) - f(X)$ is termed as the *marginal contribution of j at X*.

[2]The *second price mechanism* is an auction mechanism such that the highest bidder wins the auction and pays the second-highest bid price.

[3]Algorithm ALG has *competitive ratio* of α if for any input sequence σ, $\text{OPT}(\sigma) \leq \alpha \text{ALG}(\sigma)$ where OPT is an optimal offline algorithm.

[4]We can add any kernel function $k(t)$ to the integral as $f(X) = \int_{-\infty}^{\infty} g(X; t)k(t)dt$ to give a weighting to the time.

[5]We say that a function $p : \mathbb{R} \to [0, 1]$ is rapidly decreasing if $\int_{-\infty}^{\infty} p(t)dt < \infty$
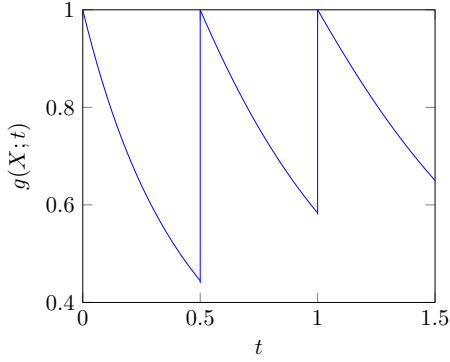
Figure 1: Forgetting curve of our model. It captures the repetition effect.

where t_j is the arrival time of impression j and λ_j, β_j, and γ_j are the parameters. The parameters would be determined by the features of the impression j. Then, the forgetting curve for user u when he is exposed to *multiple* advertisements for impressions X_u is defined as follows:

$$g_u(X_u; t) = 1 - \prod_{j \in X_u} (1 - p_j(t)). \qquad (2.4)$$

As shown below, this function captures the repetition effect of the forgetting curve.

The definition of function (2.4) is motivated by *study-phase retrieval theory* [Greene, 1989]. This assumes that the previous presentations (study-phase) are retrieved later, and this causes the repetition effect of the forgetting curve. Here, (2.4) shows that the probability of recalling the advertisement at time t is the probability of recalling at least one presentation that is shown for impression $j \in X$, which constitutes the same idea as the study-phase retrieval theory.

2.3 Properties of the Objective Function

Psychological Justification   First, a simple numerical experiment is performed to see that the function is consistent with the standard findings in cognitive psychology.

Suppose that a user u is exposed to three impressions $X_u = \{1, 2, 3\}$ at time $t_1 = 0$, $t_2 = 1/2$, and $t_3 = 1$. We set the parameters of power law (2.3) as $\lambda_j = 1$, $\beta_j = 1$, and $\gamma_j = 2$ for all impressions. Then the corresponding forgetting curve $g_u(X_u; t)$ is obtained as Figure 1, which looks similar to the forgetting curve that we can see in psychological literatures [Ebbinghaus, 1913]. This shows that the proposed function captures the two properties of forgetting curves. Specifically, the function successfully captures the repetition effect, i.e., the decline on the second learning is slower than that on the first learning, and consequently decline on the third learning is slower than that on the second learning.

Algorithmic Property   From an algorithmic viewpoint, the most important property of our objective function f is *monotone* and *submodular*. Therefore, the problem expressed in (2.1) is a kind of *online knapsack constrained monotone submodular maximization problem*. This allows the construction

of an efficient approximation algorithm, which will be presented in Section 3.

Proposition 1. The objective function $f : 2^V \to \mathbb{R}$ that is defined by (2.2) and (2.4) with any integrable functions p_j is monotone submodular.

Proof. We use the fact that the sum of monotone submodular functions is also monotone submodular.

First, we fix time t. Then, $g_u(X; t) = \mathbb{E}[\min\{|X \cap R|, 1\}]$ where $R \subseteq V$ is a random set in which each element j appears with probability $p_j(t)$ and the expectation is with respect to R. Since $\min\{|X \cap R|, 1\}$ is monotone submodular in X for any R, its expectation, g_u, is also monotone submodular in X[6]. Since f is obtained by taking the integral of g_u over t, it is also monotone submodular. $\square$

Function Evaluation   Evaluating the marginal contribution of f is often required to implement an algorithm for the submodular maximization problem. For the proposed objective function (2.2), we have

$$f(j|X) = \int_{-\infty}^{\infty} g_u(j|X_u; t)dt$$
$$= \int_{-\infty}^{\infty} p_j(t) \prod_{i \in X_u} (1 - p_i(t))dt \qquad (2.5)$$

where u is the user for impression j.

