Optimal Strategies for Reviewing Search Results

Jeff Huang
Information School
University of Washington
aaai@jeffhuang.com

Anna Kazeykina
Computational Mathematics and Cybernetics
Moscow State University
kazanna@bk.ru

Abstract

Web search engines respond to a query by returning more results than can be reasonably reviewed. These results typically include the title, link, and snippet of content from the target link. Each result has the potential to be useful or useless and thus reviewing it has a cost and potential benefit. This paper studies the behavior of a rational agent in this setting, whose objective is to maximize the probability of finding a satisfying result while minimizing cost. We propose two similar agents with different capabilities: one that only compares result snippets relatively and one that predicts from the result snippet whether the result will be satisfying. We prove that the optimal strategy for both agents is a stopping rule: the agent reviews a fixed number of results until the marginal cost is greater than the marginal expected benefit, maximizing the overall expected utility. Finally, we discuss the relationship between rational agents and search users and how our findings help us understand reviewing behaviors.

Introduction

Web search engines typically return numerous results for a search query, but it is impractical to review every result before clicking. This is expected, since considering the first few results is often sufficient for finding a satisfying link, while lower ranked results are rarely relevant. Indeed, empirical studies have shown this is how users interact with the search results page (Cutrell and Guan 2007; Joachims et al. 2007).

As a user scans the search results, they are continuously deciding whether they have reviewed enough results, and when it is time to select a link to click. We explore this as a decision problem by determining the costs and benefits of reviewing results. We compare the optimal strategies under two different sets of assumptions about how users judge search results: individually vs. pairwise comparisons. We model these assumptions using two rational agents: the Comparison Agent and the Threshold Agent; the agents' objectives are to maximize the probability of finding a satisfying result while minimizing cost. The *Comparison Agent* is a hypothetical searcher based on observed behavior where users viewed a large number of results and decided from these which result was most likely to be satisfying (Miller and Remington 2004; Klöckner, Wirschum, and Jameson 2004; Brumby and Howes 2008). Likewise, this agent only compares result snippets relatively, and picks the perceived best result after completing its review. This seems reasonable since people can easily compare two text snippets to determine which is more promising; people often compare between choices rather than judge each option individually in daily decisions (Ariely 2009). The *Threshold Agent* is based on observed behavior from (Miller and Remington 2004; Young 1988; Fu and Pirolli 2007), where users decide at each point whether to select the curent result, abandon the query, or continue reviewing results. The goal of this agent is to simply find a result which appears to be satisfying. This agent represents a person who predicts whether the underlying web page after clicking the result will meet their need. For both agents, it is this perceived quality of a result which they consider. From this, we can model pre-click behavior on the search results page but not post-click behavior or actual satisfaction from the result.

The strategy for both agents is shown to be an optimal stopping rule: an agent sequentially reviews results starting from the top, and decides when to stop. Once stopped, it can select either the current result, or an earlier result. The optimal number of results the agent reviews is a function of the cost of reviewing results and the result quality.

This research is prescriptive rather than descriptive; the problems are solved for an optimal rational agent rather than describing observed user behavior. This approach differs from existing empirical work in this area and adds a theoretical perspective. The primary contribution is a solution to the optimal strategy problem posed for reviewing search results given the assumptions for each agent. The secondary contribution is a comparison of the relationship between the two agents and web users.

Related Work

Search Theory Problems

In a SIGIR plenary address on economics and search, (Varian 1999) introduces the concept of optimal search behavior and presents different search theory problems[1] that can be applied to information retrieval. Taking this approach, we

[1]See (Weitzman 1979) for examples of classic search theory problems.

use concepts of benefit and cost to find the optimal strategy for reviewing web search results.

Our problem was inspired by the well-known secretary problem (Freeman 1983) of finding the optimal strategy for an employer to pick the best candidate among N candidates interviewed in no particular order. After each interview, the employer decides whether to hire the candidate, but only benefits from hiring the best one. The optimal strategy has been shown to be a stopping rule: interview N/e candidates and pick the first candidate from the remaining candidates who is better than any previous candidate. (Aumann et al. 2008) generalizes search problems to include unknown costs. Our work is specific for the domain of search results and our search cost is situationally dependent.

Eye-Tracking Studies

Aula et al. (Aula, Majaranta, and Räihä 2005) divide searchers into two types: economic and exhaustive evaluators. Economic evaluators quickly decide their next action after reviewing a small number of results. Exhaustive evaluators review a large number of results before determining their next action. (Klöckner, Wirschum, and Jameson 2004) qualitatively find 65% of users employ a depth-first search strategy in reviewing results (clicking the first interesting result), 15% of users employ a breadth-first search strategy (reviewing all results before clicking), and 20% of users look ahead for several results before deciding which to click. Cutrell and Guan (Cutrell and Guan 2007) find that users review at least 3 or 4 results, even when clicking on the first or second result. They also note that reviewing results has a different cost for different types of search: longer snippets benefit users in information queries but hamper their performance in navigational queries.

