The Role of Value of Information Based Metareasoning in Adaptive Sponsored Search Auctions

A thesis presented by

Jimmy J Sun

to
Computer Science
and
Economics
in partial fulfillment of the honors requirements
for the degree of
Bachelor of Arts
Harvard College
Cambridge, Massachusetts

April 2, 2007

Contents

Chapter 1

Introduction

1.1 Overview

The Internet has seen the advent of personalized, context-sensitive advertising on a scale never before seen. One of the most ubiquitous and powerful manifestations of this has been sponsored search advertising, where advertisers pay for their advertisements to appear next to specific keyword search terms in large search engines such as Google and Yahoo!. When a user searches for a relevant keyword, she typically receives a page with her query results in the middle, and a side bar or banner of "sponsored searches," that is, advertisements of companies that have paid to be associated with her keyword. For example, a user searching for "laptop" on Google receives advertisements from Hewlett-Packard, Dell, Sony and Toshiba along with the query results concerning the One Laptop per Child initiative and Laptop magazine. Clicking on one of these advertisements redirects the user to a website specified by the advertiser, whereupon the advertiser is charged a fee by the search engine. This process allows advertisers fine grained targeting of their advertisement campaigns to maximize exposure to relevant, interested potential customers and reduce pointless exposure to uninterested viewers.

These advertisements have a high variability in value and desirability depending on the position and placement of the advertisement, as one placed in a banner at the top of the page is much more likely to be clicked on than the same ad appearing at the bottom of the page. This makes the traditional Internet banner advertising sales model of setting a fixed price per time shown (impression) unsuitable, as it requires the search engine or provider to maintain a large and highly variable set of prices. Auctions have provided an elegant method for price-setting and position allocation based on advertiser input. One of the earliest and simplest methods used is a rank-by-bid mechanism, whereby advertisers are ranked by their bids in decreasing order, with the highest bidding advertisement occupying the highest and most desirable slot, the second highest bidder occupying the second slot, etc.

More complex allocation mechanisms are currently used by most sponsored search providers. In particular, there are a large number of allocation mechanisms which rely on the calculation of an advertisers "rank score", typically the product of the advertiser's bid and a proxy value for its quality or relevance. The most straightforward of these mechanisms is to rank advertisers by the product of their bid and their estimated click-through-rate (CTR), the estimated probability that a given user who is shown the ad will click on it. Indeed, Google's AdSense auctions, in a more transparent incarnation, used precisely this mechanism [6].

The calculation and estimation of click-through-rates are difficult tasks, since the number of clicks an advertisement receives depends both on the advertiser itself, as an advertisement from Sony is (presumably) more likely to be clicked on than an advertisement from Sketchy-Computers.com, and on the position in which the advertisement is placed. All else equal, an advertisement placed in a desirable top slot will receive more clicks than in a lower slot, increasing the amount of information gained about the advertisement and amplifying its effect on net revenue. and With no prior information about advertisers, for example, with a new keyword suddenly becoming relevant for advertisers (e.g. new marketing strategy, new product launch, etc.), a sponsored search provider is forced to quickly learn about the CTRs of the advertisers who submit bids. This can be done through simple deterministic methods such as rotating advertisements through the top slot for some period of time, so that all advertisements are given some maximal exposure, and estimating the CTR based on how many clicks are received in that time period. However, at the same time, the sponsored search provider wants to maximize revenue. This is the classic exploration/exploitation trade-off, as advertisers need to both learn about the advertiser specific CTR and maintain desirable revenue properties.

We attempt to balance these two important goals by exploring a game theoretic model for sponsored search auctions and applying a value of information based metareasoning process in order to rationally decide how to adaptively set auction parameters to concomitantly learn about advertiser click-through-rates and maintain high revenue. The agents are advertisers, each with different per-click valuation, who submit bids to appear in an array of sponsored searches, with the top slot in the array most desirable for an advertiser, and slots decreasing in desirability as they appear further down the page. We focus on a simple and elegant family of mechanisms explored by Lahaie and Pennock that weight advertiser bids by a "squashed click-through-rate estimate," that is, an estimated CTR raised to some exponent. This allows the auction to place variable weights on the importance of CTRs and to set a number of different allocations without drastically changing the rules and equilibria of the auction. While allowing the sponsored search provider greater flexibility in choosing allocations, Lahaie and Pennock further note that this family of mechanisms is also important because it is often the case that the revenue maximizing allocation is neither rank-by-bid or rank-by-revenue, but rather some intermediate mechanism [14]. We are able to characterize equilibrium advertiser behavior by following Varian and Edelman, Ostrovsky and Schwarz in investigating a specific subset of the Nash equilibria of the auction game called symmetric Nash equilibria or locally envy-free equilibria [24, 6]. These equilibria allow us to calculate equilibrium bids and most importantly, equilibrium expected revenue, which will prove to be a critical component of the algorithms we consider, as it allows algorithms to reason meaningfully about current and future revenue streams.

With the ability to calculate expected equilibrium in hand, we extend the single-period advertisement auction to a multiple stage setting, where each time period the provider sets and publishes a squashing factor, advertisers bid their equilibrium bids and revenue and information is received according to the resulting allocation. We then apply a decision theoretic metareasoning process in order to optimally learn about how to adaptively set auction parameters to refine click-through-rate estimates and maintain high revenue. By metareasoning, we mean an algorithm utilizing value of information calculations to evaluate the attractiveness of a set of potential actions (in this case, squashing factors and allocations) based not only on their current period expected revenue, but also on the effect of the information that they will accrue on future revenue.

We adapt the metareasoning framework presented by Russell and Wefald to the multi-stage sponsored search auction setting [20]. This process involves careful consideration of the information properties of the sponsored search domain, as information gathering (computational) and actions in the "real world" (external) are not decoupled as they traditionally are in metadeliberation problems. Comparing the revenue performance and information gathering capabilities of our metareasoning algorithm to heuristic greedy exploitation and exploration algorithms allows us to begin to understand the power and potential of value of information metareasoning for sponsored search auctions.

1.2 Motivation

Keyword advertising using the basic sponsored search model is an already highly valuable and still rapidly growing source of revenue for the Internet advertising industry. Edelman estimates that in 2005, over 98 percent of Google's $6.14 billion revenue was from AdWords, its implementation of sponsored search [6]. While Google dominates the sponsored search industry, sponsored search auctions contributed over half of Yahoo!'s $5.26 billion in 2005 revenue. The industry is still growing with new entrants ranging from giants such as Microsoft to small firms carving out niches with new auction rules or advertisement placement transparency such as Quigo technologies [23].

Central to the efforts of these new entrants and to the efforts of current leaders Google and Yahoo! to maintain their dominance are novel and more effective allocation mechanisms, optimizing both auctioneer revenue and efficient advertisement allocation. An important aspect of this will be the allocation of advertisements for novel keywords when little is known about the associated advertisers. Google currently addresses this problem by using their search relevance score as (essentially) a prior estimate of at least the relative ordering of click-through-rates. As the revenue-maximizing allocations require accurate estimates of click-through-rates, the process of quickly and effectively refining information will allow the sponsored search provider to more readily adapt to and provide the best allocation mechanisms for novel keywords. Being the best provider of sponsored search for a valuable new keyword can secure advertiser participation in the most important and dynamic part of a keywords relevant advertising life, when advertisers are both willing to pay the most and users are most likely to click on these advertisements.

Direct application potential aside, the unique and elegant auction mechanisms used in the sale of sponsored search provide an interesting case study on the adoption of market-based tools in a growing and dynamic industry. Alongside well known applications of mechanism design such as FCC radio spectrum auctions and Roth's medical residency matching process, sponsored search auctions represent a triumph of economically inspired design. Understanding the theoretic basis for the effectiveness of sponsored search auctions and the characteristics of the application environment can inform the development of market-based mechanisms for myriad other applications and industries.

Finally, the adaptation of a metareasoning process to this environment, which has a unique and complex informational structure is extremely interesting in its own right [20]. Combining lessons from value of information and decision-theoretic reasoning and well known problems such as the multi-armed bandit problem, we are able to use a wide array of research and results to inform our understanding of metareasoning in a novel domain [1, 12, 3, 2]. This work is a compelling study in applying metareasoning to a domain whose informational structure entails a number of complications in value of information calculations, the resolution of which informs the general applicability of metareasoning to environments with complicated and highly interrelated information structures.

1.3 Primary Contributions

The primary technical contribution of this work is the adaptation and application of a value of information based metareasoning process to rationally deliberate about how to adaptively set allocation mechanisms in sponsored search auctions. Using a game theoretic framework, we adapt the traditional principles of metareasoning, which generally rely on a very specific and often restrictive set of assumptions, to the peculiarities of the sponsored search setting. By carefully considering the way in which a sponsored search advertiser learns from a set of clicks and impressions, we adapt a framework which traditionally views exploration and exploitation as completely decoupled to a domain in which they are powerfully and inextricably interwound.

Empirically, we demonstrate the benefits in increased revenue achievable through value of information based metareasoning, and characterize the performance of the algorithm in comparison to heuristic benchmark algorithms. In addition, we begin to understand the effect of changing a number of of parameters on the performance of the algorithm. Finally, we discuss the application potential and some promising extensions to the realistic faithfulness and performance of our models and algorithms.

1.4 Outline

Chapter 2 provides a history of Internet advertising, and presents the game theoretic framework with which we analyze sponsored search auctions, and within which our algorithms will operate. Drawing primarily from work done by Varian, Edelman et al., and Lahaie and Pennock, it develops to requisite tools for analyzing a multi-stage sponsored search auction game [24, 6, 14]. Chapter 2 additionally presents the two heuristic algorithms, myopic revenue maximization and variance based exploration which will serve as benchmarks for the value of information based metareasoning algorithm. Chapter 4 presents the metareasoning principles of Russell and Wefald, and adapts them for the unique information landscape and action types of the sponsored search game [20]. It also presents our value of information calculations and the metareasoning algorithm itself. Chapter 5 presents performance results, comparing the revenue performance of the metareasoning algorithm to myopic revenue-maximization, and comparing its information gathering power to variance based greedy exploration. Chapter 5 also investigates the performance of the algorithms under a varying set of tunable parameters. Finally, chapter 6 concludes by discussing future work, possible extensions and the application potential of metareasoning to real world sponsored search auctions.

Chapter 2

Sponsored Search and Internet Advertising

Our understanding of sponsored search auction mechanisms begins with understanding the unique history and technological trajectory of Internet advertising. Drawing inspiration from successively more refined advertising systems, we present a game theoretic model based on the work of Varian, Edelman and Lahaie and Pennock for the sponsored search auction, and characterize the equilibrium bidding behavior of advertisers, which allows us to calculate expected revenue for a given allocation [24, 6, 14]. This ability to characterize equilibrium revenue properties will prove critical to the algorithms we develop in the next chapter for deliberating in the multi-stage sponsored search auction problem.

2.1 Online Advertising

By 2004, Internet advertising was a $9.6 billion industry, eclipsing such advertising stalwarts as outdoor billboard and banner advertising, and comprising some 80% of the magazine advertisement industry and over 50% of the radio advertisement industry. In addition, Internet research providers DoubleClick and Nielsen//Netratings have calculated that the growth rate of internet advertising is 31.5% annually, more than three times the growth rate of television advertising and almost five times the growth rate of U.S. GDP [4].

The history of online advertising is one of trial and error and rapid change. In little more than 10 years, Internet advertising has evolved from simple, contract-based advertising mechanisms that are simply the online analogue to traditional ad media to dynamic and unique allocation systems that allow online advertising to provide significantly more relevant, targeted and effective advertising than other methods. Edelman notes that the evolution of market mechanisms in Internet advertising is especially dynamic when compared to other notable market mechanisms such as FCC radio spectrum auctions and medical residency matching programs due to high competitive pressure, negligible or low barriers to entry, low cost of experimentation and dynamic and adaptable technology [6].

Current methods for online advertising are of particular interest not only for their economic and industrial impact, but also as a case study on the adoption and evolution of market mechanisms in a very real and valuable setting. Modern dynamic methods of sponsored search advertising are a testament to the elegance and power of auction mechanisms that allow for the efficient and relevant pricing and display of advertisements. While understanding these auctions is interesting solely from a mechanism design perspective, the rapid growth and increasing importance of online advertising necessitates such an understanding in order to effectively design and analyze the performance of the next generation of Internet advertising tools.

2.1.1 History

In order to understand the current state of Internet advertising and the industry landscape in which sponsored search ad auctions occupy a central role, and to better anticipate future trajectories and applications of online marketing, it is instructive to examine the history of Internet advertising.

The earliest and still most ubiquitous online advertising mechanism is banner ads, displayed in eye-catching locations on websites. Banner advertisements are the online analogue to traditional ad media such as fixed billboards. HotWired, WebConnect and NetScape pioneered most of these efforts in 1994, cooperating with large company advertisers such as AT&T, MCI, Sprint and Volvo to display banner ads on their web portal [21]. These advertisements were typically negotiated in large, static contracts, and advertisers were charged each time their banner was shown (per impression). The high barriers to entry associated with contract negotiation and partnerships with large companies meant that new entrants were slow, even with negligible technological barriers. Banner ads began to evolve in smarter ways with companies such as Focalink Communications introducing basic context-sensitive banner advertising, in order to begin to attempt to focus the torrent of online advertising toward prospective customers [21].

The banner advertising model was the only and remains the dominant advertising model today, although the simple image or text ads of 1994 have been largely replaced with Flash and large or moving image banners. In 2007, according to Nielsen//NetRatings, banner advertising in its various incarnations commanded over 80% of the Internet advertising market [16]. In high growth advertising sectors such as consumer and retail goods, however, sponsored search has begun to overtake banners as the dominant online advertising tool. The effectiveness of consumer goods advertisements in particular is greatly enhanced when placed in a relevant context, which sponsored search provides more directly and effectively than other methods.

The first incarnation of the context sensitive sponsored search type of online advertising was in 1997 by GoTo, later becoming Overture and acquired by Yahoo. Overture introduced several novel ideas that would become critically important to modern Internet advertising.

- *Context Sensitivity* – By associating advertisements with keywords for the first time, Overture offered advertisers the chance to be exposed to users at least tangentially interested in their product, increasing the effectiveness of their advertisements by properly situating them in a potentially valuable environment.

- *Pay Per Click* – Rather than charging advertisers per fixed number of showings or impressions (1000 was the typical lot number), Overture began to charge advertisers per click. This allowed advertisers to pay only when they had accrued some benefit from the advertisement, and was prized by advertisers as a way to maximize the effectiveness of marketing.

- *Auction Mechanism* – Overture utilized a "generalized" first price auction for the sale of their advertising slots. The highest bidding advertiser would be placed in the most prominent and desirable position with the other advertisers arrayed by decreasing bid in the balance of the slots. Each time an advertisers ad was clicked on, they would be charged their most recent bid.

2.1.2 Keyword advertising – Yahoo's Overture and Google's AdWords

The "generalized" first price auction mechanism used by Overture bears closer examination, as its weaknesses informed the development of the auction mechanisms used today. The first price auction is highly unstable, due to ability and utility of rational advertisers to rapidly and repeatedly change their bids. The following example illustrates this.

Example 1. *Consider an auction with two slots, the top slot recieving twice the number of clicks of the bottom, and three advertisers with advertiser A having value $1 per click, B, $2 and C, $3. If, in order ensure that he obtains a position, advertiser B bids $1.01, ensuring that A will never enter, then C's best response is to bid $1.02. But then it is in B's best* *interest to bid \$1.03. This cycle continues until C bids \$2, at which point B's best response is to return to bidding \$1.01, and this process repeats.*

This instability led many advertisers to develop automated bidding agents which would quickly and repeatedly refine bids in order to quickly change advertisement placements. Not only is this socially inefficient, as agents waste computational power and effort in designing and deploying bidding robots, but as Edelman demonstrates, in the case where agents have unequal computational power or response lag time, the process of repeated bid changes results in decreased revenue for the sponsored search provider when agents cannot properly or quickly respond to inefficient bids. In addition, McAdams argues that the costs incurred by advertisers in developing these bidding agents are directly reflected in decreased bids, resulting further in decreased revenue for the provider [15].

The undesirable volatility of prices and inefficiency of allocations in the generalized first price auction led Google to introduce a different mechanism with its advertising system AdWords. AdWords implements a second price auction over multiple positions, which Edelman calls a "generalized second price" (GSP) auction, so that each advertiser, when his ad receives a click, pays the bid of the advertiser appearing directly below him. This generalizes the standard second price auction in the sense that a GSP auction for one slot is precisely a second price auction. This mechanism is informed by the observation that in a generalized first price auction, if an advertiser in slot s does not want to switch slots, then he would like to minimize his payment while maintaining his position – thus his optimal bid would be whatever the advertiser in slot $s + 1$ bids, plus a marginal amount. GSP ensures that no matter the actual bid, the payment of an advertiser is this optimal amount.

By setting advertiser payments equal to the bid of the advertiser appearing one slot lower, the mechanism decouples an advertisers own bid, which determines his placement, and his payment, which will always be the allocation preserving minimum bid of the generalized first price auction. This results in more stable prices, as an advertiser unilaterally changing his bid while maintaining his current slot does not affect his payment.

