

Optimal Ad Ranking for Profit Maximization

Raju Balakrishnan
rajub@asu.edu

Subbarao Kambhampati
rao@asu.edu

School of Computing and Informatics
Arizona State University
Tempe AZ 85287

ABSTRACT

Despite the enormous commercial importance of on-line advertisements (ads), there has been little work done to clarify the basis for ranking and displaying them. Most existing methods rank ads as if the user views each of them in isolation. We will consider a more realistic user model that induces three mutual influences between displayed ads: (i) positional bias (for viewing ads placed higher up) (ii) similar ad fatigue (which reduces interest in an ad when similar ads have already been displayed above it) and (iii) browsing impatience (which accounts for the user abandoning the ad viewing based on the ads already seen). We will show that in general, when the inter-ad similarity is taken into account, optimal ranking is NP-hard. Ignoring inter-ad similarity, we state and prove the optimal ranking function for sorting the ads that is sensitive to the other two factors. We will show that the known ad ranking strategies correspond to restricted special cases of our ranking function. We also provide simulation studies that establish the effectiveness of its generality.

Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: Online Information Services

1. INTRODUCTION

Online keyword advertisements (ads) are a multi-billion dollar business. The basic idea is to show a ranked list of advertisements to a user searching for related information. Individual vendors bid for certain keywords. When these keywords are seen to be relevant to the user's search, the search engines will show, typically in a separate pane, a list of these ads. If the user clicks on any of the ads that are shown, then the search engine gets the bid amount as the revenue.

Confirmed details of the exact ranking strategies used by the various search engines are of course hard to get. Yahoo!

is said to rank its ads in terms of their bid amount;¹ while Google and Microsoft are said to rank the ads in terms of the expected profit [11, 1], assessed as product of relevance and bid amount. There is little work done towards explaining which models are optimal under what assumptions.

The few existing academic efforts on ad ranking all implicitly assume that the list of displayed ads can be decided by selecting individual ads in isolation (see Section 2). Unfortunately, such a ranking is insensitive to the mutual influences between the displayed ads induced by the typical page browsing patterns of users (c.f. [4, 5, 3]):

Positional Bias: There is a bias for viewing ads positioned higher up in the list.

Similar Ad Fatigue: The perceived relevance of an ad is reduced by the presence of similar ads higher up in the list [9].

Browsing Abandonment: The user is not compelled to view all the ads and may abandon browsing at any point of time either because of impatience or satiation. The abandonment likelihood may thus depend on the (number and content of the) ads already viewed.

In this paper, we aim to investigate the optimal ranking of ads, that is sensitive to the mutual influences between displayed ads induced by the user browsing patterns. We start by showing that optimal ranking to maximize expected profit considering similar ad fatigue is NP-hard. For the case where similarities between the displayed ads can be ignored, we provide a simple ranking function for displaying ads that correctly combines the bid amount, ad relevance and abandonment probabilities. After proving the optimality of this ranking, we will show how it subsumes the ranking strategies said to be used by Yahoo!, Google and Microsoft. Our analysis shows these latter ranking strategies to be optimal only under significantly restrictive assumptions. Finally, to gain an understanding of the practical impact of the ranking, we provide a simulation study that compares the proposed strategy to the Google and Yahoo! models.

While the immediate motivation for our work is ad ranking, we believe that the methodology can also be adapted to ranking in other e-commerce scenarios such as vendor recommendations.

2. RELATED WORK

¹This strategy was adapted by Yahoo! from Overture, but said to have changed in the past year.

Ranking is an extensively researched problem in IR. While there has been some work in the IR community on browsing model sensitive ranking of search results, it is not directly applicable to ad ranking because the latter is influenced by the competing pulls of ad relevance and ad profit. Robertson [12] claimed that a retrieval order based on Probability Ranking Principle (PRP) leads to the largest number of relevant documents in a set of retrieved documents than any other policy. Later Gordon and Lenk [7, 8] identified the required assumptions for the optimality of the ranking according to PRP.