The preceding studies attribute different user behavior on search result pages to the users' inherent personalities. However, inconsistencies between strategies found in those studies preclude a unified theory. We believe that user behaviors are dependent on situational and dispositional factors in addition to personality. Thus, theoretical models can provide a holistic perspective of actual behaviors by considering factors such as search engine quality and importance of query.

Cognitive Browsing Models

Several studies from human-computer interaction present browsing models derived from cognitive psychology. In a model of optimal strategy in application menus, Young's analysis (Young 1988) proposes Bayesian updating: a rational user would decide after each step whether to continue reviewing a link or to select the current link. We show that for search results, it is in fact optimal for both agents to decide upfront the number of results to review. SNIF-ACT (Fu and Pirolli 2007) is a cognitive browsing model that also explores user satisficing behavior. They use a metric called information scent, quantified as the Pointwise Mutual Information between words in the information need and link text on a page. Brumby and Howes (Brumby and Howes 2008) apply the ACT-R cognitive theory in two experiments to show that link selection depends on the quality of previously reviewed results. They find that a satisfying link is more likely to be selected if the previously reviewed results were highly unsatisfying and conversely, users review more results when the general quality of search results is high. Miller and Remington construct a browsing model (Miller and Remington 2004) to contrast two plausible user strategies for initially selecting the link on a page to click. Based on a rationality principle, they propose that users either employ a threshold strategy, where they immediately click a link which seems satisfying enough (above a threshold), or a comparison strategy, where the user reviews a fixed set of links and selects the best one. Our two agents are based on these two strategies as innate characteristics.

Click Data Analysis

Wang et al. define a Pskip value as the probability a user will skip a result they review on the search results page (Wang, Walker, and Zheng 2009). Their model is based on the assumption that the first relevant result follows a geometric distribution. From this, they derive search metrics such as average search length, mean reciprocal rank, estimated search length, and clickthrough rate. More recent work by Wang et al. in (Wang, Gloy, and Li 2010) predict user views and clicks from click logs using a partially observable markov model. This model closely agrees with eye-tracking studies and can account for ads and positional bias on the search results page.

Model Formulation

The problem we study is derived from similar search theory problems: we want to model the agent's optimal strategy for selecting a *satisfying result* while accounting for cost. We formulate this as finding the strategy which maximizes the expected utility of the agent which equals the benefit minus cost. The benefit in this model only occurs when the agent successfully chooses a satisfying result, while the cost increases linearly with the number of results reviewed. We introduce a cost coefficient, k, which represents the effort an agent puts into reviewing each result to improve their chances of finding a satisfying result. This parameter makes the model flexible for tuning to each search.

The agent's strategy can be shown to be a stopping rule by proving that the expected utility is a unimodal function. In other words, the function is monotonically increasing and then monotonically decreasing, and so the stopping point is at a local maximum. This is a classic example of a normative (or prescriptive) decision theory problem, where the objective is to maximize the expected utility function. We will begin defining this more formally.

If n is the number of the results the agent reviews, the utility to the agent is

$$U(n) = B(n) - C(n)$$

where $B(n)$ and $C(n)$ are the benefit and cost functions of the whole review process. In the optimal strategy, the agent chooses to review up to N, which maximizes the expected utility, $\mathbb{E}U(N)$. The equation becomes,

$$\mathbb{E}U(N) = \mathbb{E}B(N) - \mathbb{E}C(N) \tag{1}$$

Power Law of Satisfying Results

The probability the first satisfying result is in the ith position decreases. This is expected because search engines sort the returned results by the descending probability of satisfying the user. We postulate that the first satisfying result occurs at position i with a power law probability distribution, like other properties on the web (Faloutsos, Faloutsos, and Faloutsos 1999; Huberman et al. 1998). This postulate can be supported experimentally by using the distribution of click logs as a proxy for result quality. Both the AOL click logs (Pass, Chowdhury, and Torgeson 2006) and data from (Joachims et al. 2007) showed that for user click positions, the power law fits better than any other common distribution, typically with $R^2 \geq 0.99$. The power law of result quality should hold even in an ideal search engine because result quality is subjective; a search engine can only hope to satisfy the most users with its top result, the second most users with its second result, etc. This postulate expressed formally is,

$$\mathbb{P}(\sigma = i) \propto \frac{1}{i^b} \quad \forall i \in \mathbb{N} \tag{2}$$

where $\mathbb{P}$ defines a probability distribution function. σ is the position of the first satisfying result, where $\sigma = 1$ is the first (highest ranked) result returned. b is the exponent to our power law distribution; intuitively, it represents the dropoff for the distribution of the first satisfying result. In AOL click logs, we find $b \approx 1.9$.