Example 2. *Continuing the example from above, we see that if all agents bid truthfully, advertiser A appears in slot 1 paying \$2, the bid of advertiser B, who appears in slot 2, paying \$1, the bid of advertiser C. Advertiser C cannot benefit from changing her bid, because any bid lower than \$1 will result in her continued exclusion, and any bid above \$1 will result in negative profits. Advertiser B cannot benefit from changing his bid, because any bid lower than \$1 will result in his exclusion from the auction and 0 profits, and any bid above \$2 will result in negative profits, and any bid in the range \$1 - \$2 will not change his allocation. Advertiser A cannot benefit from changing her bid since any bid above \$2* *will not change her position, and any bid below \$2 but above \$1 will result in her moving down one slot and recieving half as many clicks, while making less than double the profits per click. Thus, bidding true valuations is an equilibrium of the generalized second price auction in this example. Note that this is not generally the case, but the structure of click differences in the slots has resulted in an equilibrium. It is clear, however, that the second price mechanism results in more stable prices due to the fact that an interval of bids will result in the same slot allocation.*

Edelman notes that GSP is structurally similar to a Vickrey auction, but the characteristics and properties of the auction are very different. Importantly, unlike the closed-bid second price auction, truthful revelation is not generally an equilibrium, nor is there usually a dominant strategy solution [6]. Yahoo! also adopted the generalized second price auction mechanism, and both Yahoo! and Google use some form of GSP in the current incarnations of their sponsored search advertising.

2.2 Position Auction Game Theory

A more formal game theoretic examination of the properties of the generalized second price auction mechanism for sponsored search sharpens our understanding of the equilibrium properties of the auction, and allows us to characterize expected equilibrium revenue. Varian and Edelman, Ostrovsky and Schwarz independently analyze the game theoretic properties of these algorithms and arrive at similar results, although with different goals in mind [24, 6]. We primarily follow the exposition of Varian in this work, as it has an appealing ground up construction and derivation of equilibrium advertiser behavior. We additionally follow the work of Lahaie and Pennock, who extend the work of Varian to allow for arbitrary weighting of agent bids [14].

We formally model the sponsored search auction as a n-player game whereby k advertising slots are allocated to n advertisers. Advertisers with quasi-linear utility want to maximize their revenue, which is a product of the clicks they receive and the per click profit. Say advertiser s appears in slot s, renumbering if necessary. Then his payoff for obtaining these clicks at price p per click is

$$u_s(p) = z_s(v_s - p) \tag{2.1}$$

We assume that the observed click-through-rate $z_s = e_s x_s$ is a separable combination of advertiser effect e_s, which captures how likely a user who views an advertisement will click it, and x_s, the position effect measuring how desirable the slot that the advertisement appears in is. Note that when we refer to an advertiser's click-through-rate, we mean specifically the advertiser e_s, which is independent of slot or allocation, rather than the observed click-through-rate.

We are working towards the characterization of equilibrium advertiser behavior and revenue. We begin by examining the two most intuitive and currently most commonly implemented auction mechanisms, which will inform the understanding of the more general squashing factor family of auction allocation rules studied by Lahaie and Pennock [14]. Following Varian and Edelman et al., we then explore the Nash equilibrium conditions, and then focus on a subset of these equilibria, the symmetric or locally envy-free equilibria, which allow us to derive bounds on equilibrium bids and revenue [24, 6]. Understanding advertiser equilibrium behavior allows us to focus on the decision-making of the sponsored search provider and to take the behavior of advertisers as fixed and exogenously given.

2.2.1 Rank-by-Bid and Rank-by-Revenue

The GSP mechanism presented in section 2.1.2 is typically known as a *rank-by-bid* mechanism, as advertisers are allocated positions based on decreasing order of bids. The auction system used by Google has a further wrinkle which distinguishes it from the standard GSP setting. Rather than simply ordering advertisers by bid, Google adjusts their bid by weighting by a *relevance score*, which captures both the quality of the advertisement and the likelihood that it will be clicked on. In the case where this score is exactly the probability that the advertisement will be clicked on, conditional on it appearing in a given slot (so that the weight of all advertisements is calibrated to a common metric), this mechanism orders advertisers in decreasing order of expected revenue to the sponsored search provider.

Examining these two early forms of actual sponsored search position auctions and their game theoretic properties with a more formal treatment will inform the understanding of general sponsored search auctions.

In the Yahoo/Overture generalized second price auction, advertisers submit bids $b_1, \cdots, b_n$, and are allocated slots in decreasing bid order, and pay the bid of the advertiser that appears one slot lower. As an example, with $n = 5$ bidders, $k = 4$ slots, and $b_1 > b_2 > \cdots > b_5$, we have the following rank-by-bid allocation

Position	Value	Bid	Price	Position Effect	Advertiser Payoff
1	v_1	b_1	b_2	x_1	$e_1 x_1 (v_1 - b_2)$
2	v_2	b_2	b_3	x_1	$e_2 x_2 (v_2 - b_3)$
3	v_3	b_3	b_4	x_1	$e_3 x_3 (v_3 - b_4)$
4	v_4	b_4	b_5	x_1	$e_4 x_4 (v_4 - b_5)$
5	v_5	b_5	0	0	0

This allocation scheme is optimal for the auctioneer when value and advertiser effect are highly correlated. For example, in searches for products with high brand awareness, large brands are likely to both be willing to pay more and have higher probability of being clicked on. If, however, V and E are uncorrelated or negatively correlated, this scheme performs poorly, as advertisers with high value are placed higher, but are clicked on less often, resulting in lower revenue.

In earlier, more transparent versions of Google's AdWords, bids are ranked by score $e_s v_s$ or the relative ranks of expected revenue [6]. The *rank-by-revenue* allocation scheme attempts to address some of the shortcomings of the *rank-by-bid* scheme, and allocates to maximize expected revenue. In this scheme, advertisers pay the least price necessary for their rank to remain unchanged. That is, advertiser s pays according to

$$e_s b_s = e_{s+1} b_{s+1} \Rightarrow p_s = \frac{e_{s+1}}{e_s} b_{s+1} \tag{2.2}$$

Current implementations of AdWords remain true to the spirit of the rank-by-revenue mechanism, although rather than maintaining a transparent estimate of click-through-rates, AdWords now uses the relevancy rankings and algorithms of Google Search to proxy the estimated click-through-rate.

2.2.2 Equilibria of the Sponsored Search Auction

Rank-by-bid and rank-by-revenue are two special cases of the more general allocation mechanism which we explore here. Lahaie and Pennock noted that these two allocation schemes can be generalized by introducing an advertiser-specific weight w_s on bids [14]. Following their work, we primarily consider weights that are "squashed" click-through-rate estimates, that is, of the form e_s^γ so that advertisers are ranked by $e_s^\gamma v_s$. Rank-by-bid has $\gamma = 0$, where click-through-rates are maximally squashed so that they have no effect, and rank-by-revenue has $\gamma = 1$, where click-through-rates have maximal effect on determining the allocation. Advertiser s again pays the bid necessary to remain in position s, that is,

$$p_s = \left(\frac{e_{s+1}}{e_s}\right)^{\gamma} b_{s+1} \tag{2.3}$$

These prices reduce in the rank-by-bid case with $\gamma = 0$ to $p_s = b_{s+1}$ and in the rank-by-revenue case with $\gamma = 1$ to $p_s = \left(\frac{e_{s+1}}{e_s}\right) b_{s+1}$, consistent with our earlier definitions.

The complete information equilibria of the sponsored search auction can be understood by examining the equilibria of the corresponding simultaneous move game of complete information[1]. Advertiser s sets a bid to maximize utility $u_s = (v_s - p_s)x_s e_s$. Note that, in comparing the utility of appearing in slot s against the utility of appearing in slot t, the advertiser effect e_s appears in both utility functions, and can thus be cancelled out.

In a Nash equilibrium, no advertiser wishes to switch out of his current position either by trading up or trading down. Since an advertisers payment and utility changes only if he alters his bid significantly enough to change the allocation, swapping slots is the only way in which an advertiser can potentially deviate in this game. Suppose we have the allocation presented above with $\gamma = 0$ so we are using the rank-by-bid allocation scheme. If the advertiser in position 3 decides to move up a slot, he sets $b_3 > b_2$, and then pays b_2 per click. If he decides to move down a slot, he sets $b_3 < b_4$ and pays b_5 per click. Note that when swapping up to slot t, bids are determined according to current b_t, but when swapping down to slot t, bids are determined according to current b_{t-1}. This results in the following price structure and equilibrium from Varian [24]:

$$p_{st} = \begin{cases} \left(\frac{e_{t+1}}{e_s}\right)^{\gamma} b_{t+1} & \text{for } s \leq t \\ \left(\frac{e_t}{e_s}\right)^{\gamma} b_t & \text{for } s > t \end{cases} \tag{2.4}$$

When $t = s$ (i.e. in the equilibrium allocation where all advertisers occupy their respectively numbered slot) we will drop the second subscript and refer to the equilibrium price p_{ss} simply as p_s.

Definition 2.2.1. A Nash equilibrium in the sponsored search allocation game is a set of bids $B = \{b_1, \cdots, b_n\}$ and associated prices $P = \{p_1, \cdots, p_n\}$ that satisfies, for all agents s

$$\begin{aligned} (v_s - p_s)x_s &\geq (v_s - p_{st})x_t \text{ for slots } t > s & (2.5) \\ (v_s - p_s)x_s &\geq (v_s - p_{s(t-1)})x_t \text{ for slots } t < s & (2.6) \end{aligned}$$

[1] Although Varian argues that the complete information assumption is not unreasonable in this setting due to the transparency of bidding and auction mechanisms, the analysis here extends naturally to an incomplete information case with Bayesian Nash equilibria [24]

Clearly, there are a large range of bids that are Nash equilibria, since marginal bid changes do not directly change allocations or personal utility. In the example above, any bid $b_3 \in (b_4, b_2)$ does not change the allocation or payments of advertiser 3, and is thus part of some supporting set of Nash equilibrium bids. In general, the set of supporting bids can be found through a linear program.

2.2.3 Symmetric Nash or locally envy-free equilibria

Varian introduces the concept of the symmetric Nash equilibrium in order to simplify the set of supporting bids, and to make more tractable the analysis of the position auction [24]. Symmetric Nash equilibria admit a concise characterization of equilibrium advertiser bids, and thus allow us to calculate equilibrium expected revenue.

Definition 2.2.2. A symmetric Nash equilibrium is a set of bids $B = \{b_1, \cdots, b_n\}$ and associated prices $P = \{p_1, \cdots, p_n\}$ such that

$$(v_s - p_s)x_s \geq (v_s - p_{st})x_t \text{ for all s and t} \tag{2.7}$$

Equivalently,

$$v_s(x_s - x_t) \geq p_s x_s - p_{st} x_t \text{ for all s and t} \tag{2.8}$$

Since only the inequalties for $t < s$ change from the definition of Nash equilibrium to symmetric Nash equilibrium, we see that the primary difference is in the properties of the bid made by an agent who is pushed out of his slot by another agent trading up.

Following Varian, we first verify that the set of symmetric Nash equilibria are indeed a subset of the Nash equilibria. This confirms the suitability of the symmetric equilibrium solution concept as the basis for a plausible set of advertiser behaviors.

Proposition 3. *If a set of bids B and associated prices P satisfies the symmetric Nash equilibrium conditions, then it satisfies the Nash equilibrium conditions.*

Proof. Consider advertiser s and $t = k + 1$ in the SNE conditions. Then

$$(v_s - p_s)x_s \geq (v_{k+1} - p_{s(k+1)})x_{k+1}$$

but since $k + 1$ is an excluded bidder, $x_{k+1} = 0$, thus $v_s \geq p_s$. Then, we use 2.8, the alternative form of the SNE inequality to obtain

$$\begin{aligned} v_s(x_s - x_{s-1}) &\geq p_s x_s - p_{s-1} x_{s_1} \\ p_{s-1} x_{s-1} &\geq p_s x_s + v_s(x_{s-1} - x_s) \\ &\geq p_s x_s + p_s(x_{s-1} - x_s) \\ &= p_s x_{s-1} \end{aligned}$$

This demonstrates price monotonicity in a symmetric Nash equilibrium, $p_{s-1} > p_s$ for all slots s. Then, it is easy to see that the SNE inequality

$$(v_s - p_s)x_s \geq (v_s - p_{st})$$

implies

$$(v_s - p_s)x_s \geq (v_s - p_{s(t-1)})$$

□

The particularly relevant property of symmetric equilibria that makes for more tractable analysis of the position auction is the fact that one needs only check one position up and one position down from s to verify that the inequalities hold for all s. This allows us to solve for bounds on equilibrium bids with a simple recursive solution, rather than the linear programming solution necessary for general Nash equilibria. The following propositions extend Varian's proof to allow for weighted bids, as described in Lahaie and Pennock [24, 14].

Lemma 4. *In a symmetric Nash equilibrium, agents are ordered according to the weighted scoring mechanism, that is, in order of decreasing $e_s^\gamma v_s$.*

Proof. Consider advertisers and slots s and t. In equilibrium, each prefers her own slot to the other

$$\begin{aligned} v_s(x_s - x_t) &\geq p_s x_s - p_{st} x_t \\ v_t(x_t - x_s) &\geq p_t x_t - p_{ts} x_s \end{aligned}$$

we can rewrite the inequality for s as follows

$$\begin{aligned} v_s(x_s - x_t) &\geq \left(\frac{e_{s+1}}{e_s}\right)^\gamma b_{s+1} x_s - \left(\frac{e_{t+1}}{e_s}\right)^\gamma b_{t+1} x_t \\ e_s^\gamma v_s(x_s - x_t) &\geq e_{s+1}^\gamma b_{s+1} x_s - e_{t+1}^\gamma b_{t+1} x_t \end{aligned}$$

Similarly, the inequality for t can be rewritten as

$$e_t^\gamma v_t(x_t - x_s) \geq e_{t+1}^\gamma b_{t+1} x_t - e_{s+1}^\gamma b_{s+1} x_s$$

Adding these two inequalities gives

$$(e_s^\gamma v_s - e_t^\gamma v_t)(x_s - x_t) \geq 0$$

showing that x_s and x_t are ordered the same way as $e_s^\gamma v_s$ and $e_t^\gamma v_t$. □

Proposition 5. *If bid set $B = \{b_1, \cdots, b_n\}$ satisfies the symmetric Nash equilibrium inequalties for slots $s-1$ and s and s and $s+1$, then it satisfies them for all $t \neq s$. Without loss of generality, let $s = 2$ so that $s-1 = 1$ and $s+1 = 3$.*

Proof. The SNE inequalities are

$$v_1(x_1 - x_2) \geq p_1 x_1 - p_{12} x_2$$

$$v_2(x_2 - x_3) \geq p_2 x_2 - p_{23} x_3$$

By proposition 4, $e_1^\gamma v_1 > e_2^\gamma v_2 \Rightarrow \left(\frac{e_1}{e_2}\right)^\gamma v_1 > v_2$. Then we can rewrite the second inequality as

$$\begin{aligned} \left(\frac{e_1}{e_2}\right)^\gamma v_1(x_2 - x_3) &\geq p_2 x_2 - p_{31} x_3 \\ v_1(x_2 - x_3) &\geq \left(\frac{e_2}{e_1}\right)^\gamma \left(\left(\frac{e_3}{e_2}\right)^\gamma b_3 x_2 - \left(\frac{e_4}{e_2}\right)^\gamma b_4 x_3\right) \\ v_1(x_2 - x_3) &\geq \left(\frac{e_3}{e_1}\right)^\gamma b_3 x_2 - \left(\frac{e_4}{e_1}\right)^\gamma b_4 x_3 \end{aligned}$$

Rewriting the first inequality as

$$v_1(x_1 - x_2) \geq \left(\frac{e_2}{e_1}\right)^\gamma b_2 x_1 - \left(\frac{e_3}{e_1}\right)^\gamma b_3 x_2$$

and adding the two inequalities, we obtain

$$v_1(x_1 - x_3) \geq \left(\frac{e_2}{e_1}\right)^\gamma b_2 x_1 - \left(\frac{e_4}{e_1}\right)^\gamma b_4 x_3$$

or

$$v_1(x_1 - x_3) \geq p_1 x_1 - p_{13} x_3$$

demonstrating that the SNE inequality for swapping down holds for 1 and 3. The proof for swapping up from 3 to 1 is similar. This demonstrates the transitivity of the SNE inequality conditions, and shows that verifying $s-1$ and $s+1$ is sufficient for demonstrating that the inequalities hold for all s. □

2.2.4 Equilibrium bids and prices

We can now explicitly solve for equilibrium prices and bids by considering the symmetric decision of s and $s-1$. Since s does not not want to swap up to slot $s-1$ in a SNE,

$$\begin{aligned}
x_s(v_s - \left(\frac{e_{s+1}}{e_s}\right)^{\gamma} b_{s+1}) &\geq x_{s-1}(v_s - \left(\frac{e_s}{e_s}\right)^{\gamma} b_s) \\
x_s(v_s e_s^{\gamma} - e_{s+1}^{\gamma} b_{s+1}) &\geq x_{s-1}(v_s e_s^{\gamma} - e_s^{\gamma} b_s) \\
x_{s-1} e_s^{\gamma} b_s &\geq e_s^{\gamma} v_s (x_{s-1} - x_s) + x_s e_{s+1}^{\gamma} b_{s+1}
\end{aligned}$$

Since $s-1$ also does not want to swap down to slot s,

$$\begin{aligned}
x_{s-1}(e_{s-1}^{\gamma} v_{s-1} - e_s^{\gamma} b_s) &\geq x_s(v_{s-1} e_{s-1}^{\gamma} - e_{s+1}^{\gamma} b_{s+1} \\
x_{s-1} e_s^{\gamma} b_s &\leq e_{s-1}^{\gamma} v_{s-1}(x_{s-1} - x_s) + x_s e_{s+1}^{\gamma} b_{s+1}
\end{aligned}$$

We can combine these inequalities to bound the possible values for the bid of advertiser s, b_s in a symmetric Nash equilibrium.

$$e_s^{\gamma} v_s (x_{s-1} - x_s) + x_s e_{s+1}^{\gamma} b_{s+1} \leq x_{s-1} e_s^{\gamma} b_s \leq e_{s-1}^{\gamma} v_{s-1}(x_{s-1} - x_s) + x_s e_{s+1}^{\gamma} b_{s+1} \tag{2.9}$$

We can find a symmetric Nash equilibrium set of bids by recursively finding a sequence that satisfies the inequalities 2.9.