The ad placement problem has attracted considerable research, mainly due to the popularity of search engine advertisements. Feng *et al.* [6] compare the ad-placement strategies of different search engines. Richardson *et al.* [11] tackle the problem of predicting the click-through rates for rarely clicked ads and Clarke *et al.* [2] examine the influence of caption features on click-through rates of the ads. Mehta *et al.* [10] propose a placement of ads that also considers the budget limits of each bidder and the bid amount of ads, formulating the problem as an online bipartite matching problem. All of these efforts ignore the mutual influence between ads induced by the user browsing patterns.

Craswell *et al.* [3] empirically evaluated different models of user search result browsing to analyze the positional bias of click-through rates. They found that the cascade model—a model assuming linear traversal through the results ending with a clicked result—predicts the observed bias accurately. We adapt this model with extensions for our work.

3. PROBLEM SETUP

Our aim is to rank the ads to maximize expected profit. The ranked placement of ads is denoted as an ordered set A , $A = \langle a_1, a_2, \dots, a_n \rangle$, where the subscripts denote the position of a_i in the ranked list starting from the highest ranked result first. The profit of ad a_i is denoted by $\$(a_i)$.

We use the basic cascade browsing model suggested by Craswell *et al.* [3] augmented by two extensions. In the basic cascade model the user views search results from top to bottom, deciding whether to click each result before moving to the next. Each ad a_i in the ranked list, is either clicked with probability equal to its relevance $R(a_i)$ or skipped with probability $(1 - R(a_i))$. A user who clicks an ad never comes back, and a user who skips always continues. Clicking on a result means the user must have skipped the results above and decided to click on the result. That is, in this model, probability of clicking a_i ($P_c(a_i)$) is,

$$P_c(a_i) = R(a_i) \prod_{j=1}^{i-1} (1 - R(a_j)) \quad (1)$$

The static distribution of rapidly decreasing view probabilities in eye tracking studies [5], and the dynamic behavior of eye fixations gradually disseminating down the search results [4] are direct implications of cascade model.

Three unrealistic assumptions in this model, as partly acknowledged by Craswell *et al.* [3] are:

- A1. The user keeps going down until he clicks a result. (In reality, user may abandon checking the search results without finding the required results.)
- A2. The click probability of an ad is dependent only on its position and is independent of the ads above it.

(In reality, the presence of similar ads in the higher position reduces the click probability.)

- A3. The user will always click only one document. (In reality, multiple clicks are commonly observed in ads and search.)

We base our analysis on a model that relaxes all three assumptions. To relax A1, we allow for the user abandoning search without finding a relevant result. The abandonment probability, $\gamma(a_i)$, is modeled as a function of the result a_i at the position. Incorporating this probability, Equation 1 changes as,

$$P_c(a_i) = R(a_i) \prod_{j=1}^{i-1} [1 - (R(a_j) + \gamma(a_j))] \quad (2)$$

Since the abandonment probability and number of ads clicked may vary depending on the nature of the application, we make no assumptions about the nature of the function $\gamma(a_i)$ or the number of selections in deriving the proposed ranking strategy.²

To relax A2, we generalize the relevance function $R(a_i)$ to include the mutual influence of similar ads on relevance. The relevance of an ad for the user is reduced by a similar ad placed above in the list, and this reduced residual relevance is denoted by R_r . Residual relevance depends on the set of results higher in the ranked list. Formally, the residual relevance $R_r(a_i | \langle a_1, a_2, \dots, a_{i-1} \rangle)$ is the probability of user clicking a_i , after seeing a_i and all the ads above a_i . Substituting residual relevance R_r for relevance function R in Equation 2 gives us context dependent click probabilities. For brevity we denote $R_r(a_i | \langle a_1, a_2, \dots, a_{i-1} \rangle)$ by $R_r(a_i)$ in the rest of this paper.

Using the residual relevance we can rewrite click probability P_c as

$$P_c(a_i) = R_r(a_i) \prod_{j=1}^{i-1} [1 - (R_r(a_j) + \gamma(a_j))] \quad (3)$$

Regarding A3, though we derive our initial ranking model assuming single click, we prove that the proposed ranking is also optimal for multiple clicks.

The user model may be schematically represented as a flow graph as shown in Figure 1. Labels on the edges refer to the probability of the user traversing them. Each vertex in the figure corresponds to a view epoch (see below), and the flow balance holds at each vertex. Starting from the top ad, the probability of the user clicking the first ad is $R(a_1)$ and probability of him abandoning browsing is $\gamma(a_1)$. The user goes beyond the first ad with probability $1 - (R(a_1) + \gamma(a_1))$ and so on for the subsequent results.