Even for Wickelgren's power law (2.3), the above integral (2.5) cannot be evaluated analytically because it involves the hypergeometric function $_2F_1$. Therefore, numerical integration [Davis and Rabinowitz, 2007] must be employed. If $p_j(t)$ is smooth except $t = t_j$, the function $f(j|X)$ has finitely many discontinuities at $t = t_i$ for $i \in X_u$ and is smooth and rapidly decreasing in each interval. Thus, a standard numerical integrator (e.g., Gauss–Lobatto integrator) works well by splitting the integral into the intervals.

The complexity for evaluating the marginal contribution is $O(|X_u|)$. In a real-world problem, this is not very large because it is inefficient to display the advertisement to the same person very often.

3 Algorithm

In this section, we propose an algorithm for the bid optimization problem. We derive our algorithm in the following two steps. First, we suppose that the winnable price c_j is known before making the bid, and derive an algorithm for this case (Section 3.1). Then, we convert the algorithm to the case that the winnable price c_j is unknown (Section 3.2).

3.1 Online Knapsack Constrained Monotone Submodular Maximization Problem

Suppose that the winnable price c_j (the highest bid made by the other advertisers) is known before making a bid. Then,

---

[6]In fact, the submodularity of $g_u(X; t)$ is immediately obtained from the observation that it has the same form of the function used in the budget allocation problem [Alon *et al.*, 2012], which is known to be submodular.

Algorithm 1 Algorithm for online knapsack constrained monotone submodular maximization problem.

1: $X = \emptyset$
2: **for** $j = 1, 2, \ldots$ **do**
3:     **if** $f(j|X) \geq c_j \Psi(c(X)/B)$ **then**
4:         Take j-th impression by paying c_j, i.e., $X \leftarrow X \cup \{j\}$
5:     **end if**
6: **end for**
7: **return** X

---

since we can win the auction by bidding marginally greater than c_j, we can fix our bidding price to 0 (ignore the impression) or c_j (take the impression), and hence the problem is reduced to an *online knapsack constrained monotone submodular maximization problem*:

$$\begin{array}{ll} \text{maximize} & f(X) \\ \text{subject to} & c(X) \leq B, \\ & X \subseteq V. \end{array} \qquad (3.1)$$

where the impressions $j \in V$ arrive in a one-by-one manner. Because the problem involves the standard online knapsack problem, some assumptions are necessary to arrive at a provable competitive ratio [Marchetti-Spaccamela and Vercellis, 1995]. Here, we assume the following conditions:

1. The price of each impression is very small compared to the knapsack capacity, i.e., $c_j/B \leq \epsilon$ for all $j \in V$.
2. The lower and upper bounds of marginal efficiency are known, i.e., U and L such that for any $j \in V$ and $X \subseteq V \setminus \{j\}$, $L \leq (f(j|X))/c_j \leq U$.

$$(3.2)$$

The first condition is usually satisfied in real-time bidding setting. The second condition is "approximately" satisfied by learning the bounds from a training set in practice, but it may cause some difficulty in theory.

The proposed algorithm is shown as Algorithm 1. This is a "submodular version" of the algorithm for the online knapsack problem proposed by Zhou et al. [Zhou *et al.*, 2008]. Let

$$\Psi(z) = (Ue/L)^{z/(1-\epsilon)} (L/e) \qquad (3.3)$$

be a threshold function where z denotes the fraction of the current budget, and e is the base of the natural logarithm. The algorithm bids impression j if its marginal efficiency $f(j|X)/c_j$ exceeds the current threshold $\Psi(z)$. This algorithm has the following theoretical guarantee.

Theorem 2. Given assumptions (3.2), Algorithm 1 has a competitive ratio of $(1 + \kappa + O(\epsilon)) \log(Ue/L)$, where $\kappa \in [0, 1]$ is the curvature of f[7].