2.2.5 Lower Recursive Solutions

Of particular interest are the boundary conditions of advertiser bids, of which the lower bound is most directly relevant. We can solve for a closed form solution, bids b^L for the lower recursive solutions

$$b_s^L x_{s-1} e_s^{\gamma} = e_s^{\gamma} v_s (x_{s-1} - x_s) + x_s e_{s+1}^{\gamma} b_{s+1} \tag{2.10}$$

by noting the base case of the recursion occurs at the first excluded bidder. Since there are k slots, $x_s = 0$ for $s > k$. The lower bound for $s = k+1$ gives us

$$\begin{aligned}
b_{k+1}^L x_k e_{k+1}^{\gamma} &= e_{k+1}^{\gamma} v_{k+1}(x_k - x_{k+1}) + x_{k+1} e_{k+2}^{\gamma} b_{k+2} \\
&= e_{k+1}^{\gamma} v_{k+1} x_k \\
\Rightarrow b_{k+1}^L &= v_{k+1}
\end{aligned}$$

The reasoning for this is the same as in the standard Vickrey auction. For the first excluded bidder, it is optimal to bid her value because bidding lower will not change her payments, and if, for some reason, a higher bidder leaves, she will make a profit. We then obtain:

$$b_s^L x_{s-1} e_s^{\gamma} = \sum_{t \geq s} v_t e_t^{\gamma} (x_{t-1} - x_t) \tag{2.11}$$

Varian notes that the lower recursive solution is particularly compelling because it captures the following simple yet intuitive behavioral strategy [24]: If I am in position s what is the optimal bid to set so that if I happen to exceed the bid above me b_{s-1} by a marginal amount I will make at least as much profit as I do now? We find that solving for this optimal bid is precisely the lower recusion:

$$\begin{aligned} (v_s e_s^{\gamma} - e_{s+1}^{\gamma} b_{s+1}) x_s &\geq (v_s e_s - e_{s+1} b_s^*) x_{s-1} \\ b_s^* x_{s-1} e_s &= e_s^{\gamma} v_s (x_{s-1} - x_s) + x_s e_{s+1}^{\gamma} b_{s+1} \end{aligned}$$

2.2.6 Symmetric Nash equilibrium Revenue

Using the lower bounds on bids derived above, we can calculate total advertiser revenue. Recalling that agent s pays $\left(\frac{e_{s+1}}{e_s}\right)^{\gamma}$, we first multiply equation 2.11 by e_s/e_s^{γ} to obtain $x_{s-1} e_s b_s^L$, estimated revenue for each agent. We then sum over all slots $s = 1 \cdots k$ to obtain total equilibrium revenue:

$$R^L = \sum_{s=1}^{k} \sum_{t=s}^{k} \left(\frac{e_t}{e_s}\right)^{\gamma} v_t e_s (x_{t-1} - x_t) \tag{2.12}$$

The ability to estimate total equilibrium revenue will prove crucial in our analysis of meta-deliberation and learning in sponsored search.

Chapter 3

Repeated Position Auctions and Heuristic Algorithms

Armed with a characterization of equilibrium advertiser behavior and revenue, we can now extend the sponsored search auction game to a multi-period model. This will allow our algorithms to make inter-temporal decisions, accepting lower current revenue in favor of better information and higher future revenue. We first present the details of the multi-period model, and then present two heuristic algorithms which attempt to naively solve, respectively and separately, the exploration and exploitation problems.

3.1 Repeated sponsored search auctions

The sponsored search auction is not a one-shot auction. Generally, a given allocation will persist for some fixed number of hours, days, weeks, etc., and then the auction will be repeated, and a new allocation will be implemented. We thus extend the sponsored seach auction to a multi-period game. This extension will allow for meaningful intertemporal decision-making, and is an environment where metareasoning and the value of information become important.

3.1.1 Multi-period sponsored search auctions

The multi-period sponsored search auction model proceeds as follows: In each time period, the auctioneer chooses a squashing factor γ, which, in combination with CTR estimates $\hat{E}$ and advertisers' equilibrium bidding behavior, gives an allocation. We assume throughout that the advertisers bid according to the lower recursive solution of the symmetric Nash equilibrium presented above. Advertisements receive impressions according to this allocation, with the top advertisement receiving some set amount of effective impressions, and advertisements in lower slots receiving a discounted amount of effective impressions. The auctioneer receives both revenue and information from the auction, with each click providing both revenue and a sample for CTR estimates. The following outline summarizes the multi-period model of the sponsored search auction.

1. Using current estimates of click-through-rates $\hat{E}_t$, based on number of clicks and effective impressions observed in time periods $1 \cdots t-1$, the auctioneer chooses a squashing factor γ
2. Advertisers bid according to the equilibrium behavior specified in the lower recursive solution of the symmetric Nash equilibrium
3. From advertisers' weighted bids and click-through-rate estimates, the auctioneer sets an allocation
4. Advertisers receive impressions based on the allocation, and auctioneer receives revenue and uses the actual numbers of clicks accrued by each ad to refine next-period CTR estimate $\hat{E}_{t+1}$

The relationship and potential trade-off between information and revenue becomes clear here. The advertisement that is placed in the top slot will receive the largest number of effective impressions, offering the most information and allowing the auctioneer to learn most effectively about the true click-through-rate. However, this is not necessarily, and indeed, is unlikely to be the revenue maximizing allocation.

3.1.2 Click-through-rate Estimation

Ex ante, the sponsored search provider has no information about the true click-through-rates of each advertiser. Thus, the auctioneer must estimate CTRs according to the number of clicks and impressions he observes for each advertisement. The provider starts off with very naive symmetric prior estimates for the true click-through-rates E of the advertisers and maintains estimates $\hat{E}$ of the advertisers' click-through-rates by tallying the total number of clicks received and the total number of effective impressions.

Impressions in a sponsored search auction are markedly different from impressions in banner advertising, as all showings are not equal in the sponsored search setting. Effective impressions attempts to capture in the count of impressions the effect of appearing in a more or less desirable position. If the relevant keyword was searched for 100 times, then all advertisements appearing in the sponsored search auction have 100 technical impressions. However, only the advertisement in the top slot will receive the full amount of effective impressions - the advertisements appearing in lower slots will receive only some fraction of the number of technical applications. One can think of this as only a certain percentage of users will even look at ads appearing in lower positions, but all will at least look at the top ad. We thus factor in the effect of the position an advertisement appears in into the estimates of that advertisements click-through-rate. This allows CTR estimates to purely reflect the advertiser effect, and to be independent of the positions in which that advertisement appeared in the past. We thus maintain counts c^t of clicks and i^t of effective impressions as the provider's estimates of advertiser click-through-rates.

3.2 Myopic revenue maximization

The first of the heuristic algorithms we consider is a greedy revenue maximizing algorithm that calculates expected revenue with respect to current CTR estimates for a number of alternative squashing factors and chooses the squashing factor and associated allocation with the highest expected revenue. The utility of the squashing factor family of mechanisms becomes clear when we consider that, although rank-by-bid and rank-by-revenue are perhaps the most intuitive and simple of the ranking mechanisms, they do not necessarily maximize revenue. Although rank-by-revenue results in more efficient allocations, in that advertisers are ordered according to their expected contribution to total revenue, weighting bids, and thus payments by click-through-rate also decreases the amount of revenue obtained by the auction. Conversely, while rank-by-bid maximizes payments, it often results in suboptimal allocations. There is often some intermediate squashing factor which maximizes expected revenue that balances allocative efficiency with maximizing revenue [14].

We have already noted above that with highly correlated value and click-through-rate, rank-by-bid is optimal, as weighting by click-through-rate in any form is redundant, and serves only to decrease the amount of revenue. Similarly, in highly negatively correlated cases, rank-by-revenue becomes optimal, as it becomes difficult to meaningfully rank advertisements otherwise. However, for non-correlated or incompletely correlated cases, there is often an intermediate squashing factor that is optimal. Lahaie and Pennock note that by varying the correlation of advertiser value and click-through-rates, we can observe this phenomenon in effect [14]. Figure 3.1 replicates their results, and shows normalized expected revenue vs. squashing factor γ for differing Spearman correlations (-1, 0, 1) between value

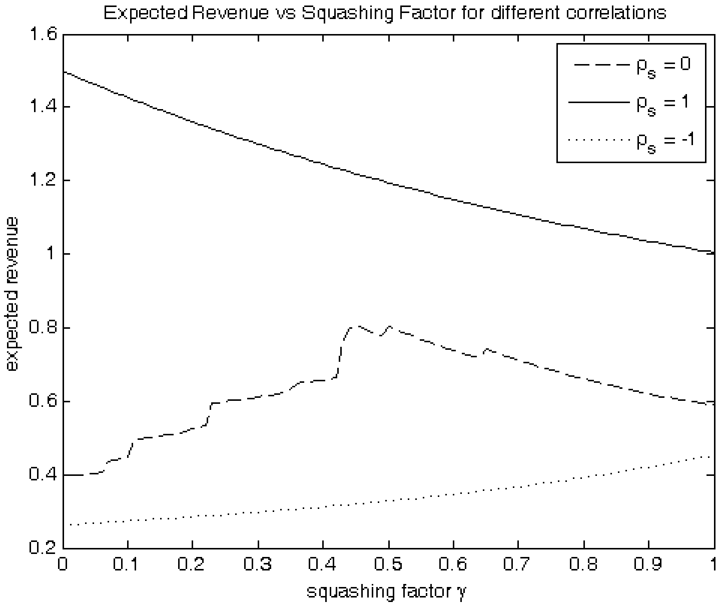

Figure 3.1: Expected Revenue vs. Squashing Factor for differing Spearman correlations ρ_s

and click-through-rate.

Greedy revenue maximization is a process of choosing the optimal squashing factor γ, conditional on current click-through-rate estimates and advertiser value. The myopic algorithm that we present here chooses the squashing factor that maximizes expected revenue, conditional on current CTR estimates. This is a multi-step process:

1. Identify potential squashing factors, a candidate set of different allocations.
2. Simulate expected revenue for each of these squashing factors with uncertain click-through-rate estimates. Since our click-through-rate estimates are uncertain, noisy estimates, we maintain a model of knowledge uncertainty, and use a Monte-Carlo method to simulate the uncertainty.
3. Choose squashing factor with highest average simulated expected revenue. This is our myopic expected revenue maximizing squashing factor and allocation.

3.2.1 Identifying potential squashing factors

We can create a set of potential squashing factors Γ by calculating the pairwise swap squashing factors at which two advertisers will switch positions, and then identifying which of these are associated with a allocative shift in the full game.

Consider two advertisers with private values v_1, v_2 and CTR e_1, e_2. The squashing factor γ at which these two advertisers switch positions satisfies

$$e_1^{\gamma} v_1 = e_2^{\gamma} v_2 \Rightarrow \gamma log(e_1) + log(v_1) = \gamma log(e_2) + log(v_2) \Rightarrow \gamma = \frac{log(v_2/v_1)}{log(e_1/e_2)}$$

Thus, for any two advertisers s and t, we can calculate their *swap gamma* $\gamma_{st}^{swap} = \frac{log(v_t/v_s)}{log(e_s/e_t)}$. For advertisers $1...n$, the matrix of swap gammas

$$\Gamma^{swap} = \begin{bmatrix} 0 & \gamma_{12}^{swap} & \cdots & \gamma_{1n}^{swap} \\ \gamma_{21}^{swap} & 0 & \cdots & \gamma_{2n}^{swap} \\ \vdots & & & \\ \gamma_{n1}^{swap} & \gamma_{n2}^{swap} & \cdots & 0 \end{bmatrix}$$

gives us all possible gammas where a shift in allocations can occur. We can then constrain the entries in Γ^{swap} to be within $[\gamma_{min}, \gamma_{max}]$. Most frequently, $\gamma_{min} = 0$ and $\gamma_{max} = 1$, since this gives the logical border cases of rank-by-bid ($\gamma = 0$) and rank-by-revenue ($\gamma = 1$).

We then determine the allocations of the position auction using each of these squashing factors and our current CTR estimates. This associates each squashing factor γ_{ij}^{swap} with an allocation A_{ij}. Most of these allocations are not unique. Suppose without loss of generality for illustrative purposes that there are allocations $A_1, \cdots, A_n$ with γ_{ij}^{swap} resulting in allocation A_i for all j. We select the minimum and maximum squashing factors that result in each allocation as our candidate set of squashing factors Γ. That is, (with swap superscript suppressed):

$$\Gamma = \{\gamma_{1a}, \gamma_{1b}, \gamma_{2a}, \gamma_{2b}, \cdots, \gamma_{na}, \gamma_{nb}\} \tag{3.1}$$

where $\gamma_{ia} = min\{\gamma_{i1}, \cdots \gamma_{in}\}$ and $\gamma_{ib} = max\{\gamma_{i1}, \cdots \gamma_{in}\}$ for all i.

3.2.2 Revenue Simulation with Uncertain CTRs

With a set of of alternative squashing factors Γ, we can begin to reason about the expected revenue of setting these squashing factors and associated allocations. Since our estimates of CTRs are noisy and uncertain, our estimates of expected revenue will be correspondingly uncertain. We ensure that we properly reason about expected revenue by explicitly modeling the uncertainty of our knowledge, and by using simulation to evaluate potential uncertain outcomes.

We model the uncertainty of our knowledge of the true click-through-rate as a normal distribution. It is important to note that this does not mean we believe the true CTR is distributed normally. Rather, it means that we view the knowledge we have obtained provides a normally distributed uncertain estimate of the true CTR. This characterization of uncertainty is consistent with work done on refining uncertain estimates through sampling from various distributions [27]. The CTR estimate for each advertiser s is composed of a number of impressions i_s and a number of clicks c_s. Our knowledge of the true CTR is thus modeled as a random variable distributed $N\ (\hat{\mu_s}, SE_s)$ where mean $\hat{\mu_s} = \frac{c_s}{i_s}$ and standard error

$$SE_s = \sqrt{\frac{c_s(1-\hat{\mu_s})^2 + (i_s - c_s)\hat{\mu_s}^2}{(i_s - 1)i_s}}$$

This is the standard definition of standard error $(\hat{\sigma}/\sqrt{n})$ of a sample drawn from a population with estimated standard deviation $\hat{\sigma}$ where the input vector is binary. We thus refine estimates of click-through-rates by receiving more impressions and more clicks, increasing the sample size and reducing the standard error.

We use a Monte-Carlo simulation to evaluate the possible revenue from our estimated CTRs. For a large number of steps, we draw for each advertiser s a simulated CTR e_s^{sim} from the $N(\hat{\mu_s}, SE_s)$ distribution, and calculate the expected revenue for each set of click-through-rates with respect to this set of CTR estimates. The mean over the repeated simulations of the revenue for each squashing factor gives us simulated expected revenue with uncertain CTRs.

The myopic algorithm then simply chooses the squashing factor γ^0 with the highest simulated expected revenue, and executes the corresponding allocation.

This process for simulating the uncertainty allows for some degree of exploration and dynamism even in the myopic algorithm. When the estimate for a specific click-through-rate is highly noisy, that is, it has a high standard error, there is a large range of possible values that it can take on in the simulation, and it can affect expected revenue in significant ways. This allows the algorithm to reason about what it does not know by repeatedly drawing a possible information state from the spectrum of uncertain information states it can be in. Advertisements for which we know very little, for which we have received a comparatively small number of effective impressions and thus high standard error are likely to significantly affect this process. Since negative click-through-rates are meaningless and zero click-through-rates are highly unlikely, simulation click-through-rates are constrained to be positive. This amounts to a upward bias in simulated click-through-rates, particularly for advertisements with noisy estimates. This correspondingly results in an upward bias in their simulated score and a higher ranking, resulting in the algorithm learning more about advertisements about which it knows very little. Since this amounts to a net upward shift in click-through-rates, this is also generally associated with increased simulated expected revenue, so that even in a myopically exploiting case, there is some impetus toward exploration and refining estimates of particularly uncertain or noisy CTRs.