4. OPTIMAL AD RANKING

Using the notations introduced, we formally define the ranking problem and derive optimal ranking in this section. The formal problem statement is,

Choose the optimal ranking $A_{opt} = \langle a_1, a_2, \dots, a_N \rangle$ of N

²Other than the assumption that the probability of user leaving the search results at a position is independent of the results above.

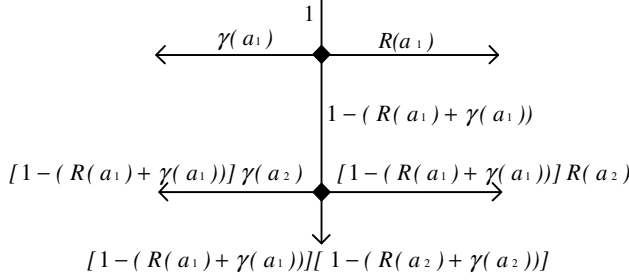


Figure 1: Flow Graph for user browsing the first two ads. The labels are the view probabilities and a_i denotes the ad at the i^{th} position

ads to maximize the expected profit

$$E(\$) = \sum_{i=1}^N \$ (a_i) P_c(a_i) \quad (4)$$

where N is the total number of ads to be ranked.

Substituting click probability P_c from Equation 2 in Equation 4 we get,

$$E(\$) = \sum_{i=1}^N \$ (a_i) R_r(a_i) \prod_{j=1}^{i-1} [1 - (R_r(a_j) + \gamma(a_j))] \quad (5)$$

Note that we use residual relevance R_r here instead of absolute relevance R .

Now we consider the optimal ranking of ads to maximize the objective function in Equation 5. Unfortunately, optimal ranking considering residual relevance turns out to be NP-Hard as proved below.

THEOREM 1. *Ranking to optimize expected profit in Equation 5 considering inter-ad similarities is NP-Hard.*

The proof is by reduction from the independent set problem. See Appendix A-1 for complete proof.

We now focus on optimal placement ignoring similarity. Replacing R_r by R , Equation 5 becomes,

$$E(\$) = \sum_{i=1}^N \$ (a_i) R(a_i) \prod_{j=1}^{i-1} [1 - (R(a_j) + \gamma(a_j))] \quad (6)$$

The optimal ranking maximizing this expected profit can be shown to be a sorting problem with a special ranking function:

THEOREM 2. *The expected profit in Equation 6 is maximum if the ads are placed in descending order of the value of the ranking function RF ,*

$$RF(a_i) = \frac{\$(a_i) R(a_i)}{R(a_i) + \gamma(a_i)} \quad (7)$$

Proof Sketch: The proof is by induction on the number of ads in the set of ads to be ranked. We consider two placements P_1 and P_2 , where P_1 is the placement according to the proposed strategy and P_2 is a competing placement. We formulate the difference of expected profit of P_2 from expected profit of P_1 as $(E(\$ (P_1)) - E(\$ (P_2)))$. We show that this difference is always greater than or equal to zero, based on

inductive hypothesis and on the fact that P_1 is ordered in descending order of RF . See Appendix A-2 for complete proof. \square

RF can be understood as the profit generated per unit click probability consumed by the ad. Referring to Figure 1, this ordering is intuitive since the top ads in the ranked list have more view probabilities, and placing ads with higher profit per consumed view probability is likely to increase profit.

The expected profit in Theorem 2 considers ranking for single click case, and Theorem 3 extends the results to multiple clicks.

THEOREM 3. *The expected profit is maximum even for multiple clicks if the ads are placed in descending order of value of RF as proposed in Theorem 2.*

Proof Sketch: We proved that ordering according to RF provides maximum expected profit for single click. Multiple clicks are the same as the user restarting her browsing from the result immediately below the last clicked ad. A simple induction on number of clicks based on this idea, using single click as base case, is sufficient to prove that the proposed placement provides maximum expected profit for multiple clicks. See Appendix A-3 for the complete proof. \square