Proof. This is obtained by carefully combining submodularity and curvature into the proof of the linear case [Zhou *et al.*, 2008]. For self-completeness, we give a full proof. First, we modify the threshold function by

$$\Psi(z) = \max\{L, (L/e)(Ue/L)^{z-\epsilon}\}. \qquad (3.4)$$

---
[7] A submodular function $f : 2^V \to \mathbb{R}$ has *curvature* $\kappa \in [0, 1]$ if for any $j \in V$ and $X \in V \setminus \{j\}$, $f(j|X) \geq (1 - \kappa)f(j)$.

This does not affect anything to the algorithm. Let z_L be such that $\Psi(z_L) = L$. This is given by

$$z_L = \frac{1}{\log(Ue/L)} + \epsilon. \qquad (3.5)$$

The algorithm bids all the impression when $z < z_L$.

Fix an input sequence V. Let X and X^* be the obtained solution and the optimal solution, respectively. Note that all the impressions in X and X^* are winnable. Let X_j and $(X^* \cup X)_j$ denote the subset of X and $X^* \cup X$ that arrive before j-th iteration, respectively. Let $z_j \in [0, 1]$ and $Z \in [0, 1]$ be the fraction of the budget filled by the algorithm at j-th iteration and the end of the algorithm.

In general,

$$f(X^*) = f(X^* \cup X) - (f(X^* \cup X) - f(X^*)). \qquad (3.6)$$

Here, each term in the right-hand side is evaluated as follows. For the first term, by telescoping and submodularity,

$$f(X^* \cup X) = \sum_{j \in X^* \cup X} f(j|(X^* \cup X)_j) \leq \sum_{j \in X^* \cup X} f(j|X_j)$$

$$= \sum_{j \in X} f(j|X_j) + \sum_{j \in X^* \cup X} f(j|X_j) \qquad (3.7)$$

For the second term, by telescoping, curvature, and submodularity,

$$f(X^* \cup X) - f(X^*) \geq (1 - \kappa) \sum_{j \in X \setminus X^*} f(j)$$

$$\geq (1 - \kappa) \sum_{j \in X \setminus X^*} f(j|X_j). \qquad (3.8)$$

Therefore

$$f(X^*) \leq \sum_{j \in X} f(j|X_j) + \sum_{j \in X^* \setminus X} f(j|X_j)$$

$$- (1 - \kappa) \sum_{j \in X \setminus X^*} f(j|X_j). \qquad (3.9)$$

Also, by telescoping,

$$f(X) = \sum_{j \in X} f(j|X_j). \qquad (3.10)$$

Therefore

$$\frac{f(X^*)}{f(X)} \leq \frac{\text{RHS of (3.9)}}{\text{RHS of (3.10)}}. \qquad (3.11)$$

Since the algorithm bid $j \in X$, the first term in the numerator (and the denominator) is evaluated as

$$\sum_{j \in X} f(j|X_j) \geq \sum_{j \in X} c_j \Psi(z_j). \qquad (3.12)$$

Also, the third term in the denominator is evaluated as

$$\sum_{j \in X \setminus X^*} f(j|X_j) \geq \sum_{j \in X \setminus X^*} c_j \Psi(z_j). \qquad (3.13)$$

Similarly, since the algorithm did not bid $j \in X^* \setminus X$, and the algorithm never rejects impressions due to the shortage of

Algorithm 2 Algorithm for bid optimization problem.

1: $X = \emptyset$
2: **for** $j = 1, 2, \dots$ **do**
3:     Bid $\tilde{c}_j = f(j|X)/\Psi(c(X)/B)$ for j-th impression
4:     **if** we won the auction **then**
5:         $X \leftarrow X \cup \{j\}$
6:     **end if**
7: **end for**
8: **return** X

the budget since $\Psi(z) > U$ if $z > 1 - \epsilon$, the second term is evaluated as

$$\sum_{j \in X^* \setminus X} f(j|X_j) \leq \sum_{j \in X^* \setminus X} c_j \Psi(z_j). \qquad (3.14)$$