By employing a formal model for uncertainty, the role of sampling to explore and reduce uncertainty becomes clear. Each effective impression an advertisement receives increases the sample size of its CTR estimates and hence both refines its estimate in reality, by bringing the mean CTR estimate closer to the true CTR, and reduces its standard error, reducing the uncertainty in the model. Since the number of total effective impressions is the same each time period the provider, by choosing an allocation chooses, within the constraints of the squashing factor allocation mechanisms, which advertisers receive which number of effective impressions. One can think of the learning process as the provider holding a certain number of "information tokens" and distributing them among the advertisers to learn about their click-through-rates.

3.2.3 Weaknesses of myopic revenue maximization

It is clear to see that with accurate CTR estimates, the estimates of expected revenue are likely to be correspondingly accurate, and the myopic algorithm executes optimal behavior, as it maximizes expected revenue. In this situation, where we have reached an information state where no additional information will materially change our beliefs and actions, the myopic policy is optimal. However, in most situations, incorrect CTR estimates will lead to incorrect and suboptimal squashing factor choices and allocations. In addition, these allocations are self-perpetuating, as they result in little to no exploration on alternative choices, so that suboptimal behavior is both unacknowledge and unremedied.

The following example illustrates the weakness of myopic algorithm due to lost opportunities from slow learning. When click-through-rate estimates are low for high value advertisers, the myopic algorithm will place these advertisements in lower slots, resulting not only in lost revenue due to inefficient allocation, but also compounding the problem by resulting in slow refinement of these incorrect estimates due to low information gains from lower slots.

Example 6. *Consider a two slot, three advertiser sponsored search auction, with the top* *slot having position effect $x_1 = 1$ and the bottom $x_2 = 0.5$ so that the top slot receives twice as many clicks as the bottom. Let $E = \{0.25, 0.5, 0.75\}$ and $V = \{$ \$0.25,\$0.50, \$0.75 $\}$ but let the estimated CTR for the third advertisement be inaccurate, and in particular, lower than the true CTR: $\hat{E} = \{0.25, 0.5, 0.1\}$. Then, consider $\Gamma = \{0, 1\}$. From equation 2.12 we have equilibrium revenue calculations*

$$\begin{aligned} R[0, E] = 0.1937 \quad &> \quad 0.1125 = R[1, E] \\ R[0, \hat{E}] = 0.08 \quad &< \quad 0.1125 = R[1, \hat{E}] \end{aligned}$$

In this case, the myopic algorithm will repeatedly set $\gamma = 1$, and the allocation [2 3 1]. This places the third advertisement, which has the highest value but most uncertain estimate, in the lower slot, whereas the optimal allocation with $\gamma = 0$ would be [3 2 1]. The greedily chosen squashing factor and allocation receives both lower revenue, and exacerbates the information problem by placing the third advertisement in a position where it is receiving very few effective impressions and thus its estimate is only very slowly being refined.

Due to the lackluster learning speed of the myopic algorithm, it can often find itself mired in a situation where it is consistently setting suboptimal allocations, but is not refining information sufficiently quickly to address its incorrect valuations.

3.3 Variance Based Exploration

At the other end of the spectrum of algorithms balancing exploration and exploitation are algorithms that exclusively explore. There are a large number of ways to explore, including entirely random exploration, targeted search and uncertainty minimization methods. We will focus on the latter of these in developing an algorithm that explores by attempting to greedily minimize standard error on click-through-rate estimates.

Recall from section 3.2.2 that we model the uncertainty of our knowledge about click-through-rates as normal, with mean the clicks to effective impressions ratio, and variance the usual standard error. Specifically, the standard error $\sigma/\sqrt{n}$ is reduced primarily by increasing the sample size, or in this case, the number of effective impressions an advertisement has received. As n tends to infinity and each advertisement becomes shown an arbitrarily large number of times, the standard error of these estimates tends to 0 and we have beliefs and estimates free of uncertainty (but that certainly still may be incorrect).

A simple and intuitive exploration algorithm can aim to reduce knowledge uncertainty each time period by greedily allocating the maximal number of effective impressions or samples to the advertisement with the highest standard error. In the "knowledge token" view, this is the natural strategy of allocating as much learning or exploration effort as possible to the advertisement about which the least is known.

Each time period, the variance based exploration algorithm takes the following steps

1. Construct potential squashing factor set Γ as in the myopic algorithm (described in section 3.2.1)

2. Calculate, for each advertisement, the standard error of the estimated click through rate as described in section 3.2.2, and select the advertisement s with the highest standard error, i.e. s such that $SE_s = \mathbf{max}_i(SE_i)$ for $i = 1, \cdots, n$.

3. Calculate, for each squashing factor, the associated allocation, and determine the number of effective impressions that will be allocated to advertisement s in that allocation

4. Choose the squashing factor that maximizes the number of effective impressions allocated to s

We see that this is essentially the exploration-maximizing analogous algorithm to the myopic revenue-maximization algorithm: rather than maximize expected revenue, we greedily maximize the number of effective impressions given to the advertisement about which the least is known.

3.3.1 Squashing factor intervals and possible allocations

The variance based exploration algorithm highlights a property of our formulation of sponsored search auctions that is relevant to all the algorithms we consider, but is of particular note here because we are very concerned with where a single advertisement, specifically, the one with highest standard error, appears. The squashing factor mechanism is elegant and general and admits an extremely clear and precise game theoretic characterization of equilibrium behavior, but it does not allow arbitrary allocations, or even the ability to tune allocations with some degree of specificity. It is generally the case that changing γ will result in a number of different shifts in the allocation.

This is of particular concern in the variance based exploration algorithm because there exist cases where it is literally impossible, given the current estimates of click-through-rates, to set any squashing factor within some predetermined interval that allows a specific advertisement to appear at all. Put more concretely, it may be that for all $\gamma \in [0, 1]$, the interval that we typically consider, the scoring mechanism places a given advertisement in places $n - k$ to n, that is, outside of the displayed slots.

This naturally leads us to consider expanding the possible squashing factor interval we consider. $[0, 1]$ is the most natural interval, as it is bounded by the two most common scoring mechanisms, rank-by-bid and rank-by-revenue, but there is no technical reason why this interval cannot be arbitrary, or indeed, why an interval is necessary at all. It is certainly possible to simply consider all of the possible swap squashing factors calculated in 3.2.1 and to allow all of these to be potentially used by the algorithm. Even without restriction to an interval, there are not an infinite number of possible squashing factors to consider – the n^2 swap squashing factors are the only ones where a shift in allocation can occur, and any squashing factors outside of the extrema of this set will simply result in the same allocation as the extrema.

In addition to providing inspiration for the consideration of wider potential squashing factor intervals, the variance based exploration algorithm offers a comparative tool for analyzing speed of convergence of click-through-rate estimates. As a simple, intuitive heuristic for greedy exploration, it serves as a foil for the myopic revenue-maximizing algorithm, and thus allows for a bounding of the behavior and performance of the metareasoning algorithms by two greedy algorithms which each separately care about exploration or exploitation.

The setting and benchmarks for our value of information based metareasoning process will be this multi-stage sponsored search auction and the heuristic greedy exploration and exploitation policies.

Chapter 4

Metareasoning and Value of Information

Metareasoning is a general principle whose goal is, as expressed by Russell and Wefald, to "provide a basis for selecting and justifying computational actions" [20]. It is thus a tool within the framework of bounded-rationality decision theory used to measure the benefit and cost of computation, and to analyze and formulate optimal computational policies. In an uncertain world, where the outcome of actions are unknown, we rely on probability theory and decision theory to inform rational decision-making. In the sponsored search setting, metareasoning is concerned primarily with the evaluation of the information properties associated with a given allocation and their effect on equilibrium revenue. The auctioneer is faced with the problem of learning about advertiser-specific properties in order to make more accurate and profitable allocations.

Our metareasoning process draws on the calculation of the value of the information gained by setting a specific squashing factor. Since the allocation and distribution of information about click-through-rates is different for each squashing factor, future information states (i.e. CTR estimates) depend greatly on the sequence of allocations selected. This leads us to the natural definition of value of information as the increased future revenue due to more accurate CTR estimates and correspondingly more optimal squashing factor selections and allocations. Our algorithm estimates value of information by simulation into the future, comparing the expected revenue of a default squashing factor with the optimal squashing factor selected with future information. We draw the greater part of our inspiration from the work of Russell and Wefald, but take particular care to adapt their principles to the unique information structure of sponsored search auctions.

4.1 Metareasoning principles

Russell and Wefald provide the framework for the construction of a metareasoning system capable of rational evaluation of computation, information and utility [20]. The key principles on which this system rests are the following:

1. Computational or information gathering actions are actions, and are thus to be evaluated as such, that is, according to their expected utility
2. The utility of a computational action or the value of information gained comes from the revision of intended actions based on changed information states.

These principles express the idea that by refining knowledge of the world, computation and information gathering cause the agent to choose a different, more beneficial course of action, thus increasing agent utility. The value of computation, or the expected utility gain that can be attributed to computation is the difference in agent utility from taking a new course of action versus a default course of action.

The classic metalevel decision problem is framed as follows. At any given point in time, an agent is faced with the following decision - either take *external action* $\alpha^0 = \textbf{arg max}$ $(\alpha_1, \cdots, \alpha_n)$ deemed to have the highest utility in the current information state, or take one of $C_1, \cdots, C_k$ *computational actions*. This distinction between a computational action which affects only an agent's internal knowledge state and not the world at large except through computational costs (passage of time, for example), and external actions, which impact the real world, is an important one. It allows the metalevel decision-making process to entirely disentagle the value of information and real utility gained from interaction with the external environment. After a computational action C_j is taken, a new (possibly but not necessarily different) optimal external action α_j^* is recommended.

Following Russell and Wefald's principle that the utility of a computation resides in its effect on the agent's choice of best action, the general form of the net utility of a computation C_j is [20]:

$$V(C_j) = U(C_j) - U(\alpha^0) \tag{4.1}$$

This is the first and basic value of information calculation that will be successively refined for use in our algorithms. Note that we still have not defined the utility of a computation, only its value in comparison to current default action. In addition, note that these calculations are with reference to *true* utility.

4.1.1 Complete and partial computations

There is an important distinction arising between *complete* and *partial* computations. A complete computation is one where recommended action α_j *must* be taken after the computation, whereas a partial computation does not result in a binding commitment. The utility calculation for a complete computation is thus simple: it is the difference in utility from new recommended action α_j^*, given that we took computation C_j, and current default best action.

$$V(C_j) = U(\alpha_j, C_j) - U(\alpha^0) \tag{4.2}$$

For partial computations, the final external action is uncertain, and may not depend directly on the computational action C_j taken. Thus the utility of computation C_j cannot be expressed solely in terms of α_j^*. Calculation of its estimated value must range over all possible paths of completing the computation and arriving at an external action. Letting $C'_1, C'_2, \cdots$ represent the possible complete computations following partial computation C_j, and let $\alpha'_1, \alpha'_2, \cdots$ be the associated optimal actions, and $P_(S'_i)$ be the probability of performing complete computation C'_i, we arrive at the following characterization of the value of a partial computation C_j.

$$V(C_j) = \sum_i P(C'_i) U(\alpha'_i, C_j \cdot C'_i) - U(\alpha^0, C_0) \tag{4.3}$$

Where we take $C_j \cdot C'_i$ to mean the computation C_j followed by the computation C'_i.

Russell and Wefald, and the tradition decision theory framework combine these equations into the following optimal metareasoning algorithm [20].

1. while $\mathbf{max}_i(V(C_i)) > 0$: take computational action C_i such that C_i maximizes $V(C_i)$

2. take external action α_i recommended by the sequence of computational actions taken in step 1. That is, take external action with highest expected utility with respect to the internal information state resulting from the computations taken.

This simple algorithm captures the essential goal of decision theory information gathering: obtain information and data through computational actions until their cost outweights their benefit, that is, when their net utility is no longer positive, and perform the best action after that.

4.2 Utility estimation in metareasoning

The above framework is useful for reasoning about a fully rational agent or omniscient external observer. In particular, it assumes the agent has access to the true utility function U, which is generally either unknowable or uncomputable for an agent with limited rationality. We can reformulate the equations from above using utility estimates $\hat{U}$ which are calculated estimations of true utility with respect to the internal computational state. Let $\hat{U}^{\mathbf{C}}$ represent the estimate of utility after sequence of computations $\mathbf{C}$. Then, when considering the value of information of computational action S_j, at a time after sequence of computations $\mathbf{C}$, equation 4.1 becomes

$$\hat{V}(C_j) = \hat{U}^{\mathbf{C}\cdot C_j}(C_j) - \hat{U}^{\mathbf{C}\cdot C_j}(\alpha^0) \tag{4.4}$$

For complete computations, we have only that the utility of action α_j^* is replaced by the estimated utility, so that equation 4.2 becomes

$$\hat{V}(C_j) = \hat{U}^{\mathbf{C}\cdot C_j}(\alpha_j^*, C_j) - \hat{U}^{\mathbf{C}\cdot C_j}(\alpha^0) \tag{4.5}$$

For partial computations, we now can relax the assumption that the agent will optimally complete the computation, and can formulate the following net value for a partial computation, which assumes that the agent maximizes percieved or estimated utility when she chooses to act[1]:

$$\hat{V}(C_j) = \sum_i \hat{P}^{\mathbf{C}\cdot C_j}(C_i) \max\nolimits_i \hat{U}^{\mathbf{C}\cdot C_j\cdot C_i}(\alpha_i^*, C_j \cdot C_i) - \hat{U}^{\mathbf{C}\cdot C_j}(\alpha^*) \tag{4.6}$$

When considering whether to take computational action C_j, the agent does not have access to the ex post utility estimates $\hat{U}^{\mathbf{C}\cdot C_j}$. Thus, it must rely on statistical knowledge gained from previous draws from its distribution in other time periods to estimate its expectation. We can thus view $\hat{V}$ as a random variable, whose expectation is, in the continuous case, calculated as an integral or approximated through a Monte Carlo simulation, or, in the case of discrete outcomes, through a summation. When we take computational action C_j, call $\mathbf{u} = \{u_1, \cdots, u_n\}$ the new utility estimates for external actions $\alpha_1, \cdots, \alpha_n$ and call $p_j(\mathbf{u})$ the joint distribution of the external actions over their new utility estimates. Recalling that, after computation C_j is complete, the highest perceived utility action will

[1] Russell and Wefald note that this is far more plausible than the assumption that the agent acts optimally with respect to true utility. Rather than assuming a fully rational agent, this assumes only that the agent maximizes expected revenue with respect to beliefs[20].

be recommended, we have the following representations, the relevant forms of the value of information characterization from Russell and Wefald [20].

$$
\begin{aligned}
\mathrm{E}[\hat{V}(C_j)] &= \mathrm{E}[\hat{U}^{\mathbf{C}\cdot C_j}(C_j)] - \hat{U}^{\mathbf{C}\cdot C_j}(\alpha^0) && (4.7)\\
\mathrm{E}[\hat{V}(C_j)] &= \int_{\mathbf{u}} \max(\mathbf{u}) p_j(\mathbf{u}) d\mathbf{u} - \int_{-\infty}^{\infty} \mathbf{u} p_{\alpha_j}(\mathbf{u}) d\mathbf{u} && (4.8)
\end{aligned}
$$

This equation captures the idea that the expected value of a computation lies in the expected increase in revenue from the new optimal action in the possible new information states over the default optimal action. This final equation will be the form of the net utility estimates that our agent maintains in order to perform rational metareasoning about its computational options and external actions. The key characteristics of this equation arise from the shift from true to estimated utility, and from ex post realization of a random variable to ex ante expectation outlined in the above sequence of equations.

4.2.1 Related work in metareasoning

Work in metareasoning and decision theory can trace its roots to the seminal work of von Neumann and Morgenstern on expected utility theory [26]. The principal tenet of von Neumann's work, the *maximum expected utility principle*, holds that perfectly rational agents always act to maximize their expected utility. Since expected utility is made in reference to agent beliefs, this principle does not require omniscience about the actual probability of events, but requires that the agent act in agreement with its internal beliefs about these probabilities. Von Neumann-Morgenstern rationality has seen widespread application in economics and related fields, forming the basis for much of game theory, mechanism design, and other agent based systems and analysis. However, real economic actors are limited by finite computational capacity and finite computational time, meaning that rarely, if ever, are they able to make fully rational decisions.

These shortcomings were pointed out by Simon as an explanation for the patently irrational behavior of real agents [22]. Simon proposed that the behavior of agents and individuals in real economic situations was largely determined by unsufficient information and reasoning power, and the corresponding inability to maximize expected utilty according to the decision-theoretic framework. Simon is usually credited with coining the term "bounded rationality," and pioneering the study of problems in which the cost and value of computation was critical. Closely related was the work of Good, where he delinates the difference between classical, perfect, or "type I" rationality, and "type II" rationality, where an agent maximizes net utility net of computational costs [8]. In a more modern conception, Russell characterizes a type II rational agent as one that after deliberating and acting, has maximized its subjective utility compared to all other deliberation/action pairs . According to Russell's interpretation of Good's work, type II rationality is intended to provide a mechanism by which optimal computational sequences could be calculated and recommended. Thus, it is a first incarnation of a metareasoning process - reasoning about reasoning and computation - designed to formally and systematically analyze the value and impact of computation [19].