Mining the parameters for RF : Since the existing ad ranking strategies already have to mine the ad relevance, the only additional parameter needed to implement RF is the abandonment probability $\gamma(a_i)$. These can be mined from click logs, similar to click through rates [11]. For example, assuming a single click model, the abandonment probability of the first ad a_1 is,

$$\gamma(a_1) = 1 - \left[\frac{OCP_2(a_2)}{R(a_2)} + R(a_1) \right]$$

where $OCP_2(a_2)$ is the observed click probability on a_2 when a_2 is displayed in the second position, i.e. if the ad a is displayed 100 times in the second position and is clicked 5 times, $OCP_2(a)$ is $\frac{5}{100}$. As the computation progresses, the abandonment probabilities will converge to unique values irrespective of the position used to compute OCP . Similarly, the abandonment probabilities of the ads in the lower positions can be mined once we know the abandonment probabilities and relevance of the ads above. Initially, the abandonment probabilities can be calculated for all the ads placed in the top positions. The ads in the positions below can be used to mine probabilities—knowing abandon probabilities of ads above—as computation progresses.

5. ANALYZING EXISTING STRATEGIES

In this section we compare the existing ad placement strategies to the ranking function RF in Equation 7. We will see that they all correspond to specific assumptions on the abandonment probability $\gamma(a_i)$.

Special Case 1: Ranking by Bid Amount: If we assume that the user never abandons browsing (i.e., $\forall_i \gamma(a_i) = 0$), then Equation 7 reduces to $RF(a_i) = \$ (a_i)$

This means that the ads are ranked purely in terms of their bid amount, a strategy that was attributed to Yahoo! (and Overture) until last year.

When $\gamma(a_i) = 0$, we essentially have a user with infinite patience and will keep browsing downwards until he finds the

relevant ad. So, to maximize profit, it makes perfect sense to rank ads by bid amount. More generally, for small abandonment probabilities, ranking by bid amount (the erstwhile Yahoo! placement) is near optimal.

Special Case 2: Ranking by Expected Bid Amount:

Another approximation to $\gamma(a_i)$ is to assume that it is negatively proportional to the relevance of the ad a_i —the more relevant the current result, the less likely the user is to abandon the search. Specifically, if we have $\forall_i \gamma(a_i) \approx k - R(a_i)$ (for some constant k between 0 and 1), then the Equation 7 reduces to,

$$RF(a_i) \approx \frac{\$(a_i)R(a_i)}{k} \propto \$(a_i)R(a_i)$$

This shows that ranking ads by their individual expected profit, as is supposedly done by Google and Microsoft, is near optimal as long as abandonment probability is negatively proportional to the relevance.

6. SIMULATION STUDIES

The discussion in the previous section shows that the existing ad placement strategies are optimal only under more restrictive assumptions on the abandonment probabilities. This suggests that the revenues (and profits) may be improved by ranking using RF . To quantify the potential increases in expected profits offered by RF , we performed a large number of simulation runs comparing it with the other two existing strategies. In our experiments we assigned the relevance values as a uniform random number between 0 and α ($0 \leq \alpha \leq 1$) and values of abandonment probabilities as uniform random between 0 and $1 - \alpha$ (This will make sure that $\forall_i (R(a_i) + \gamma(a_i)) \leq 1$). The profits (bid amounts) for ads are assigned uniformly random between \$0 and \$10. Note that uniform random distribution is the maximum entropy probability distribution and makes least assumptions about the distribution of bid amounts. The number of relevant ads (corresponding to the number of bids on a query) is set to 50. Simulated users are made to click between 1 – 10 ads. We performed the simulation for multiple clicks. The number of ads clicked is set as a random number generated in a zipf distribution with exponent 1.5, with ranks increasing from 1 to 10. This is reasonable since a power law is most intuitive for the distribution of the number of clicks.

Simulated users browse down the list. Users click an ad with probability equal to the relevance of the ad and abandon search with a probability equal to the abandonment probability of the ad. The set of ads to be placed is created at random for each run. For the same set of ads, three runs—one with each placement strategy—are performed. For each value of α , 200,000 rounds with each placement strategy are executed.