Therefore, by monotonicity of fraction, monotonicity of Ψ, and feasibility of X and X^*,

$$\frac{f(X^*)}{f(X)} \leq \frac{\sum_{j \in X^*} c_j \Psi(z_j) + \kappa \sum_{j \in X \setminus X^*} c_j \Psi(z_j)}{\sum_{j \in X} c_j \Psi(z_j)}$$
$$\leq \frac{(1 + \kappa)B\Psi(Z)}{\sum_{j \in X} \Psi(z_j)c_j} = \frac{(1 + \kappa)\Psi(Z)}{\sum_{j \in X} \Psi(z_j)\Delta z_j} \quad (3.15)$$

where Δz_j denotes the difference between the consecutive z_j for $j \in X$, i.e., $\Delta z_j = c_j/B$. The summation in the denominator is approximated by the integral, and is evaluated as follows:

$$\sum_{j \in X} \Psi(z_j)\Delta z_j \geq \int_0^{Z - \epsilon} \Psi(z)dz$$
$$= \int_0^{z_L} L dz + \int_{z_L}^{Z - \epsilon} (L/e)(Ue/L)^{z - \epsilon} dz$$
$$= z_L L + \frac{\Psi(Z - 2\epsilon) - \Psi(z_L)}{\log(Ue/L)}$$
$$= \frac{(1 - O(\epsilon))\Psi(Z)}{\log(Ue/L)}. \qquad (3.16)$$

The theorem is obtained from (3.15) and (3.16). □

Under assumptions (3.2) the curvature of f satisfies $\kappa \leq 1 - L/U$. Therefore, the following corollary is obtained by removing κ from Theorem 2.

Corollary 3. Given assumptions (3.2), Algorithm 1 has a competitive ratio of $(2 - L/U + O(\epsilon))\log(Ue/L)$. □

The competitive ratio in Theorem 2 is optimal if $\kappa = 0$ [Zhou *et al.*, 2008]. Note that when $\kappa \to 1$, we have $L/U \to 0$; therefore the approximation ratio becomes unbounded. The tightness of the competitive ratio for $\kappa \in (0, 1)$ is open.

3.2 Bid Optimization Problem

Under the second price mechanism, the algorithm proposed in the previous section can be converted to the algorithm for the bid optimization problem, i.e., it can be executed *without knowing* winnable prices. This technique is originally introduced in [Zhou *et al.*, 2008].

Our bid optimization algorithm (Algorithm 2) maintains the impressions X that we made successful bids, and for each arriving impression j, it determines the bid price by $f(j|X)/\Psi(c(X)/B)$. This value can be computed in real-time bidding setting since we know the winnable prices c_i for $i \in X$. We prove the following.

Theorem 4. Under the second price mechanism, Algorithm 1 and Algorithm 2 produce the same solution with the same payment.

Proof. Due to the second price mechanism, the payment for impressions does not depend on the bid prices. Thus we only have to show that the impressions obtained by Algorithm 1 and Algorithm 2 are the same.

We prove this by the induction. Consider the step that impression j is arrived. By the induction, so far, Algorithm 1 and Algorithm 2 obtained the same set of impression X. If j satisfy $f(j|X) \geq c_j\Psi(c(X)/B)$, Algorithm 1 takes this impression. Also, Algorithm 2 bids $\tilde{c}_j = f(j|X)/\Psi(c(X)/B)$, which is greater than or equal to c_j; thus it wins the auction and gets this impression. The same is true for the converse case $(f(j|X) < c_j\Psi(c(X)/B))$. □

This means that Algorithm 2 perfectly emulates Algorithm 1 without knowing c_j; hence, it has the same theoretical guarantees given in Theorem 2 and Corollary 3.

4 Experiments

Numerical experiments were conducted to evaluate the proposed model and algorithm.

All codes were implemented in Python 3.5. For the numerical integration, we use the adaptive Gauss–Lobatto integrator, implemented in scipy.integrate. The program was executed in a standard laptop computer (2.3GHz CPU, 8.0GB Memory).

The datasets used in the experiments are collected by our real-time bidding system. We selected a campaign of brand advertising that has sufficiently many impressions and unique users. Then, we collected bid log data, which is used to evaluate the performance of the algorithm by a numerical simulation. We also conducted a questionnaire survey to justify the our proposed psychological model.

4.1 Model Justification by Questionnaire Survey

A few days after the end of the campaign, a questionnaire survey was conducted electronically through a web survey company.[8] The survey included questions such as "Do you remember this advertisement?." The surveys were returned by 10,000 users, 2,000 of them had watched the advertisement, and 426 of them remembered the advertisement. The fraction of the returned survey by the latter users with the answer "YES" was considered to denote awareness.