Metareasoning based on the value of information, was largely pioneered by Howard in his 1966 work "Information Value Theory" which attempts to ascribe economic value to the reduction of uncertainty [11]. He introduces the critical idea that the value of information in reducing the uncertainty of outcomes depends on both the probabilistic impact of refined outcome possibilities and the economic impact of refined estimates of outcome utility. Most of modern metareasoning is based on this characterization of value of information. Reasoning about computation proceeds through reasoning about a computatational action's economic impact in reducing uncertainty and refining perception of the future.

A central area of focus for these algorithms and metareasoning has been the time-sensitive decision-making of expert recommender systems. This domain is particularly well suited to apply the principles of metareasoning - there is generally a clear distinction between computational actions and external actions, there is a well defined and evaluable utility function, and there is a natural cost of computation. Heckerman's Pathfinder, for example, is an expert system that aids in the diagnosis of lymph-node diseases, and relies on decision-theoretic tools for reasoning about the cost in patient well-being of making an immediate recommendation or continuing to perform diagnostic tests [10, 9].

In work more closely related to the sponsored search domain, Russell and Wefald successfully apply metareasoning as a search control algorithm in the more traditional computer science setting of competitive games. They demonstrate significant improved efficiency using their metareasoning framework over traditional search methods such as alpha-beta pruning [20]. Boddy and Dean successfully apply metareasoning to a robot control setting, demonstrating near-optimal control that is far more computationally inexpensive than more complete decision-theoretic reasoning. The term that Boddy uses to characterize his algorithms are "expectation-driven iterative refinement" which neatly captures the relevant principles of metareasoning [5, 3].

A second line of related work on metareasoning and the exploration/exploitation trade-off has been the multi-armed bandit problem, where a gambler must choose among a series of K arms of a slot machine, whose associated reward is drawn from a distribution that he is uncertain about. Choosing an arm advances the internal state of the arm and possibly changes its reward distribution, while the states of all other arms remain unchanged.

Gittins and Jones presented a *dynamic allocation index* for computing the value of each arm in the multi-armed bandit problem, that reduces the problem to a series of one-dimensional stopping problems [7]. When the distribution of rewards behind each arm are known, the calculation of the Gittins indices allows for the construction of an optimal policy. In addition, Katehakis demonstrated that the calculation of the Gittins indices is analogous to the well-known "restart-in-i" Markov decision process problem, and thus traditional MDP policy solution and approximations methods can be used to solve or approximate optimal policies for bandit problems in this domain [13]. Under uncertainty about reward distributions, however, there have been a large array of algorithms proposed to address the explore/exploit tradeoff of multi-armed bandit problem. These include ϵ-greedy methods, which are variants on myopic revenue maximizing with small probability of random exploration, probability matching algorithms, which ascribe a probability to each arm depending on its likelihood to be optimal, and reward interval estimation algorithms [25]. The mechanics of interval estimation algorithms are very similar to the uncertainty simulation mechanism we utilize in simulating expected revenue, in that actions which have been taken less often and about which there is more uncertainty are ascribed higher values and are thus more likely to be explored [12].

The multi-agent bandit problem was first formally described by Robbins, and has since been used to model a number of real world problems such as pharmaceutical clinical trials and adaptive routing mechanisms [17, 25]. Katehakis describes the problem in the context of project scheduling, where each time period, a project manager observes the current progression of a number of projects and allocates resources to advance the state of a project, both gaining information about that project and reward from the project itself [13]. In addition, bandit models have been successfully applied to market learning in industrial organization, buyer/seller matching in experience goods markets, and venture funding in corporate finance and asset pricing models [18, 2]. The bandit model is useful for its simplicity and decomposability, while remaining robust and powerful enough to be applied to problems of sufficient realism and applicability to be interesting.

We will draw on lessons from both work in traditional decision-theoretic metareasoning and from the exploration/exploitation tradeoff in multi-armed bandit problems in constructing an effective approach toward rational metareasoning in the sponsored search setting.

4.3 Metareasoning in sponsored search

There are a number of peculiarities of the sponsored search domain that distinguish it from both expert recommender systems and the multi-armed bandit problem. While metareasoning remains useful and powerful, this colors its application to this domain with a unique flavor and introduces a number of different problems requiring careful consideration and treatment.

4.3.1 Information richness and scarcity

Primary among the concerns in applying metareasoning to the sponsored search setting is the lack of a clear, or indeed, any delineation between computational and external actions. The fact that all actions accrue some knowledge makes extremely important the careful consideration of the informational landscape of the environment. Since the metareasoning process is no longer a binary distinction between reward-recieving, zero-information external actions and zero-reward, information gathering computations, the myopic exploitation policy is concomitantly and inadvertently exploring. In an environment where the decision-maker receives a large amount of information from the environment regardless of the action taken, there will be little benefit to metareasoning. There must be a sufficiently large potential gap in information received by the optimally exploring policy and the optimally exploiting policy in order to meaningful delineate the usefulness of information accrued by different actions.

The sponsored search setting offers a very natural and effective way to tune the amount of information gained each time period. Since the uncertainty of click-through-rate estimates is tied directly to the number of effective impressions an advertisement receives, tuning the total number of effective impressions received per time period amounts to tuning the information richness of the environment. While the absolute amount of information gained will decrease for all actions, the relative amount of information gained will become more important. This is intuitively equivalent to varying the number of “information tokens,” thus varying the amount of information the provider can gather each time period.

Typically, we will operate in a moderately information-scarce setting, as this will allow the metareasoning process to meaningfully trade off information and current-period revenue. We can think of an information-scarce environment as the sponsored search provider setting an allocation that persists for a number of hours or half a day, rather than a longer time period. In addition, since the position effects are normalized so that the top slot receives 100% of the literal impressions as effective impressions, setting a low number for the number of effective impressions per time period can also be interpreted as downweighting the percentage of technical impressions received as effective ones. In other words, this also captures the idea that that not even appearing in the top slot will guarantee that everyone will even look at an advertisement.

4.3.2 Sponsored search and the Multi-Armed bandit problem

The informational properties described above bear some similarity to the multi-armed bandit setting, where activing a given arm provides both information about its underlying distribution of rewards and also the reward itself. The bandit setting thus also does not have a meaningful distinction between computational actions and external actions. However, the literature on the MAB problem takes into account the concept of value of information in a number of different ways. These methods range from indirect, such as the greedy overvaluation of interval estimation algorithms and the weighting of optimal probabilities by times played in probability matching algorithms, to very specific, such as attempts to add an “exploration bonus” to the reward received, or to price the knowledge gained [25]. The key lesson of these efforts is that there are a multitude of different ways in which the effect of the concept of value of information and exploration can be felt, even if they are not explicitly characterized. A unifying guiding principle of all of these algorithms is the effort to internalize and encapsulate the effect of uncertainty and information within the estimates of the value of different actions.

In the MAB setting, the information gains from exploration remain clearly separated – each arm provides only information on its own underlying distribution, whereas in the sponsored search setting, any squashing factor and allocation will provide information on click-through-rates, which will inform estimates of the value for all other squashing factors and allocations. In any assessment of the value of reducing uncertainty or gaining information, the effect of the increased information on all other arms must be considered. This makes it difficult to extend the probability matching algorithms and other algorithms tailored to the multi-armed bandit domain to the sponsored search domain, as the updating of probabilities and weights does not extend to this correlated information case.

The above discussion has assumed that we are viewing the various squashing factors and their associated allocations as the “arms”. In one sense, this is the most natural conception, as the squashing factors are the actions that the sponsored search provider takes, but in another sense, this is one step more complex and removed from the basal revenue drivers, which are the advertisements themselves. If one thinks of the advertisements as the arms, then the distribution behind each arm is a simple Bernoulli. The sponsored search learning problem is then the familiar MAB problem of discovering the parameters underlying the distribution of each arm, with the change that in each time period, instead of specifying a single arm to activate, the choice of a squashing factor and allocation specifies a bundle of concomitant activations. Unfortunately, however, it has been shown that a Gittins index characterization of arms is not possible when multiple arms can or, as in this case, must be selected at each time period [2].

The multi-armed bandit problem provides a instructive framework for understanding some of the exploration/exploitation tradeoffs of the sponsored search domain. While the results from its relevant literature do not extend directly to the sponsored search auction setting, they are certainly of considerable use in formulating how one ought to think about trading off current revenue in favor of refined information.

We see that the sponsored search domain features some of the problems of both the traditional metareasoning setting and the multi-armed bandit problem. Like the MAB domain, there is no clear distinction between computational and external actions, and like the metareasoning domain, exploration provides information about the utility of all actions, not only the specific action taken.

4.3.3 Computational and external actions

Correctly calculating value of information for actions that receive both reward and information is primary obstacle that must be navigated to apply a metareasoning framework like Russell and Wefald's to the sponsored search domain, and resolving it proves to be the most critical contribution to the success and applicability of value of information based metareasoning to sponsored search auctions.

An agent action in the sponsored search setting consists of setting a squashing factor γ, which, with reference to a set of beliefs about click-through-rates, entails a certain allocation of advertisements. The setting of a squashing factor is both the mechanism by which the auctioneer agent learns about the true click-through-rates of advertisements and receives reward from the external world. When a squashing factor and associated allocation are set, the true number of clicks observed provides both reward in form of advertiser payments and information in the form of new observations for clicks and impressions that refines click-through-rate esimates. This is an environment drastically different from the traditional metareasoning domain, where computational actions do not receive any reward, and external actions do not provide the agent with any information.

This means that even when myopically exploiting, the agent is still recieving information in the form of refined click estimates for the ads that happen to be shown. The uncertainty model over click-through-rate estimates combines with this fact to ensure that a myopically exploiting algorithm actually explores a non-trivial amount, and will, given sufficient time, converge to an information state with extremely accurate click-through-rate estimates.

The problem of metareasoning then becomes a question of degree: rather than trading off immediate revenue for information in an all-or-nothing exchange, metareasoning in the sponsored search setting must consider forsaking some amount of immediate revenue in return for better information in the future. In the framework of metareasoning presented above, we no longer have computational actions $C_1, \cdots, C_k$ and external actions $\alpha_1, \cdots, \alpha_n$, but only actions $A_1, \cdots, A_m$, all of which have some associated information value and and some associated immediate revenue.

This impacts our decision-making framework in a number of places. The first is in the calculation of α^*_j, the optimal action recommended after taking complete computation C_j. Equation 4.5 tells us that the net value of a computational action C_j after taking it is simply the expected utility from the optimal action α^*_j minus the expected utility of previous default action α^0. This applies equally well to a repeated setting or a one-shot setting. In the original repeated setting, optimal external action α^*_j will never change as long as no further computational actions are taken, since no additional information is gained from taking external action α^*_j. However, in the repeated case where additional information is accrued by taking action α^*_j, optimal action will change over time. The import of this change is that all computations, even those committed to taking a subsequent optimal action, become like partial compuations, so that the first term of 4.5 must be projected over all possible future actions, each of which also will provide information that further changes optimal actions. Thus,

$$\hat{U}(C_j) = \hat{U}^{\mathbf{C}\cdot C_j}(\alpha^*_j, C_j)$$

becomes the following chimera: letting C_t be the sequence of computations leading up to time t and A^*_{t+1} be the optimal action (i.e. having highest expected utility) recommended in time $t+1$ by this sequence, the estimate of the value of an action A_j in time t is then

$$\begin{aligned} \hat{U}(A_j) &= \hat{U}^{\mathbf{C_t}\cdot A_j}(A^*_{t+1}) + \hat{U}^{\mathbf{C_{t+1}}\cdot \mathbf{A^*_{t+1}}}(A^*_{t+2}) + \cdots && (4.9) \\ &= \hat{U}^{\mathbf{C_t}\cdot A_j}(A^*_{t+1}) + \sum_{i=t+1}^{\infty} \hat{U}^{\mathbf{C_i}\cdot A^*_i}(A^*_{i+1}) && (4.10) \end{aligned}$$

In words, we consider taking action A_j and then following the policy that chooses actions in subsequent time periods to maximize expected revenue. The impact of an action A_j on the subsequent sequence of expected revenue maximizing actions is driven by the information state arrived in at time $t + 1$ after taking action A_j, as this internal state will inform the subsequent sequence of expected revenue maximizing actions. We are thus treating each action as a complete computation in that we assume it will follow a revenue maximizing policy, but also recognize that each action within the revenue maximizing policy provides additional information which may change subsequent optimal actions.

The computation of the sequence of optimal actions is generally intractable, as it requires projection or simulation until the end of time. Methods for approximating or simplifying this calcuation will prove critical to our metareasoning efforts. Acknowledging and understanding the non-standard information properties of this system proves to be the single most important factor in applying a value of information based metareasoning process to the sponsored search setting.

4.4 A Metareasoning algorithm for sponsored search auctions

Despite the potential difficulties and peculiarities of applying metareasoning to the sponsored search domain, it is clear from the weaknesses of the myopic algorithm that there is a great deal of value to be gained from an intelligent rational decision-making process. The metareasoning process attempts to calculate the value of the information gained by setting a specific squashing factor and allocation by analyzing the effect of this additional information on the optimally exploiting policy. Each squashing factor and associated allocation gives a certain number of impressions and expected clicks for each advertisement – this information refines next-period click-through-rate estimates, and if the information is valuable, a better squashing factor and allocation will be chosen then. Thus, the value of information gained by setting a specific squashing factor is the increased revenue in the future attributable to refined and more accurate click-through-rate estimates. An effective metareasoning process will trade off current-period revenue in favor of information that will allow for more informed future decision-making.

Our metareasoning algorithm consists of the following steps, which parallel the steps in Russell and Wefald's optimal control algorithm from section 4.1.1 [20].

1. Identify squashing factor set Γ and default best action γ^0 maximizing expected revenue received this time period. This process is exactly identical to the myopic revenue optimization algorithm presented in 3.2.1.

2. For each γ in Γ calculate the value of information through the following Monte-Carlo

simulation

(a) Determine the allocation of slots and advertisements, and the associated number of effective impressions accrued to each advertisement

(b) Simulate the number of clicks each advertisement will receive next period

(c) With the new simulated click-through-rate estimates, repeat the myopic revenue optimization process to find the new optimal squashing factor γ^*

(d) The value of information is the additional expected revenue gained in from setting squashing factor γ^* instead of γ^0

(e) Repeat this process for a number of time periods into the future, and for a large number of Monte-Carlo steps

3. Choose the squashing factor that has the highest sum of expected revenue and value of information

4.4.1 Complete computations and meta-greedy assumption

The metareasoning algorithm takes what Russell and Wefald call the "meta-greedy assumption," which allows us to treat all actions as complete computations [20]. The standard assumption is that the current time period is the last time period in which we will be allowed to take a computational action and receive information, and that for the rest of the future, only external actions will be taken. Since there is no distinction between external and computational actions in the sponsored search setting, this assumption is recast as: each time period, we assume that this is the last time period in which we will forgo current revenue in favor of better information, and that for the rest of the future, we will be following the myopic revenue maximizating policy. This is distinct from the traditional metareasoning framework, where due to the lack of information gained by taking external actions, once the agent commits to taking a certain action, nothing will change for the rest of time. For the metareasoning case, we assume that the agent commits to the *policy* of myopic exploitation after time t, but the squashing factors chosen in the future by this policy will vary, due to the additional information gained each time period.

Of course, we make the same assumption and follow the metareasoning process next time period, so the assumption is never technically accurate, but it allows us to meaningful reason about the value of information on the myopic exploitation policy. Technically, the meta-greedy assumption allows us to directly apply equation 4.9 without consideration of the spectrum of possible ways of completing a partial computation. This allows us flexibility to abstract away from the cascading effect of different information and the inter-temporal credit assignment problem of attempting to calculate the value of information for a sequence of partial computations.

4.4.2 Value of Information

Our concept of value of information captures the fact that information in the form of clicks and impressions affects future revenue by refining click-through-rate estimates and changing the future behavior and perception of optimal revenue-maximizing action. Recalling equation 4.9 we see that in the sponsored search setting, the estimated value of information calculation for squashing factor γ is

$$\hat{V}(\gamma) = \hat{U}^{\mathbf{C_t}\cdot\gamma}(\gamma^*_{t+1}) - \hat{U}^{\mathbf{C_t}\cdot\gamma}(\gamma^0) \tag{4.11}$$

Where, as above, C_t is the sequence of actions (i.e. squashing factors) set up to time t. In other words, when we denote $\hat{U}^{C_t}(\gamma^0)$ we mean the estimated value setting squashing factor γ^0 with respect to the information state, that is, the click-through-rate estimates after having set the sequence of squashing factors C_t in the past. And when we denote $\hat{U}^{C_t\cdot\gamma}(\gamma^0)$ we mean the estimated value γ^0 with respect to the information state at time $t+1$ where we have set squashing factor γ in time t, and refined our click-through-rate estimates based on the allocation associated with γ. This characterizes the following metareasoning intuition: "The value of information I gain from setting a squashing factor in time t is the increased revenue I would expect from taking the best action, knowing what I know now in time $t+1$, versus the increased revenue I would expect from taking the default action, *still knowing what I know in time* $t+1$." The fact that one compares the revenue from the optimal squashing factor and the default squashing factor with respect to the same information state is very important, as comparing squashing factors and estimated revenue with respect to different CTR estimates is ultimately meaningless.