The results for the three strategies are shown in Figure 2. We see that the proposed strategy (RF) is the winner for all values of α . Ranking by profit strategy loses for low values of α and reaches the optimal value for large values of α , as discussed in Section 5. Note that for $\alpha = 1$ ($\gamma(a_i) = 0$) ranking by profit placement gives exactly the same profit as the optimal placement. The profit from RF exceeds the profit from competing strategy by 40 – 50% for some values of α . For example for $\alpha = 0.3$ $Profit \times Relevance$ (competing strategy) gives an expected profit of \$5.36 while RF gives a profit of \$7.81 (exceeds by 45.7%) and for $\alpha = 0.4$

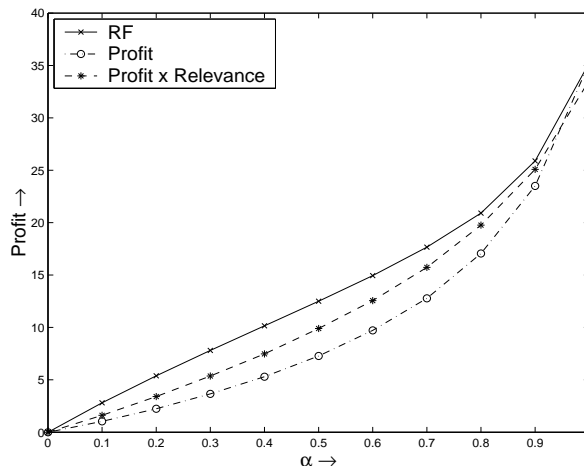


Figure 2: Comparison of proposed placement (RF) with the existing ad placement strategies. Each point corresponds to 2×10^5 simulation runs. Relevance are uniformly random in $[0, \alpha]$ and abandonment probabilities are a uniformly random in $[0, 1 - \alpha]$.

$Profit \times Relevance$ gives a profit of \$7.47, as against \$10.15 by RF (exceeds by 35.9%). This is highly significant for a multi billion dollar market and millions of users.

7. CONCLUSION

This paper addresses the problem of the optimal ranking of ads considering (i) relevance of ads (ii) reduction of click probability of ads due to the other relevant ads above (iii) abandonment of search due to user impatience (iv) reduction of click probability due to similar ads above. We prove that optimal ranking considering the inter-ad similarity (iv) is NP-Hard. Further, we propose an optimal and tractable ad placement strategy considering the other factors—factors (i), (ii), and (iii). We prove optimality of the expected profit from the proposed placement strategy. We compare the existing placement strategies to the proposed strategy and enumerate the assumptions under which the existing strategies are optimal. A simulation study verifies our results and gives an idea of the amount by which the proposed strategy can outperform the existing ones.

In future, we hope to investigate approximation schemes for considering inter-ad similarity. We are also exploring the possibility of validating the ranking model on actual search ad data. Another interesting direction would be to generalize our analysis to consider the advertiser budget limits (c.f. [10]). Finally, although we focused on ad placement, the ranking can be extended to other related e-commerce scenarios such as vendor recommendations (e.g. Amazon’s product recommendations), and search results in e-commerce portals.

Acknowledgements: This research is supported in part by the ONR grant N000140610058 and a 2008 Google research award.

8. REFERENCES

- [1] H. K. Bhargava and J. Feng. Paid placement strategies for internet search engines. In *WWW Proceedings*, pages 117–123. ACM, May 2002.
- [2] C. L. A. Clarke, E. Agichtein, S. Dumais, and R. W. White. The influence of caption features on clickthrough patterns in web search. In *Proceedings of SIGIR*, pages 135–142. ACM, July 2007.
- [3] N. Craswell, O. Zoeter, M. Tayler, and B. Ramsey. An experimental comparison of click position bias models. In *Proceedings of WSDM*, pages 87–94, February 2008.
- [4] Enquiro-Research. The eye tracking example, <http://www.enquiroresearch.com/eye-tracking-examples.aspx>.
- [5] Enquiro-Research. The eye tracking study, <http://www.enquiroresearch.com/images/eyetracking2-sample.pdf>.
- [6] J. Feng, H. K. Bhargava, and D. Pennock. Comparison of allocation rules for paid placement advertising in search engines. In *ICEC Proceedings*, pages 294–299. ACM, May 2003.
- [7] M. G. Gordon and P. Lenk. A utility theory examination of probability ranking principle in information retrieval. In *Journal of American Society of Information Science*, volume 41, pages 703–714, 1991.
- [8] M. G. Gordon and P. Lenk. When is probability ranking principle suboptimal? In *Journal of American Society of Information Science*, volume 42, 1992.
- [9] T. Hernandez and S. Kambhampati. Improving text collection selection techniques with coverage and overlap statistics. In *Proceedings of WWW*, 2005.
- [10] A. Mehta, A. Saberi, U. Vazirani, and V. Vazirani. Adwords and generalized on-line matching. In *ACM Transactions on Computational Logic*, volume 5, August 2007.
- [11] M. Richardson, E. Dominowska, and R. Ragnó. Predicting clicks: Estimating the click-through rate for new ads. In *Proceedings of WWW*, May 2007.
- [12] S. E. Robertson. The probability ranking principle in information retrieval. In *Journal Documentation*, volume 33, pages 294–304, 1977.