Points in Figure 2 shows the relationship between awareness and the time elapsed since the last time when the users watch

---

[8]Each respondent is a member of the web survey company. By using the cookies, the company can identify the number of times each respondent was exposed to the advertisement. Note that the members agree to fill in the questionnaire for the purpose of research; thus there are no ethical issue.

the advertisement. This can be interpreted as a forgetting curve for brand advertising. By manually fitting (2.3) to the points in Figure 2, we obtain the following formula:

$$p_0(t) = 0.13 + (1 + t)^{-1.3}, \qquad (4.1)$$

which is presented as the curve in Figure 2.

Figure 3 shows the relationship between the number of times the users see the advertisements and their awareness. A three-point moving average was performed for smoothing. This indicates the diminishing return property; therefore modeling by a submodular function would be reasonable.

4.2 Evaluation of Bidding Algorithm

Dataset Description   To evaluate the performance of the algorithm in a real-world setting, the bidding log data on two days (November 12 and 13, 2016) of the campaign are collected from our system. The first day's data, which comprises of 1,440,641 impressions with 731,029 unique users, is used for a training set and the second day's data, which comprises of 1,455,002 impressions with 735,560 unique users (including the users who appeared in the first day's data) is used for an evaluation set. Each impression is marked with its arrival time, user id, and winnable price.[9]

The distribution of the number of presented advertisements related to the users, which follows a power-law distribution (Figure 4). Specifically, there is a user who had too many presentations. Therefore the bid optimization is promising for this data.

Experimental Setting   For the purpose of the experiments, we determined the shape of the forgetting curve. We used the same forgetting curve function for all users as

$$p_j(t) = p_0'(t - t_j), \qquad (4.2)$$

where p_0' is a function estimated in (4.1) where the constant term is ignored since the constant term can be understood as the fraction of people who already know the advertisement.

This setting means that we ignored the difference of users, which is reasonable because we do not have sufficiently many dataset to learn the difference of the users. In addition, it is also reasonable in psychological viewpoint because there are no significant difference in the memory retention ability among the individuals [Ebbinghaus, 1913].

Algorithms   We implemented three online algorithms (`Proposed`, `FixedBid`, and `Random`) and one offline algorithm (`Offline`).

- `Proposed` is the proposed algorithm with the parameters estimated in Section 4.1. The lower and upper bounds L and U were optimized by the training set.

---

[9]Since we need the winnable prices of all impressions to evaluate the performance of the algorithms, we only use the impressions that are obtained by the successful bids. Our system used the fixed price bidding strategy; therefore the all winnable prices are at most the fixed price. This makes dataset biased.

Algorithm	Objective Value
Proposed	52188 (45%)
FixedBid	38577 (33%)
Random	40233 (34%)
Offline	116129 (100%)

Table 1: Performance of algorithms.

- `FixedBid` is the algorithm that always bids a fixed price. The fixed price is optimized by the training set. Note that this is a baseline method, which is implemented in our current system.

- `Random` is the algorithm that bids randomly for each impression. Again, the probability is determined for spending the budget. Note that this is also implemented in our current system.

- `Offline` is the offline greedy algorithm for the knapsack constrained monotone submodular maximization problem [Khuller *et al.*, 1999], which has an approximation ratio of $1 - 1/e - O(\epsilon)$. Because this algorithm determines the bidding strategy by taking future impressions into account, this will give an upper-bound of the performance.

Result   The performance (obtained advertising effect) of these algorithms are summarized in Table 1. The proposed algorithm has 1.4 times larger advertising effect (i.e., $f(X)$) than the two baseline online algorithms. This is not very large, but a solid improvement from the baseline algorithms. On the other hand, the proposed algorithm has a room to improve the practical performance since the offline greedy algorithm attains 2 times larger advertising effect.

Figure 5 shows the total amount of memory retentions, $g(X; t)$, in each time, obtained by the algorithms. Here, the area under the curve is the objective value. From this figure, we can point out that the proposed algorithm exhausted the budget in early stage, compared with other strategies. Therefore, it overlooked the efficient impression in later stage. The reason of this phenomena is that the algorithm is designed to have a guaranteed competitive ratio, i.e., it must have a provable objective value *even if the input is immediately terminated in early stage*. To overcome this issue, we may need some (possibly stochastic) information about future impressions.