Figure 4.1 shows the impact on expected revenue curves of the different information states resulting from setting different squashing factors. Equation 4.11 ensures that all comparisons between squashing factors are made on the same curve, so that the information underlying the calculation is the same.

Earlier, we had utilized a value of information calculation that did not make this distinction, calculating value of information as

$$\hat{V}(\gamma) = \hat{U}^{\mathbf{C_t}\cdot\gamma}(\gamma^*_{t+1}) - \hat{U}^{\mathbf{C_t}}(\gamma^0)$$

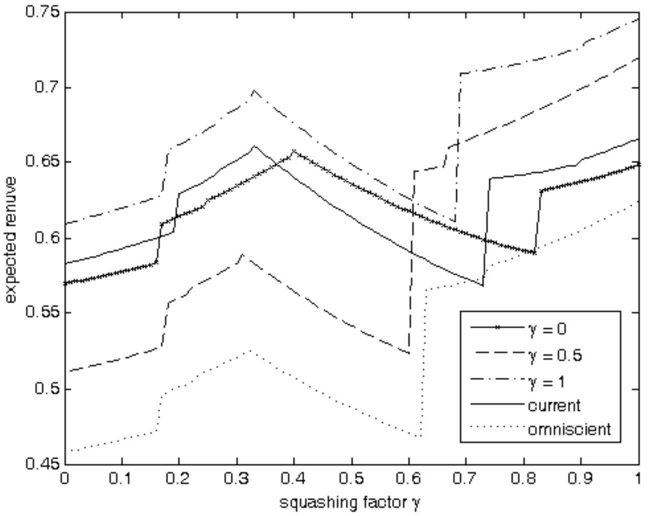

Figure 4.1: Different time $t + 1$ Expected Revenue Graphs resulting from setting different γ in time t

so that the expected revenue for new γ^* was calculated with respect to the new information state, but for default action γ^0, it was calculated with respect to the old information state. This can be seen on figure 4.1 as comparing, for example, the expected revenue of perceived optimal $\gamma = 0.3$, 0.65 with current time information (solid line) to new perceived optimal $\gamma = 1$, 0.7 with information resulting from setting $\gamma = 0.5$. This drastically understates the value of information of setting $\gamma = 0.5$, because not only does $\gamma = 0.5$ result in the identification of a new and better optimal squashing factor, it (more importantly) refines the revenue curve, bringing it closer to the omniscient curve.

We make two further assumptions in order to facilitate metareasoning, and to make the process more computationally tractable. In order to simulate the effect of the information gained from setting γ in time t on the optimal myopic exploitation policy *for the rest of time*, we apply 4.11 for times $t + 1, t + 2, \cdots$. Since it is generally infeasible to simulate this for the entire future timeline, we simulate this for a fixed number of steps into the future. This both allows the effect of the additional information accrued by setting γ in time t to be felt over a longer time horizon. Additionally, we make a “big-step” assumption that allows us to amplify the information gained by setting a specific squashing factor. The sponsored search model we use typically operates in an information-scarce setting – the discussion in section 4.3.1 outlines some of the reasons why this is both useful for the metareasoning process and plausible. However, this means that often one time step worth of additional information is insufficient to materially change CTR estimates to shift behavior. Thus, in the metareasoning process, we assume that the next allocation will persist for a longer period of time so that more effective impressions are accrued, or more “information tokens” are present for us to distribute. Note that this assumption does not change the actual rules of the auction or model, merely “tricks” the metareasoning process into thinking that it will receive more information than it actually will. This ensures that the metareasoning algorithm believes that sufficient information will be gained to change optimal revenue-maximizing behavior in the future.

4.4.3 Simulating the future and information uncertainty

When we consider $\hat{V}(\gamma)$ it is important to remember that this is an estimated value over an array of possible future information states. Recall that equation 4.7 evaluates this estimated value by taking an integral over the possible different information states we can arrive in after taking the relevant computation. Since it is not possible to solve for a closed-form solution of our revenue equation, we approximate this value through a Monte-Carlo simulation (step 2 above). We repeatedly draw a set of simulated new clicks and impressions, assuming that advertisement clicks are distributed binomially with $p = \hat{E}_t = C_t/I_t$, the mean estimate of click-through-rates, and n equal to the number of effective impressions accrued. These new clicks and impressions represent the new information state we find ourselves in after having set squashing factor γ.

Figure 4.2 illustrates this process. In time t, the metareasoning process considers setting a number of alternative squashing factors $\gamma_1, \gamma_2, \gamma_3$. The algorithm simulates the potential information states that one could arrive in at time $t+1$ after having set a specific squashing factor in time t – these are the small circles. With respect to these new information states, a new optimal squashing factor γ^* is calculated, and value of information and the utility of each squashing factor is calculated. Simulation in this way answers the question “what will I know in $t+1$ if I do this in t”? The algorithm then goes on to answer the question “if I know that at time $t+1$, what is the best thing I can do and how much better is it than what I would have done ignorantly?” This value of information characterization is the critical component of our metareasoning algorithm.

In summary, our metareasoning process utilizes a value of information calculation to reason directly about the effect of setting specific squashing factors and allocations on CTR estimates and future revenue. We adapted the traditional metareasoning framework to account for the lack of distinction between computational and external actions. By

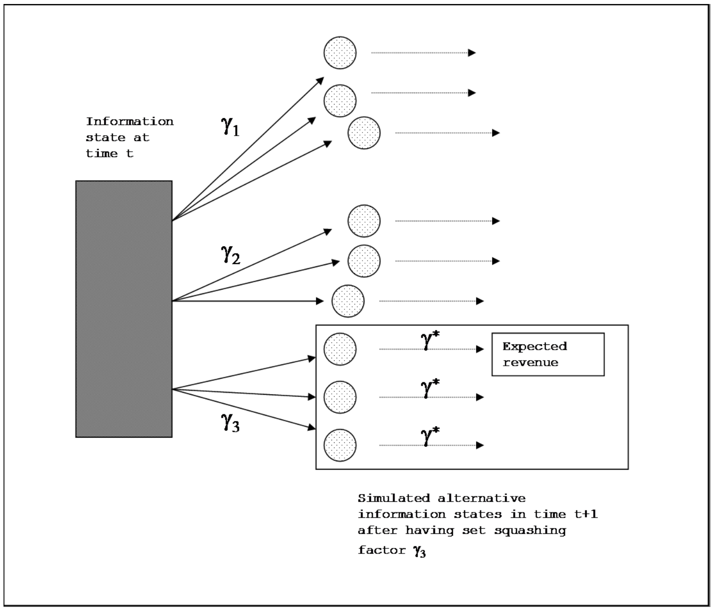

Figure 4.2: Information State Simulation Illustration

reasoning about the effect of its current actions on future information and revenue, our metareasoning algorithm is able to overcome much of the naivety of the greedy algorithms, and to rationally balance exploration and exploitation.

Chapter 5

Performance Results

Now that the concepts behind and motivation for a value of information based metareasoning algorithm for sponsored search auctions are clear, we are interested in how metareasoning performs. We are interested in a number of different metrics, the analysis of which will contribute to a more complete understanding of the suitability, applicability and potential of applying metareasoning to sponsored search auctions.

5.1 Simulation parameters

The following experiments were performed with the following default set of parameters:

- $n = 8$, $k = 5$ – it is convenient to choose $n > k$ so as to eliminate the cases of unsold auctions and
- The marginal distributions of true click-through-rates and values were beta with a = 2.7, b = 25.4 and lognormal with $\mu = 0.35$ and $\sigma^2 = 0.7$. Using a Gaussian copula, we generated correlated CTRs and values with Spearman correlation $\rho_s = -0.5$[1] A correlation of -0.5 was chosen as a value such that weighting by click-through-rate would be important, but that we would also see interesting results with intermediate squashing factors being optimal[2].
- Effective impressions – the top slot received 50 effective impressions per time period.

[1] As Lahaie and Pennock note, a copula is a function that takes two marginal distributions and gives a joint distribution. We follow in referring to Nelsen's *An Introduction to Copulas* for further exposition [14].

[2] Recall from earlier discussion and from Lahaie and Pennock that as Spearman correlation approaches -1, squashing factor 1 becomes optimal [14].

We typically assumed for the “big-step” assumption (section 4.4.2) that the top slot received 5 times as many effective impressions in the metareasoning process.

- Attention decay factor – $\delta = 1.1$, so that the second slot received $50/1.1 = 45$ effective impressions, the third, $50/(1.1)^2 = 41$, etc.

5.2 Revenue

Primarily, we are interested in the revenue properties over time of the metareasoning algorithm, principally in comparison to a myopic expected revenue-maximizing algorithm, but also in comparison to an omniscient designer with access to true click-through-rates. We are particularly interested in the evolution of revenue dynamics over time. On the exploration side, we use the variance based exploration algorithm as a benchmark heuristic for an algorithm devoted entirely to exploring, and compare the CTR estimate convergence properties of the metareasoning and myopic algorithms to it.

Due to variance in the actual number of clicks received each time period, absolute revenue is an extremely volatile measure. Expected revenue, even for an omniscient designer, is the same each time period, and the squashing factor and allocation set by an omniscient algorithm is the same, but actual revenue received is very different. Thus, in order to meaningfully look at the performance of the myopic and metareasoning algorithms by a less volatile metric, we consider the *cumulative performance ratio*, that is, the ratio of cumulative revenue obtained in time periods $1, \cdots, t$ of a given algorithm, divided by the cumulative revenue obtained in the same time period by the omniscient algorithm. This normalizes performance to the same scale, allowing us to average over a large number of runs with differing true click-through-rates and values. Figure 5.1 shows performance ratio over time for the myopic and metareasoning algorithms, averaged over 15 runs.

We see from this graph that after about 20 time steps, or some 4000 total effective impressions ($20 \sum_{i=0}^{4} 50 * (1.1)^i$) the metareasoning algorithm begins to outperform the myopic algorithm, ending at 50 time steps with a significant advantage over the myopic algorithm.

The range of performance for all time periods is narrow. Performance of the metareasoning algorithm varies only between some 97% and 90% of omniscience and stabilizes at 93% while performance of the myopic algorithm varies between 99% and 89% and stabilizes at just under 90%. The variance in cumulative performance ratio in early time periods is due to the random behavior of both algorithms in a high uncertainty information state. When CTR estimates are highly noisy, standard error of the knowledge model for CTRs

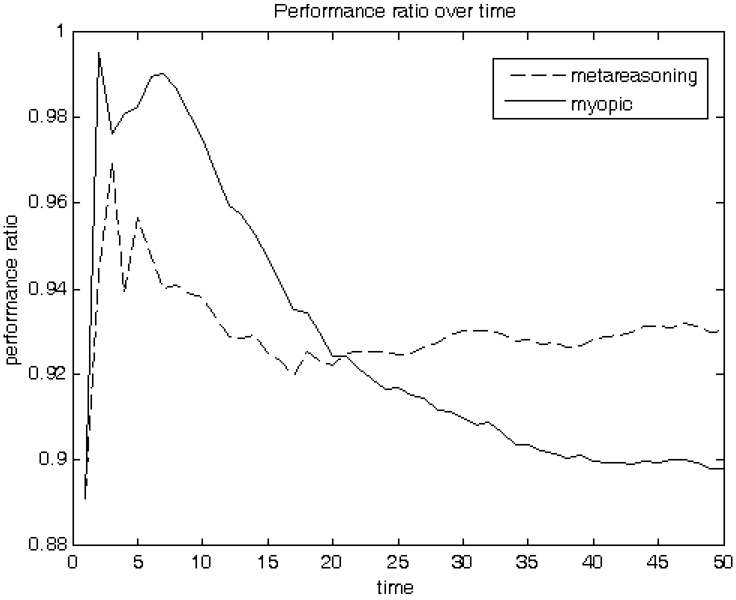

Figure 5.1: Performance ratio over time for metareasoning and myopic algorithms

is high, and decision-making is driven by the wide range of information states the Monte-Carlo simulation can arrive in 3.2.2. This is a significant contributor to the variance in decision-making and revenue in early time periods.

Interestingly, we do not observe the perhaps intuitive pattern where both algorithms significantly underperform the omniscient algorithm at the beginning, and slowly improve their performance as information increases. Rather, we see that both algorithms exhibit noisy and choppy performance in early time steps, but come closest to optimal performance there. We would like to investigate whether this is a consequence of their priors and the initialization of clicks and impressions. Since only a few number of clicks are being observed each time period for a given advertisement (consider an advertisement with CTR 10%, a fairly high value, appearing in the top slot – the expected number of clicks is still only 5), the intialized values of clicks and impressions will have a disproportionately powerful effect in the first several time periods.

We see that after this initial period, the performance of the myopic algorithm drops precipitously. This is often due to a situation highlighting the weaknesses discussed in 3.2.3 – the myopic algorithm repeatedly sets a squashing factor and allocation that is suboptimal with respect to the true click-through-rates, but that also results in an allocation where insufficient information is being learned about either excluded or low-appearing advertisements to refine estimates adequately. The algorithm receives insufficient reinforcement in the form of refined CTR estimates to quickly change its behavior. The performance decline in the middle time periods is due to a protracted period of the algorithm mired in this situation.

It is important to remember that the cumulative performance ratio metric is dominated in early periods by early-period results. That is, each time period does not have equal "weight" in determining the final shape of the graph we observe – early time periods have a disproportionately large effect. This means that the near-optimal performance in early time periods inflates perceived performance in the middle time periods. Even after estimates have stabilized and the algorithm has reached a "steady-state" of expected revenue, optimal squashing factor and associated allocation, the revenue graph will still exhibit some slope as the effect of earlier time periods becomes diluted. Thus, the most important and relevant part of cumulative performance ratio graphs is the tail end of the performance curve, after the algorithms have both plateaued and reached a steady-state where the effect of variable performance in early time periods has been sufficiently diluted.

Related to the importance of the tail end of the performance curves to analysis is the fact that there are two vehicles for the improved performance of the metareasoning algorithm. The first is *faster* convergence of CTR estimates to the "steady-state" so that the algorithm approaches its best performance as quickly as possible, and thus receives its best possible revenue for a longer period of time. The second is *better* CTR estimates at convergence. Note that, as $t \to \infty$, the CTR estimates for both algorithms will approach the true values, as the number of impressions also tends to infinity. In a sense, this is the final, but uninteresting steady-state of both algorithms. However, the quality of the intermediate state reached by the algorithms where no material change in behavior is occurring due to increased information, which is the state observed at $t = 50$ at the tail end of our performance curve, determines the increased revenue accrued by the metareasoning algorithm for the (presumably long) time between behavioral convergence and actual convergence of estimates at infinity.

5.3 Speed of convergence and accuracy of CTR estimates

To understand the reasons underlying the improved performance of the metareasoning algorithm, an analysis of the speed of convergence of CTR estimates is highly instructive. This is the most direct comparison of the information-gathering properties of the algorithms,

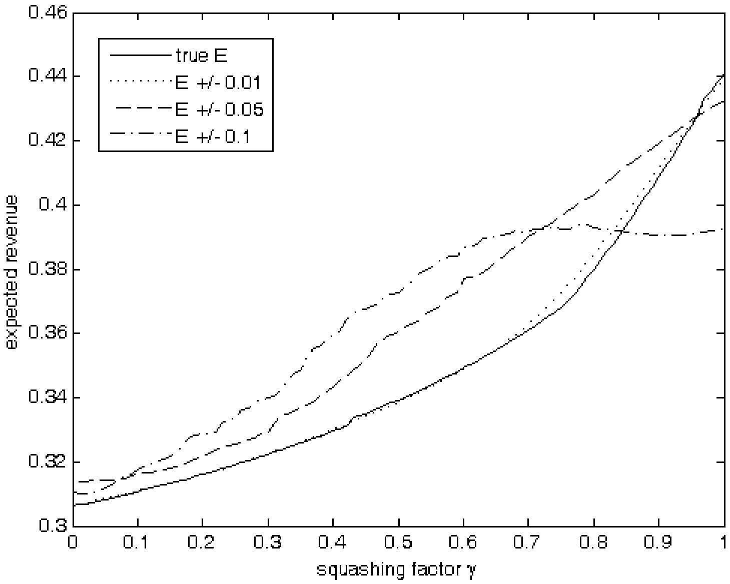

Figure 5.2: Expected Revenue vs. Gamma for perturbed click-through-rates

and relates closely to the consideration of the diminishing returns to metareasoning as time passes and click-through-rate estimates become more and more accurate. The algorithms can be seen as traversing different sets of information states, and this sequence determines the performance differences between the two.