APPENDIX

A-1 Proof of Theorem 1

PROOF. Independent set problem can be formulated as a ranking problem considering similarities. Consider an unweighed graph G of n vertices $\{a_1, a_2, \dots, a_n\}$ represented as an adjacency matrix with a value 1 in forward diagonal. This conversion is clearly polynomial time. Now, consider the values in the adjacency matrix as binary similarity values between the ads to be ranked. Let the ads have the same relevance, profitability and abandonment probability. In this set of n ads from $\{a_1, a_2, \dots, a_n\}$, the optimal ranking will have k pairwise independent ads as the top k ads. But the set of k independent ads corresponds to an independent set in graph G . \square

A-2 Proof of Theorem 2

PROOF. Induction on number of ads in the set to be ranked.

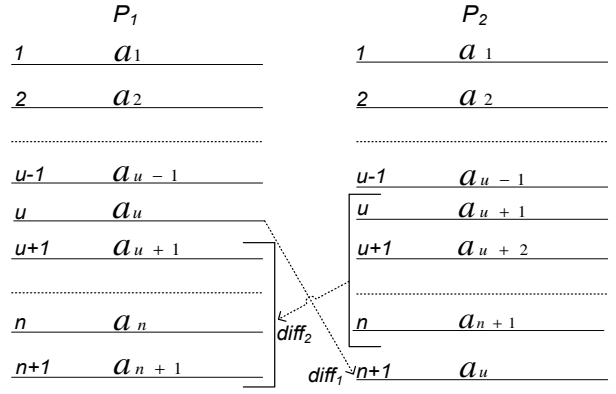


Figure 3: Two ad-placement, P_1 optimal for $n+1$ ads according to the proposed strategy, and P_2 maximal profit ordering based on inductive hypothesis and the $(n+1)^{th}$ newly added ad a_1 out of order. $diff_1$ is decrease in profit due to change in position of a_1 downwards, and $diff_2$ is increase in profit due to raise in position of ads— a_u to $a_{(n+1)}$ —by one in P_2 .

Base Case: single ad, ordering does not matter, true.

Inductive Hypothesis: The profit for a set of n ads is maximized if they are ordered in the proposed order.

Induction: For a set of $n+1$ ads, Consider two sequences of placements

$$P_1 = \langle a_1, a_2, \dots, a_{n+1} \rangle$$

$$P_2 = \langle a_1, a_2, a_{n+1} \rangle$$

Let P_1 be the placement in proposed order and P_2 be the placement out of order. We have to prove $E(\$1) \geq E(\$2)$ where $E(\$1)$ and $E(\$2)$ are expected profits from P_1 and P_2 respectively.

Since P_1 is in the proposed order, $RF(a_{n+1}) \leq \forall_i RF(a_i)$. As our proof is to handle the worst case scenario, we have to prove the hypothesis for the maximum value of $E(\$2)$. Inductive hypothesis says that the profit is maximum for the set of size n if they are in descending order of RF . So for a worst case scenario we assume that the n ads in P_2 are in order of descending RF , and $(n+1)^{th}$ out of order. Since we are having the same set of ads in P_1 and P_2 we name them using the position in P_1 . This placement is shown in Figure 3. Here P_1 and P_2 are same except for the fact that a_u is present in the last position in P_2 . instead of at the u^{th} position in P_1 . The entire set of ads a_{u+1} to a_{n+1} in P_1 is shifted up by one position in P_2 . So we need to prove that the reduction in profit due to change in position in a_u (which is denoted by $diff_1$ in Figure 3) is greater than increase in profit due to lifting of one position for a_{u+1} to a_{n+1} ($diff_2$ in Figure 3). Note that the profit generated by a_1 to a_{u-1} is exactly same for P_1 and P_2 .