5 Related Work

Bid Optimization Problem   The bid optimization problem is a very actively studied problem in real-time bidding, and several formulations and algorithms have been proposed [Zhou *et al.*, 2008; Mehta *et al.*, 2007; Chakrabarti and Kale, 2015; Buchbinder and Naor, 2009; Cai *et al.*, 2017; Zhang *et al.*, 2016; 2014]. However, as mentioned in Section 1, there is a paucity of studies that focus on brand advertising. Mathematically, the existing studies formulated the advertising effect as a sum of effects of the impressions; thus they ignored the correlation effect between the impressions. In brand advertising, this correlation is important due to the repetition effect of the forgetting curve; thus we need a new formulation.
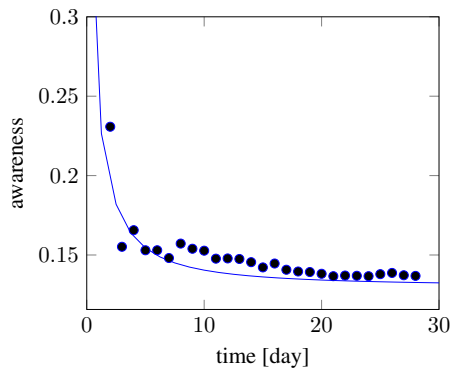
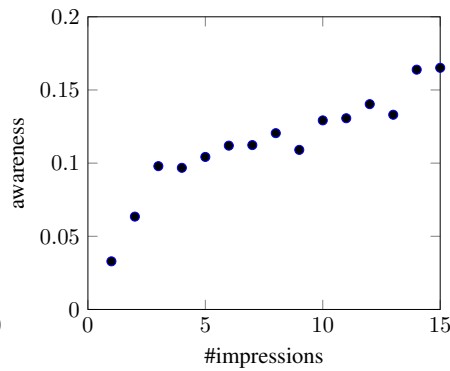Figure 2: Forgetting curve in a real campaign.

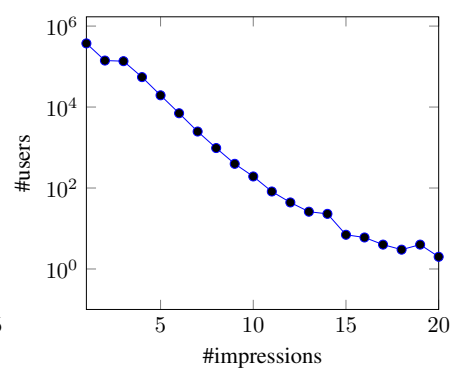Figure 3: Histogram of the number of advertisements presented.
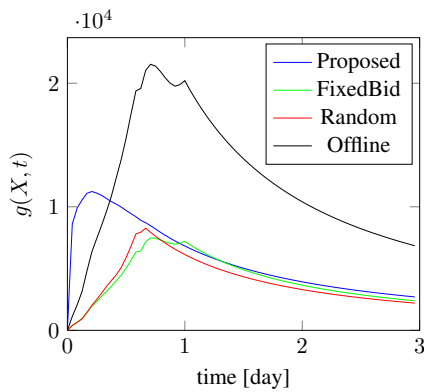
Figure 4: Histogram of the number of impressions.

Figure 5: Forgetting curves $g(X; t)$ of each algorithm. The campaign is stopped at $t = 1$.

Online Knapsack Constrained Monotone Submodular Maximization We formulated the bid optimization problem as an online knapsack constrained monotone submodular maximization problem and proposed the algorithm. Our algorithm is based on the algorithm for the online knapsack problem proposed by [Zhou *et al.*, 2008] since it does not assume anything about the arrival order of items. This is an important property for us since the impressions have natural ordering according to the arrival time; hence, we cannot use the algorithms that have better approximation guarantee under the random arrival assumption (e.g., [Lueker, 1998] and [Babaioff *et al.*, 2007]).