We view the convergence properties of the variance based exploration algorithm as a benchmark against which to compare the speed and accuracy of the metareasoning and myopic algorithms. While the variance based exploration algorithm is not an optimal exploration algorithm, for reasons including but extending beyond the squashing factor interval limitations discussed in 3.3.1, it represents a natural heuristic pure exploration strategy.

The accuracy of click-through-rate estimates is extremely important because it determines not only the expected revenue estimates of each of the algorithms, but also factors directly in the calculation of allocations. Differences in the internal information state of algorithms results in different perceptions of expected revenue and subsequently divergent decisions.

We can observe the deterioration of revenue estimation quality by observing the graph of expected revenue to squashing factor for perturbed click-through-rate estimates (figure

5.2). The "true E" graph shows estimated revenue calculated using the true click-through-rate, and each of the other graphs shows estimated revenue calculated using the true click-through-rate perturbed by a random disturbance drawn uniformly from $[-0.01, 0.01], [-0.05, 0.05], [-0.1, 0.1$ respectively. Figure 5.2 demonstrates that poor quality estimates result in greatly skewed estimates of revenue. Indeed, as these graphs represent the internal information state of the sponsored search provider, they show that in order for the provider's internal information state to be close to the true state of the world, it is necessary for estimates to be quite accurate, less than 5% away from true click-through-rates. This also highlights a weakness of the myopic revenue-maximization algorithm. When estimates are poor, the perceived revenue-maximizing squashing factor γ^0 is likely to be quite different from the true revenue maximizer, and the algorithm is thus likely to be selecting a highly suboptimal squashing factor and allocation.

It is interesting, but not surprising, to note that the largest errors in expected revenue estimation occur for high γ, where the effect of click-through-rates are the strongest, and thus the impact of errors in CTR estimates are most keenly felt. We would thus expect an algorithm which garners accurate estimates to perform particularly well in an environment where higher squashing factors are likely to be optimal. As discussed in section 3.2 and shown in figure 3.1, when true value and click-through-rate are negatively correlated, the impact of intelligent allocations becomes most important, and it is in these domains that we expect an algorithm with accurate and fast estimation refinement capabilities to perform well.

We can begin to understand the click-through-rate estimate convergence properties of the myopic and metareasoning algorithm by examining their view of the world, that is, the revenue predictions made with respect to CTR estimates at different time periods. For us, the relevant views are the expected revenue vs. squashing factor graphs, or what each algorithm "thinks" the revenue landscape looks like. This graph determines primarily the actions of each algorithm at a given time step, and which squashing factors are regarded as attractive and which are not. These graphs are for a representative set of click-through-rates and values drawn from the distributions explained above, and are an example of the way the algorithms regard the world for one specific incarnation of the world. The performance graphs, in contrast, represent an average over a number of runs.

Figure 5.3 shows the relevant graphs for the two algorithms at times $t = 2, 5, 25, 50$. "True E" in both graphs refers to the "true" state of the world, that is, expected revenue graphs using true click-through-rates in determining allocations and payments. Estimates are highly inaccurate after one time period in $t = 2$ for both algorithms, but begin to approach the shape of the true revenue curve as early as $t = 5$. By $t = 50$, CTR estimates for

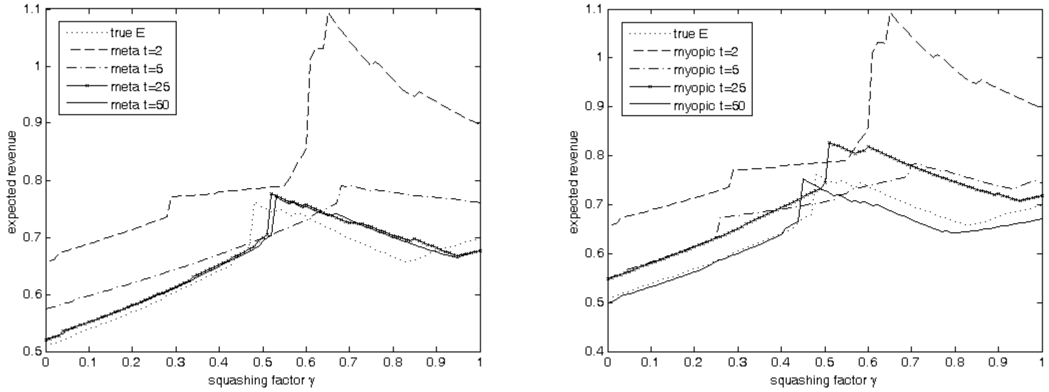

Figure 5.3: **(a)** Metareasoning algorithm "World Views" at various times **(b)** Myopic algorithm "world views"

both algorithms have converged to a reasonable approximation of true CTR, and the revenue graphs are correspondingly similar to the true revenue graph. This provides evidence for the plateau of diminishing returns for metareasoning – by this time, as estimates for both algorithms are very similar and accurate, there is very little additional benefit to be gained from metareasoning. Indeed, the value of information for all actions becomes very low at this point, since additional information does not materially change current estimates, and the behavior of the two algorithms converges as well. The difference in the revenue graphs at $t = 25$ between the two algorithms demonstrates the faster convergence of the metareasoning algorithm – the metareasoning graph at $t = 25$ does not materially differ from the graph at $t = 50$, whereas the myopic algorithm graph is more inaccurate.

This demonstrates the first vehicle for the improved performance of the metareasoning algorithm – faster behavioral convergence to steady state estimates.

5.4 Variance Based Exploration and squashing factor intervals

The variance based exploration (VBE) algorithm presented in section 3.3 is a very naive algorithm. Its behavior is determined entirely by the allocation of a single advertisement, which causes it to lose much of the complexity and reasoning power in an environment as interrelated and interdependent as a sponsored search auction. In addition, it completely disregards revenue, which, at the end of the day, is what a sponsored search provider is concerned with.

For these reasons, VBE is unsuitable as a self-contained adaptive sponsored search auction algorithm. Figure 5.4a, the performance ratio curves for the algorithms run with potential squashing factors drawn from $[-2, 2]$ rather than $[0, 1]$ as is typically the case, demonstrates the poor revenue performance of VBE in comparison to the metareasoning and myopic algorithms.

These performance curves are additionally interesting because they demonstrate that even for the expanded set of allocation mechanisms considered with the expansion of potential squashing factors to a wider interval, the metareasoning algorithm outperforms myopic revenue maximization. This shows that the metareasoning process can meaningfully navigate a larger landscape of potential squashing factors, as the average size of Γ more than doubles when the algorithm expands from $[0, 1]$ to $[-2, 2]$. This is further encouraging evidence that the performance of the metareasoning algorithm is not merely a coincidence of the particular problem formulation we have chosen, but is instead applicable to and valuable in a wide range of related dynamic sponsored search auction mechanisms.

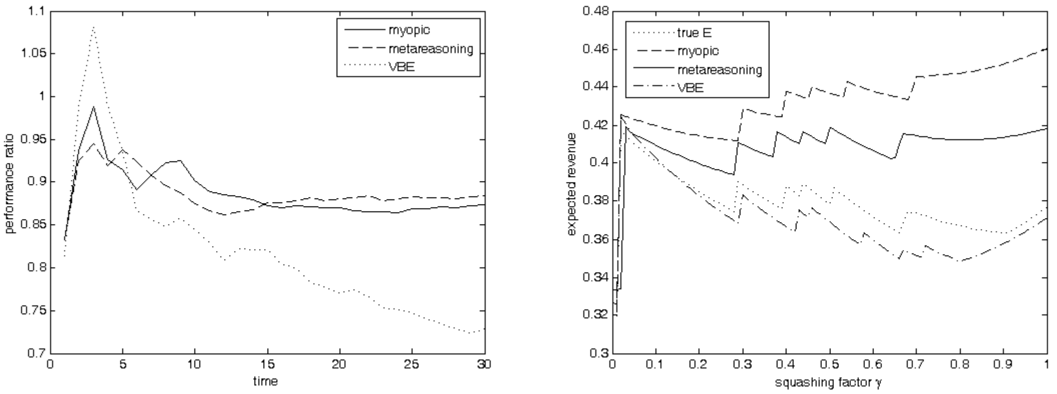

Figure 5.4: **(a)** Performance Ratios for VBE, metareasoning and myopic algorithms **(b)** VBE, metareasoning and myopic "world views" at $t = 30$

While the revenue properties of the VBE algorithm are lackluster, and expectedly so, of greater interest is its use as a benchmark for information convergence properties. Figure 5.4b demonstrates the improved informational properties of the VBE algorithm at time $t = 30$, again, for a representative example set of click-through-rates and values. All 3 algorithms have properly captured in some sense the correct pattern of peaks and troughs. The graph from the VBE estimates is markedly more accurate than that of the other two algorithms, but the metareasoning algorithm results in estimates and a revenue graph that is again significantly more accurate than the myopic algorithm. It is striking that the VBE algorithm maintains a high degree of accuracy even when γ is high and the effect of accurate (or inaccurate) CTR estimates is most sharply felt.

Although the metareasoning estimates are markedly more accurate than those of the myopic algorithm, they are still significantly different from true estimates, and fall dramatically short of the accuracy of even the simple variance based exploration heuristic. The two parts of figure 5.4 indicate that value of information based metareasoning, by focusing primarily on revenue and changes in revenue effected by changes in information state, falls significantly short of optimal or even greedy exploitation in information quality, but reaches a state of sufficiently accurate information to obtain desirable revenue properties.

5.5 Click and Impression priors

There are a number of logical and intuitive choices for priors, that is, initial counts for clicks and impressions. Perhaps the most natural is 0 clicks and 0 impressions, as this reflects the true state of knowledge – the provider has neither shown any ads nor received any clicks from these ads. However, in this case, the standard error is infinite, making it difficult to model uncertainty using the method described in 3.2.2. We can loosely categorize the possible types of priors as "pessimistic" to "optimistic" and "ephemeral" to "persisting", where "pessimistic" means the prior assigns low click-through-rate estimates to all advertisements (low click initialization), and "optimistic" means high click-through-rate estimates, while "ephemeral" means that the effect of the prior will quickly be outstripped by new clicks and impressions, and "persistent" means that the effect of the prior initializations will be felt for a number of time steps. The following table makes this somewhat clearer by giving examples of possible initial values c and i for each "type" of prior

	Ephemeral	$\cdots$	Persistent
Pessimistic	$c = 0, i = 1$	$c = 0, i = 5$	$c = 0, i = 10$
$\vdots$	$c = 1, i = 2$	$c = 2, i = 5$	$c = 5, i = 10$
Optimistic	$c = 1, i = 1$	$c = 5, i = 5$	$c = 10, i = 10$

In the balance of our simulations, the priors were optimistic and ephemeral.

Of particular interest is the intermediate ephemeral prior, $c = 1$, $i = 2$, as this prior maximizes standard error (short infinity). In this sense, this prior captures the idea that we have maximum uncertainty at $t = 0$. This allows for the greatest amount of randomness in uncertainty simulation and thus spans the largest possible set of alternative information states. Intuitively, this should allow for the most dynamic and natural exploration of uncertain click-through-rates.

We found through simulation with several different combinations of priors that changing priors does not drastically change performance. The performance ratio curves for both algorithms were extremely similar for other priors, except that for persistent priors, the metareasoning algorithm generally took longer to overtake the myopic algorithm, as it takes more time periods for the learned click-through-rates to dilute the effect of the priors. This was particularly the case for the optimistic prior, since a large number of initial clicks plays a disproportionately large effect in determining CTR estimates in early stages. In addition, we found that modifying prior counts for clicks and impressions does not affect the near-optimal performance of early time periods.

5.6 True distributions and value/click-through-rate correlation

The tests above have used the same beta and lognormal marginal distributions and Gaussian copula for generating true click-through-rates and value, and additionally have all used Spearman correlation -0.5 between CTR and value. This is akin to choosing a single keyword to model on, as the true distribution of CTRs and values will likely vary significantly, between keywords. For example, it may be the case that for a keyword such as “soda” dominated by a few large brands, value and CTR may be highly correlated in an extremely top heavy distribution, where the top advertisements have both high value and high CTR (Coca-Cola and Pepsi, for example), whereas the rest of the advertisements have both low value and low CTRs. Conversely, it is possible for a keyword purchased by a large number of homogeneous advertisers (such as printer paper, for example) to have uncorrelated and uniformly distributed values and CTRs. While it is not easy to predict a priori the distribution and correlation structure of any given keyword, it is clear that it is possible for a wide range of different keyword-specific distributional structures to exist.

The marginal distributions used in our simulations were taken from Lahaie and Pennock, and are reasonable models of true Yahoo! data for a specific high-activity keyword [14]. In order to verify that our results are not simply due to our particular choice of parameters and distributions, we also repeated the simulation for uniformly distributed, uncorrelated value and click-through-rates, and obtained the comparable results shown in figure 5.5.

The success of value of information based metareasoning with this extremely general set of values and click-through-rates is indicative of its applicability to a wide range of keywords with different particular relationships between value and CTR.

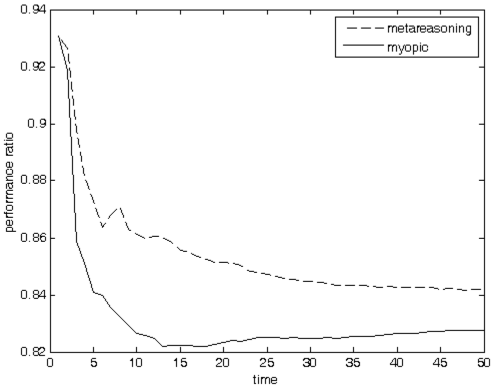

Figure 5.5: Performance ratio over time for uniformly distributed and uncorrelated values and CTRs

5.7 Information Density

As discussed in section 4.3.1 the information density of the environment plays a significant role in determining the potential benefit of metareasoning. In an extremely information rich environment, the myopic algorithm is able to explore sufficiently merely as a byproduct of greedily maximizing expected revenue, as each allocation offers a large number of impressions to all slots and allows for rapid refinement of estimates, even without active consideration of the information or exploration properties of an action. Above, we were in a moderate informations setting, where the top advertisement receives 50 effective impressions per time period, with the number of impressions per slot exponentially decreasing for lower slots. Figure 5.6a shows performance with 250 effective impressions per time period, a high information setting, and demonstrates that the performance of the metareasoning algorithm deteriorates in comparison to the myopic algorithm in this environment. Here the relative value of forgoing current revenue for better information is lower, as no matter what allocation is made, a great deal of information is obtained simply due to the large number of impressions.

It is somewhat surprising that both the metareasoning and myopic algorithms do more poorly in a high information setting than the moderate information setting we have typically explored. Both algorithms stabilize at around 88% performance ratio. This can be attributed in large part to the secondary effect of the way in which we chose to implement or model information density. By using the number of effective impressions per time period as the tuneable parameter information richness, we also introduce the secondary effect that any allocative mistakes are multiplied in their effect on revenue. If one algorithm's allocation had, for example, swapped the top two advertisements in the omniscient allocation, they would lose revenue for two reasons, first, the more valuable advertisement would receive fewer clicks, and second, the advertisement in the top slots per-click payment would be mispriced. The magnitude of the former of these effects does not depend on the number of effective advertisements and total number of clicks, but the latter does. The effect of these per-click mispricings is exacerbated in our high information setting by the increased number of effective impressions and clicks each time period.

We also investigated the performance of the two algorithms in a low information case, where only 10 effective impressions were allocated to the top slot each time period. In this setting, neither algorithm is able to learn effectively, as, with so few effective impressions per time period, estimates are not refined drastically enough from time period to time period.

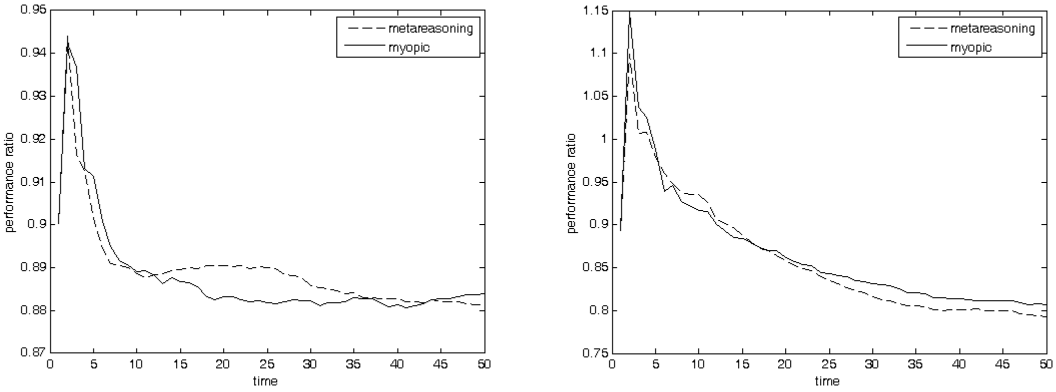

Figure 5.6: **(a)** Performance ratios in information rich environment **(b)** Performance ratios in information poor environment

We see that the performance of a value of information based metareasoning algorithm is, unsurprisingly, highly sensitive to the information structure of the domain. Both too much and too little information are highly deleterious to the performance of our metareasoning algorithm.