We have to prove $diff_1 \geq diff_2$. Let us denote the expected profit of an ad a_i (or a group of ads) from P_1 as $\$1(a_i)$ and expected profit from P_2 by $\$2(a_i)$.

$$\$1(a_u) = R(a_u)\$ (a_u) \prod_{i=1}^{u-1} [1 - (R(a_i) + \gamma(a_i))]$$

$$\$2(a_u) = R(a_u)\$ (a_u) \frac{\prod_{i=1}^{n+1} [1 - (R(a_i) + \gamma(a_i))]}{1 - (R(a_u) + \gamma(a_u))}$$

(Denominator accounts for the absence of a_u)

$$\begin{aligned} diff_1 &= \$_1(a_u) - \$_2(a_u) \\ &\text{Substituting by values of } \$_1(a_u) \text{ and } \$_2(a_u), \end{aligned}$$

$$\begin{aligned} &= R(a_u)\$(a_u) \left[\prod_{i=1}^{u-1} [1 - (R(a_i) + \gamma(a_i))] - \frac{\prod_{i=1}^{n+1} [1 - (R(a_i) + \gamma(a_i))]}{1 - (R(a_u) + \gamma(a_u))} \right] \\ &\text{Taking the common factor outside,} \\ &= R(a_u)\$(a_u) \prod_{i=1}^{u-1} [1 - (R(a_i) + \gamma(a_i))] \times \\ &\quad \left[1 - \prod_{i=u+1}^{n+1} [1 - (R(a_i) + \gamma(a_i))] \right] \quad (\text{A-1}) \\ &\quad \left(\text{Note: } \frac{\prod_{i=u}^n [1 - (R(a_i) + \gamma(a_i))]}{1 - (R(a_u) + \gamma(a_u))} = \prod_{i=u+1}^{n+1} [1 - (R(a_i) + \gamma(a_i))] \right) \end{aligned}$$

Let S be sequence of ads from a_{u+1} to a_{n+1}

$$\begin{aligned} \$_1(S) &= \sum_{i=u+1}^{n+1} R(a_i)\$(a_i) \prod_{j=1}^{i-1} [1 - (R(a_j) + \gamma(a_j))] \\ \$_2(S) &= \frac{\sum_{i=u+1}^{n+1} R(a_i)\$(a_i) \prod_{j=1}^{i-1} [1 - (R(a_j) + \gamma(a_j))]}{1 - (R(a_u) + \gamma(a_u))} \end{aligned}$$

$$\begin{aligned} diff_2 &= \$_2(S) - \$_1(S) \\ diff_2 &= \sum_{i=u+1}^{n+1} R(a_i)\$(a_i) \prod_{j=1}^{i-1} [1 - (R(a_j) + \gamma(a_j))] \times \\ &\quad \left[\frac{1}{1 - (R(a_u) + \gamma(a_u))} - 1 \right] \quad (\text{A-2}) \end{aligned}$$

Dividing $diff_1 - diff_2$ by common factor, Let

$$\begin{aligned} \Delta &= \frac{diff_1 - diff_2}{\prod_{i=1}^{u-1} [1 - (R(a_i) + \gamma(a_i))]} \\ \Delta &= R(a_u)\$(a_u) \left[1 - \prod_{i=u+1}^{n+1} [1 - (R(a_i) + \gamma(a_i))] \right] - \\ &\quad \left\{ \sum_{i=u+1}^{n+1} R(a_i)\$(a_i) \prod_{j=u}^{i-1} [1 - (R(a_j) + \gamma(a_j))] \times \right. \\ &\quad \left. \left[\frac{1}{1 - (R(a_u) + \gamma(a_u))} - 1 \right] \right\} \end{aligned}$$