To best of our knowledge, there is no algorithm that can be applied to the present formulation of bid optimization problem. Specifically, it is not possible to apply the algorithms for the *knapsack constraint submodular secretary problem* [Bateni *et al.*, 2010; Kesselheim and Tönnis, 2016] because they assume the arrival order of items is random. Also, it is not possible to apply the algorithm for *streaming knapsack constrained submodular maximization problem* [Kumar *et al.*, 2015] because the algorithm keeps more than one solutions, which is impossible to use in real-time bidding systems.

Forgetting Curves For the function modeling, in cognitive psychology, following the seminal study by Ebbinghaus [Ebbinghaus, 1913], several models for the forgetting curve were proposed [Greene, 1989; McClelland *et al.*, 1995; Murre *et al.*, 2013]. However, most of the models focused on understanding the "actual mechanism" of human memory. The present study is the first study that formulates memory retention as a computationally tractable model.

6 Conclusion

In this study, we proposed a new formulation of the bid optimization problem for brand advertising. We introduced a psychologically reasonable and computationally tractable objective function. Then we derive a bid optimization algorithm with a provable competitive ratio. The validity of our objective function is justified by the questionnaire survey, and the performance of the algorithm is evaluated by the bid log data collected from a real time bidding system.

There are several directions for future research. First, from a psychological viewpoint, it is important to propose a more realistic (and computationally tractable) model for the forgetting curve. Second, from a theoretical viewpoint, future studies could develop an algorithm that achieves better competitive ratio or prove a matching lower bound. Third, from a practical viewpoint, as shown in Section 4, the proposed algorithm has the potential to improve performance by using the information of remaining time. This may require another formulation of the problem, because the competitive ratio guarantees the worst case input, which can terminate immediately. Finally, apart from computational advertising, it is interesting to develop applications of the proposed objective function and algorithm in other areas such as a learning system for education.

Acknowledgements

References

[Aaker and Biel, 2013] David A Aaker and Alexander Biel. *Brand equity & advertising: advertising's role in building strong brands*. Psychology Press, 2013.

[Alon *et al.*, 2012] Noga Alon, Iftah Gamzu, and Moshe Tennenholtz. Optimizing budget allocation among channels and influencers. In *Proceedings of the 21st International Conference on World Wide Web*, pages 381–388, 2012.

[Averell and Heathcote, 2011] Lee Averell and Andrew Heathcote. The form of the forgetting curve and the fate of memories. *Journal of Mathematical Psychology*, 55(1):25–35, 2011.

[Babaioff *et al.*, 2007] Moshe Babaioff, Nicole Immorlica, David Kempe, and Robert Kleinberg. A knapsack secretary problem with applications. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, pages 16–28. Springer, 2007.

[Baddeley, 1999] Alan D Baddeley. *Essentials of human memory*. Psychology Press, 1999.

[Bateni *et al.*, 2010] MohammadHossein Bateni, MohammadTaghi Hajiaghayi, and Morteza Zadimoghaddam. Submodular secretary problem and extensions. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 39–52. Springer, 2010.

[Buchbinder and Naor, 2009] Niv Buchbinder and Joseph Naor. Online primal-dual algorithms for covering and packing. *Mathematics of Operations Research*, 34(2):270–286, 2009.

[Cai *et al.*, 2017] Han Cai, Kan Ren, Weinan Zhang, Kleanthis Malialis, Jun Wang, Yong Yu, and Defeng Guo. Real-time bidding by reinforcement learning in display advertising. In *Proceedings of the 10th International Conference on Web Search and Data Mining*, 2017.

[Chakrabarti and Kale, 2015] Amit Chakrabarti and Sagar Kale. Submodular maximization meets streaming: Matchings, matroids, and more. *Mathematical Programming*, 154(1-2):225–247, 2015.

[Chen and He, 2003] Rong Chen and Feng He. Examination of brand knowledge, perceived risk and consumers' intention to adopt an online retailer. *Total Quality Management & Business Excellence*, 14(6):677–693, 2003.

[Davis and Rabinowitz, 2007] Philip J Davis and Philip Rabinowitz. *Methods of numerical integration*. Courier Corporation, 2007.

[De Mooij and Hofstede, 2010] Marieke De Mooij and Geert Hofstede. The hofstede model: Applications to global branding and advertising strategy and research. *International Journal of Advertising*, 29(1):85–110, 2010.