The primary result of our simulations is captured in the tail end of the performance graphs in figure 5.1: the improved revenue properties of the metareasoning algorithm in comparison to a greedy revenue maximizing heuristic. Examining click-through-rate estimates and perceived revenue curves over time partially informs the better performance of the metareasoning algorithm. Although slower to converge to accurate estimates of true CTRs than a greedily exploring heuristic, the metareasoning algorithm leverages value of information to focus learning on valuable advertisements and more quickly converge to accurate click-through-rate estimates that inform valuable squashing factor selections and allocations. Parameter tuning revealed that the information landscape of the sponsored search auction, which may vary from keyword to keyword, is critical in determining the success and utility of metareasoning. Our results demonstrate that while there is certainly significant potential in increased revenue to be gained from metareasoning, careful consideration of keyword or auction-specific properties is crucial.

Chapter 6

Discussion and Conclusion

We have investigated the sponsored search advertising industry, a rapidly growing segment of Internet marketing that is particularly interesting due to its tremendous growth potential and because of its unique history in adopting and refining auction mechanisms as the dominant method of sale. This represents both a striking divergence from traditional methods of advertising, both online and otherwise, and is a success story of applying dynamic, customer-based pricing to a complex sale. Following Lahaie and Pennock, we extended traditional auction mechanisms to a squashing factor family of allocation mechanisms, which allows the sponsored search provider to set a variety of different allocations and payments useful for both revenue and information properties [14]. Using an simple and elegant yet powerful and representative game theoretic model due to Varian, we were able to characterize equilibrium bidding strategies of the advertisers, and thus were able to calculate expected equilibrium revenue to the sponsored search provider [24]. Our world is the extension of this game to a multiple-period model, with agents playing a specific equilibrium policy, and the provider possessing uncertain and noisy information about the advertisers. We developed a model for reasoning about this uncertainty through simulation, and provided means for greedily maximizing perceived revenue or perceived information gain. We then adapted a metareasoning framework to rationally deliberate about sacrificing present revenue in favor of improved information and higher future revenues. In empirical simulation, we demonstrated the value of metareasoning and analyzed revenue, information and convergence properties under a varying set of modeling parameters, including information richness, prior estimates and true underlying CTR and value distributions. Most importantly, we found that the information landscape of the domain plays a significant determining role in the value of metareasoning, and should be the first and most important consideration of any attempt to apply a value of information metareasoning process to sponsored search auctions.

6.1 Application Potential

It is somewhat difficult to discuss the application potential of a metareasoning algorithm to current incarnations of sponsored search auctions, since the two dominant forms of sponsored search, Google's AdWords and Yahoo's Overture system both use systems more complex than simple click-through-rate estimation. Google leverages the strength of its search algorithms to tailor a relevance score that is both a proxy for and an extension to click-through-rate estimates. The game theoretic framework employed in our analysis is an instructive one, concerned primarily with analytical tractability and descriptive elegance rather than faithfulness to reality. That said, the performance improvements due to metareasoning and careful consideration of value of information are non-trivial. Utilizing the principles of metareasoning and carefully learning about advertisements is of great use to sponsored search providers. Rather than treating relevance and click-through-rate estimation as primarily a search problem, there is a certain value in regarding it as a joint search and statistical sampling and machine learning problem.

An actual implementation of value-of-information based metareasoning, even beyond the simple algorithms presented here, requires the careful consideration and analysis of the real-world barriers and impediments to such an effort.

6.1.1 Advertiser and user barriers

The rank-by-revenue and rank-by-bid allocation mechanisms have received a great deal of attention and usage in industry because of their simplicity and elegance. There is a significant barrier to entry arising from the difficulty in garnering support and acceptance of the more complex and opaque mechanisms of the squashing factor family of allocation rules. Indeed, this reluctance to adopt or accept a novel set of mechanisms is not mere conservatism or technophobia. Lahaie and Pennock have done an in-depth analysis of the effect and potential of using the squashing factor family of allocation rules, and found that, while intermediate choices of squashing factor may be revenue-optimal for the sponsored search provider, these optimal squashing factors may not result the most efficient or relevant allocations, where efficiency is measured by total revenue to both the provider and the advertisers, and relevance is measured by the total effective click-through-rate of the allocation [14]. Since efficiency and relevance are the primary metrics determining user satisfaction for both the advertisers and the end-user to whom the advertisements are displayed, deviating from optimal levels of efficiency and relevance is likely to foment discontent and lead to customer attrition.

Lahaie and Pennock propose one solution of setting acceptable thresholds for efficiency and relevance loss and maximizing revenue with respect to these constraints [14]. It is unclear how much these constraints would hamper the ability of the metareasoning algorithm to adequately explore. Certainly by limiting the possible interval of squashing factors to a narrower band, the ability of the VBE algorithm to explore is severely impaired. The fact that metareasoning similarly outperforms the myopic algorithm in both the wider [-2,2] interval and the narrower [0,1] interval is heartening evidence that metareasoning is able to do well in a narrower band, but as the number of alternative allocations and squashing factors usable by metareasoning decreases, the performance gap between the two algorithms will close as well.

6.1.2 Real information density

We have shown that the performance of metareasoning deteriorates in extreme information landscapes, where exploration either becomes moot or impossible. Unfortunately, we cannot tune the information landscape of the real world. Ensuring a constant and uniform number of impressions per time period can be achieved somewhat tautologically by defining the length of a time period with respect to the number impressions accrued. However, the non-uniformity of time period length in real time is jarring for advertisers, resulting in customer dissatisfaction as well as lagged or suboptimal bid adjustments. This process would require not only that advertisers be introduced to a novel and somewhat complex set of allocation mechanisms, but also be told that the relevant and active allocation mechanism will be consistently changing, and further, that it will be changing at irregular, unpredictable and non-transparent intervals. The assumption that advertisers will be able to instantly adapt to their equilibrium bids becomes even more troublesome in this case.

Aside from the problems introduced by impression-based time period definition, information density is also determined by the attention decay rate, and the number of effective impressions received by advertisements in lower slots. This is also likely to be quite different from keyword to keyword, and, while taken as exogenously given in our model, must be learned by the sponsored search provider.

Any attempt to apply a value of information based metareasoning process to sponsored search auctions must carefully evaluate and understand the information properties of the keywords to which it will be applied, since it is likely to be the primary factor determining the success or failure of metareasoning.

6.2 Future Work and Extensions

The applications of value of information based metareasoning and the general adaptive auction framework presented here are merely first tentative forays into applying this type of dynamic control algorithms to the sponsored search setting. There are a host of further refinements and extensions improving the performance of the algorithms and adapting models to more realistic settings.

6.2.1 Dynamic algorithm swapping

The weaknesses of the VBE and myopic algorithms and their complementary strengths naturally raise the idea of switching from exploration to exploitation after a certain level of estimate accuracy has been reached. This hybrid algorithm would result in rapid exploration in early time periods, followed by greedy revenue maximization when estimates are accurate and the myopic algorithm is likely to be most successful. This approach leverages the respective strengths of the myopic and VBE algorithms while minimizing the effect of their weaknesses. By exploring when uncertain and exploiting when confident of our knowledge, this algorithm represents perhaps the most natural solution to balancing exploitation and exploration: explore until we know enough, and then exploit.

The primary obstacle facing this approach is the determination of when to flick the switch from exploring to exploiting. Optimally, we would like to switch from exploring to exploiting as soon as the perceived optimal squashing factor and allocation with respect to current estimates is the same as the true optimal squashing factor and allocation. This is difficult to determine a priori, however, as the algorithm does not have access to the omniscient designers decision. This process suggests a further level of meta-control, where at each time step, a meta-algorithm selects not which ranking mechanism and allocation to execute, but rather which algorithm, perhaps among value of information metareasoning algorithm, myopic revenue maximization, VBE or a number of others, to execute, which in turn selects an allocation.

Such an effort is a first step towards reconciling the heuristic appeal of VBE and myopic revenue maximizing algorithms and their obvious and glaring weaknesses. The complementary strengths and weaknesses of the algorithms presented here naturally invites an attempt to marginalize their weaknesses and capitalize on their strengths by utilizing each when most appropriate.

6.2.2 Multi-Armed Bandit extensions and Exp3

An alternative method for addressing the weaknesses of VBE and myopic exploration draws its inspiration from one of the "probability matching" algorithms developed for multi-armed-bandit problems, the **Exp3** algorithm of Auer et al [1]. In the MAB setting, Exp3 associates scores with each arm, at each time period randomly choosing an arm weighted by a probability that is a combination of their relative weight vs. all arms and a random exploration factor. The weight of the arm chosen is then updated by scaling by an exponential factor of the normalized reward received [1].

The Exp3 algorithm is appealing for our domain because it represents one of the simpler MAB algorithms that meaningfully balances exploration and exploitation and its weighting mechanism is intuitively and simply applicable to our conception of arms as advertisements. However, since we do not directly "pull" arms in our sponsored search auction setting, extension of Exp3 requires careful consideration of how to map a set of weights on arms to a set of probabilities over squashing factors, which are the actions considered by the provider. One method is to calculate the probability score of each advertisement as in Exp3, and to weight each potential squashing factor by the sum of the scores of the advertisements that would appear, weighted by position.

Our first attempts at implementing this extension to Exp3 proved fruitless, as the way in which advertisement probability scores were aggregated to generate probability scores for squashing factors did not correctly maintain the probability properties that should be associated with arms in the traditional bandit setting. In addition, the revenue received by an advertisement varies from time period to time period, depending on the allocation and squashing factor chosen. This means that learning the reward distribution underlying an advertisement arm is not a simple process of learning the CTR parameter to a Bernoulli distribution, but is rather far more complex and may not be parameterizable.

6.2.3 Time dependent values and click-through-rates

In this work, we have assumed a constant set of values and click-through-rates for advertisers, so that their internal valuations and the innate quality or relevance of their advertisements does not change over time, and only their equilibrium bids change. In reality, advertiser-specific values are likely to change as their marketing requirements evolve over time. As a very simple but illustrative case, AdWords allows advertisers to specify a per-day budget for advertising. After this budget is exhausted, the advertisers advertisement will no longer be displayed. This is essentially a case where the advertisers valuation drops to

0. Alternatively, in a case where a company is launching a new product, they are likely to intially have a very high valuation for advertising as market share and brand recognition are established. Time dependent valuations are interesting because of their clear real life motivation and revenue implications, but does not present addition direct problems to learning, as our game theoretic analysis assumes that all valuations are publically known.

Advertiser specific click-through-rates are also likely to change over time. In a simple case, assume some constant number of searchers who query a certain keyword, and assume that no one will ever click the same advertisement twice. In this case, the click-through-rates of all advertisers will decrease over time, with the advertisers in the top slots decreasing more quickly, as they begin to exhaust the "pool" of clicks available. There are also a large number of exogenous factors, news reports, brand awareness drives, etc., that can cause drastic shifts in the appeal and relevance of a given company's advertisement. Time dependent click-through-rates means that metareasoning and exploration is more important, as we never reach a plateau of extremely accurate estimates, because true CTRs become a constantly moving target. However, it also means that a value information based metareasoning is more difficult, as the perceived gains from setting a certain allocation may not come to fruition, or that a great deal of importance was attributed to learning the click-through-rate of an advertisement that suddenly becomes irrelevant due to changes in its true click-through-rate or value. Dynamically changing valuations could also lead to this latter problem, where, after significant effort is invested in learning about a specific advertisement with the intention of continued exploitation, only to have its value change.

An extension of this work to allow for some of the time dependent dynamism so characteristic of and essential to Internet advertising goes far in capturing some of the complexity of online marketing abstracted away in this work.

6.2.4 Advertiser Behavior

We have assumed throughout, following Lahaie and Pennock and Varian, that advertisers always bid the lower recursive solution to the symmetric Nash equilibrium [14, 24]. Introducing dynamic advertiser bidding agents using heuristic or behavioral strategies would be an extremely interesting extension in the multi-stage auction system[1]. Our current metareasoning process relies heavily on equilibrium analysis to calculate expected revenue, thus introducing heuristic bidding strategies would require careful re-evaluation of the value of information calculations and learning methods. Modeling advertisers as adaptive learning agents places the system in a stochastic game setting where advertisers and auctioneer are

[1] Acknowledgements are owed to Ivo Parashkevov and Jie Tang for illuminating discussion on this topic.

learning in parallel about each other and responding to each others play. While this is certainly a very interesting and generalizing extension of the multi-period sponsored search auction, we anticipate that as agent strategies become more complex and equilibrium analysis becomes more difficult, value of information based metareasoning becomes less feasible, and more traditional forms of game theoretic or statistical intertemporal learning may be applied.

6.3 Concluding Remarks

The application of the consideration of the value of information and a rational decision-making process drawing on this calculation to a sponsored search setting is interesting both as an academic and practical venture. While we would certainly regard attempting to directly apply the metareasoning algorithms presented here to a real sponsored search auction as premature, it is clear that there is significant untapped potential in considering the value of information gained from sponsored search auctions, and taking this into consideration when generating allocations. The application of metareasoning to this domain demonstrates that it is possible and indeed, quite useful to consider utilizing this form of rational decision-making in domains to which it does not immediately seem perfectly well suited. There is potentially a significant and valuable role for value of information based metareasoning as a control algorithm for adaptive sponsored search auctions, one that should receive future attention from both industry and academia.

Bibliography

[1] Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2003.

[2] Dirk Bergemann and Jusso Välimäki. Bandit problems. Technical Report 1551, Yale University Cowles Foundation Discussion Papers, January 2006.

[3] Mark S. Boddy. Solving time-dependent problems: A decision-theoretic approach to planning in dynamic environments. Technical Report CS-91-06, Brown University, 1991.

[4] Rick E. Brunner. The decade in online advertising, 1994-2004. *DoubleClick Online Research*, July 2005.

[5] Thomas Dean. Decision-theoretic control of inference for time-critical applications. Technical report, Brown University, Providence, RI, USA, 1989.

[6] Benjamin Edelman, Michael Ostrovsky, and Michael Schwarz. Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords. *NBER Working Papers Series*, 2005.

[7] John C. Gittins and David M. Jones. A dynamic allocation index for the discounted multiarmed bandit problem. *Biometrika*, 66(3), 1979.

[8] Irving J. Good. *Good Thinking*. University of Minnesota Press, Minneapolis, Minnesota, 1976.

[9] David E. Heckerman, Eric J. Horvitz, and Blackford Middleton. An approximate nonmyopic computation for value of information. *IEEE Transactions on Patterns Analysis and Machine Intelligence*, 15:292–298, 1993.

[10] David E. Heckerman, Eric J. Horvitz, and Bharat N. Nathwani. Toward normative expert systems: The Pathfinder project. *Methods of Information in Medicine*, 31(2):90–105, June 1992.

[11] Ronald A. Howard. Information value theory. *IEEE Transactions on Systems Science and Cybernetics*, SCC-2:22–26, 1966.

[12] Leslie P. Kaelbling. *Learning in Embedded Systems.* The MIT Press, May 1993.

[13] Michael N. Katehakis and Jr. Arthur F. Veinott. The multi-armed bandit problem: decomposition and computation. *Mathematics of Operations Research*, 12(2), 1987.

[14] Sébastien Lahaie and David M. Pennock. Revenue analysis of a family of ranking rules for keyword auctions. In *ACM Conference on Electronic Commerce 2007*, July 2007.

[15] David McAdams and Michael Schwarz. Who pays when auction rules are bent? *MIT Sloan Working Papers Series*, 4607-06, June 2006.

[16] Nielsen//NetRatings. Nielsen//Netratings 2007 online advertising report, February 2007.

[17] Herbert Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58:527–535, 1952.

[18] Michael Rothschild. A two-armed bandit theory of asset pricing. *Journal of Economic Theory*, 9:185–202, 1974.

[19] Stuart J. Russell, Devika Subramanian, and Ronald Parr. Provably bounded optimal agents. In Ruzena Bajcsy, editor, *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence (IJCAI-93)*, pages 338–344, Chambéry, France, 29– 3 1993. Morgan Kaufmann publishers Inc.: San Mateo, CA, USA.

[20] Stuart J. Russell and Eric Wefald. Principles of metareasoning. In Ronald J. Brachman, Hector J. Levesque, and Raymond Reiter, editors, *KR'89: Principles of Knowledge Representation and Reasoning*, pages 400–411. Morgan Kaufmann, San Mateo, California, 1989.

[21] Roy Schwedelson and Jay Schwedelson. The history of Internet advertising. *DM News*, April 1997.

[22] Herbert Simon. *Models of Bounded Rationality.* MIT Press, Cambridge, Massachusetts, 1982.

[23] Louise Story. An ad upstart challenges Google. *The New York Times*, February 2007.

[24] Hal R. Varian. Position auctions. Unpublished Article, 2006.

[25] Joannès Vermorel and Mehryar Mohri. Multi-armed bandit algorithms and empirical evaluation. In *ECML*, pages 437–448, 2005.

[26] John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, New Jersey, 1947.

[27] William Walsh, David C. Parkes, and Rjarshi Das. Choosing samples to compute heuristic-strategy Nash equilibrium. In *Proceedings of the Fifth Workshop on Agent-Mediated Electronic Commerce*, 2003.