We have to prove $\Delta > 0$

Simplifying and dividing by $R(a_u) + \gamma(a_u)$

$$\begin{aligned} r\Delta &= \frac{R(a_u)\$(a_u)}{R(a_u) + \gamma(a_u)} \left[1 - \prod_{i=u+1}^{n+1} [1 - (R(a_i) + \gamma(a_i))] \right] - \\ &\quad \left\{ \sum_{i=u+1}^{n+1} R(a_i)\$(a_i) \prod_{j=u+1}^{i-1} [1 - (R(a_j) + \gamma(a_j))] \right\} \quad (\text{A-3}) \end{aligned}$$

Since a_{u+1} to a_{n+1} is placed below a_u in P_1 ,

$$\forall_{i=u+1}^{n+1} \frac{R(a_i)\$(a_i)}{R(a_i) + \gamma(a_i)} \leq \frac{R(a_u)\$(a_u)}{R(a_u) + \gamma(a_u)}$$

$$\begin{aligned} &\Rightarrow \forall_{i=u+1}^{n+1} R(a_i)\$(a_i) \\ &\leq \frac{R(a_u)\$(a_u)}{R(a_u) + \gamma(a_u)} (R(a_i) + \gamma(a_i)) \quad (\text{A-4}) \end{aligned}$$

Note that the value of $r\Delta$ in A-3 is strictly decreasing with value of $\$(a_i)R(a_i)$ factor in second term. Using the condition in constraint A-4 for $\$(a_i)R(a_i)$ in second term of Equation A-3 we get,

$$\begin{aligned} r\Delta &\geq \frac{R(a_u)\$(a_u)}{R(a_u) + \gamma(a_u)} \left[1 - \prod_{i=u+1}^{n+1} [1 - (R(a_i) + \gamma(a_i))] \right] - \\ &\quad \left\{ \sum_{i=u+1}^{n+1} \frac{R(a_u)\$(a_u)}{R(a_u) + \gamma(a_u)} (R(a_i) + \gamma(a_i)) \prod_{j=u+1}^{i-1} [1 - (R(a_j) + \gamma(a_j))] \right\} \end{aligned}$$

Canceling Common Factor from RHS,

$$\begin{aligned} &1 - \prod_{i=u+1}^{n+1} [1 - (R(a_i) + \gamma(a_i))] - \\ &\quad \sum_{i=u+1}^{n+1} (R(a_i) + \gamma(a_i)) \prod_{j=u+1}^{i-1} [1 - (R(a_j) + \gamma(a_j))] \end{aligned}$$

For the above function, substitute the values

$$r_i = 1 - (R(a_{u+i}) + \gamma(a_{u+i}))$$

Now, *RHS* is,

$$\begin{aligned} RHS &= 1 - (r_1.r_2\dots r_n) - [(1 - r_1) + (1 - r_2).r_1 + \\ &\quad (1 - r_3).r_1.r_2 + \dots + (1 - r_n).r_1.r_2.r_{n-1}] \\ &= 1 - (r_1.r_2\dots r_n) + [(1 - r_1) + (r_1 - r_1.r_2) + \\ &\quad (r_1.r_2 - r_1.r_2.r_3) + \dots \\ &\quad + r_1.r_2\dots r_{n-1} - r_1.r_2\dots r_{n-1}.r_n] \end{aligned}$$

Terms will cancel out and this function will be 0 on simplification, irrespective of the values of the r_i s. This means

$$r\Delta \geq 0$$

i.e the difference $E(\$_1) - E(\$_2) \geq 0$. This completes the induction. \square

A-3 Proof of Theorem 3

PROOF. Induction on number of clicks.

Base Case: Single click, proved in Theorem 2.

Inductive Hypothesis: True for n clicks.

For $n+1^{th}$ click, user starts browsing down at the position next to the n^{th} clicked ad. let n^{th} clicked ad be a_c . Since there is only one click remaining, optimal ordering of ads is in the descending order of RF by base case. But we know that the ads above a_{c+1} (a_1 to a_c) are no longer relevant to the user since he has seen them already. Since the relevance values of these ads are zero now (hence RF s of these ads are also zero), the optimal order is to remove the ads from a_1 to a_c and reorder the remaining ads in descending order of RF . Since the relevance and abandonment probabilities of these ads—ads except a_1 to a_c —remain unchanged, the resulting optimal sequence will be the sub-sequence of original ranking starting from a_c to the end of the list, which is exactly the sequence the user traverses in original ranked list after n^{th} click on a_c i.e. the original ranked list is optimal for $n+1$ clicks. \square