[Ebbinghaus, 1913] Hermann Ebbinghaus. *Memory: A contribution to experimental psychology (Original work published in 1885)*. Number 3. University Microfilms, 1913.

[Greene, 1989] Robert L Greene. Spacing effects in memory: Evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(3):371, 1989.

[Hansen and Christensen, 2003] Flemming Hansen and Lars Bech Christensen. *Branding and advertising*. Copenhagen Business School Press DK, 2003.

[Kesselheim and Tönnis, 2016] Thomas Kesselheim and Andreas Tönnis. Submodular secretary problems: Cardinality, matching, and linear constraints. *arXiv preprint arXiv:1607.08805*, 2016.

[Khuller *et al.*, 1999] Samir Khuller, Anna Moss, and Joseph Seffi Naor. The budgeted maximum coverage problem. *Information Processing Letters*, 70(1):39–45, 1999.

[Kumar *et al.*, 2015] Ravi Kumar, Benjamin Moseley, Sergei Vassilvitskii, and Andrea Vattani. Fast greedy algorithms in mapreduce and streaming. *Transactions on Parallel Computing*, 2(3):14, 2015.

[Lueker, 1998] George S Lueker. Average-case analysis of off-line and on-line knapsack problems. *Journal of Algorithms*, 29(2):277–305, 1998.

[Marchetti-Spaccamela and Vercellis, 1995] Alberto Marchetti-Spaccamela and Carlo Vercellis. Stochastic on-line knapsack problems. *Mathematical Programming*, 68(1-3):73–104, 1995.

[McClelland *et al.*, 1995] James L McClelland, Bruce L McNaughton, and Randall C O'reilly. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological review*, 102(3):419, 1995.

[Mehta *et al.*, 2007] Aranyak Mehta, Amin Saberi, Umesh Vazirani, and Vijay Vazirani. Adwords and generalized online matching. *Journal of the ACM*, 54(5):22, 2007.

[Murre *et al.*, 2013] Jaap MJ Murre, Antonio G Chessa, and Martijn Meeter. A mathematical model of forgetting and amnesia. *Frontiers in psychology*, 4:76, 2013.

[Park and Stoel, 2005] Jihye Park and Leslie Stoel. Effect of brand familiarity, experience and information on online apparel purchase. *International Journal of Retail & Distribution Management*, 33(2):148–160, 2005.

[Randall, 2000] Geoffrey Randall. *Branding: A practical guide to planning your strategy*. Kogan Page Limited, 2000.

[Rawson and Dunlosky, 2013] Katherine A Rawson and John Dunlosky. Relearning attenuates the benefits and costs of spacing. *Journal of Experimental Psychology: General*, 142(4):1113, 2013.

[Wickelgren, 1974] Wayne A Wickelgren. Single-trace fragility theory of memory dynamics. *Memory & Cognition*, 2(4):775–780, 1974.

[Wixted and Carpenter, 2007] John T Wixted and Shana K Carpenter. The wickelgren power law and the ebbinghaus savings function. *Psychological Science*, 18(2):133–134, 2007.

[Zeff and Aronson, 1999] Robbin Lee Zeff and Bradley Aronson. *Advertising on the Internet*. John Wiley & Sons, Inc., 1999.

[Zhang *et al.*, 2014] Weinan Zhang, Shuai Yuan, and Jun Wang. Optimal real-time bidding for display advertising. In *Proceedings of the 20th International Conference on Knowledge Discovery and Data Mining*, pages 1077–1086, 2014.

[Zhang *et al.*, 2016] Weinan Zhang, Yifei Rong, Jun Wang, Tianchi Zhu, and Xiaofan Wang. Feedback control of real-time display advertising. In *Proceedings of the 9th International Conference on Web Search and Data Mining*, pages 407–416, 2016.

[Zhou *et al.*, 2008] Yunhong Zhou, Deeparnab Chakrabarty, and Rajan Lukose. Budget constrained bidding in keyword auctions and online knapsack problems. In *Internet and Network Economics*, pages 566–576. 2008.

[Zielske, 1959] Hubert A Zielske. The remembering and forgetting of advertising. *the Journal of Marketing*, pages 239–243, 1959.