We introduce p_0 representing the probability that there are no satisfying results for a query. This is a realistic possibility when the search engine does not produce a result from its index that satisfies the user, or when there is a mismatch between a searcher's information need and the query.

Our power law distribution based on the assumption in (2) can be defined,

$$p_i = \mathbb{P}(\sigma = i) = \frac{1-p_0}{\zeta(b)} \frac{1}{i^b} \quad \forall i \in \mathbb{N}$$

where $\zeta(b) = 1 + \frac{1}{2^b} + \frac{1}{3^b} + \ldots$; $\zeta(b)$ is commonly referred to as the Riemann zeta function. The term $\frac{1-p_0}{\zeta(b)}$ serves to normalize the probability distribution.

Comparison Agent

The Comparison Agent can judge the relative quality of results but not determine which result is satisfying before clicking. In other words, it knows whether one link is better than another, and therefore can tell which previously reviewed link is the best thus far. This agent does not determine from a result's snippet whether the link will actually satisfy an information need. An equivalent framing of this assumption is: the agent has to determine the number of results to review in advance, N, and must review those N results even when a satisfying result occurs before the agent has completed reviewing N results. The optimal strategy for this agent is solved by finding this N with the maximum expected utility, which we call N_C. We now formally state the problem and solution for this agent.

There is only benefit to the agent if the first satisfying result is within the N results reviewed. The value of the user's benefit is 1 when they select a satisfying result and 0 otherwise.

$$B(N) = \begin{cases} 1 & \text{if } \sigma \leq N \\ 0 & \text{if } \sigma > N \end{cases}$$

The expected benefit is thus,

$$\mathbb{E}B(N) = \mathbb{P}(\sigma \leq N). \tag{3}$$

Let k be the cost of reviewing each search result. This coefficient represents the value of time and effort spent evaluating a result relative to the value of a satisfying result. The cost function can be written,

$$C(N) = kN.$$

Theorem 1. *A Comparison Agent with coefficient k maximizes expected utility by reviewing $N_C = \left\lfloor \sqrt[b]{\frac{1-p_0}{k\zeta(b)}} \right\rfloor$ results.*

Proof. To prove that the optimal strategy is to review N_C results, we must first show that $\mathbb{E}U(N)$ is unimodal.

Lemma 1. *The Comparison Agent's expected utility, $\mathbb{E}U(N)$, is a unimodal function.*

Proof. Unimodality can be proved by showing the following properties: (1) if the function decreases, it decreases indefinitely, and (2) there is a point where the function decreases.

We expand the expected utility equation (1),

$$\begin{aligned} \mathbb{E}U(N) = \mathbb{P}(\sigma \leq N) - kN &= \sum_{j=1}^{N} \mathbb{P}(\sigma = j) - kN \\ &= \frac{1-p_0}{\zeta(b)}(1 + \frac{1}{2^b} + \ldots + \frac{1}{N^b}) - kN. \end{aligned}$$

In this form, we can see that if $\mathbb{E}U(N) > \mathbb{E}U(N+1)$, then $\mathbb{E}U(N) > \mathbb{E}U(M) \quad \forall M > N$. We can also see that $\mathbb{E}U(N)$ decreases at some point because $\mathbb{E}U(1)$ is a finite number and $\mathbb{E}U(N) \to -\infty$ when $N \to \infty$. $\square$

Since $\mathbb{E}U(N)$ is unimodal, N_C must be the point where $\mathbb{E}U(N)$ is at its maximum[2] if it is a local maximum,

$$\mathbb{E}U(N) \geq \mathbb{E}U(N-1), \mathbb{E}U(N) > \mathbb{E}U(N+1). \tag{4}$$

(4) expressed in terms of expected benefit and cost,

$$\mathbb{E}B(N) - k \geq \mathbb{E}B(N-1), \mathbb{E}B(N) > \mathbb{E}B(N+1) - k.$$

Thus, N is the optimal number of results to review when

$$\frac{1-p_0}{\zeta(b)(N+1)^b} < k \leq \frac{1-p_0}{\zeta(b)N^b} \tag{5}$$

which intuitively describes the point when the marginal cost surpasses the marginal benefit of reviewing an additional result; hence, it is suboptimal to review an additional result. Solving for N in (5), we obtain $N = \left\lfloor \sqrt[b]{\frac{1-p_0}{k\zeta(b)}} \right\rfloor$ as the optimal stopping point. $\square$

[2] If $\mathbb{E}U(N) = \mathbb{E}U(N-1)$, i.e., there is no difference in expected utility for reviewing an additional result, we assume they will review that result, only to produce simpler expressions.

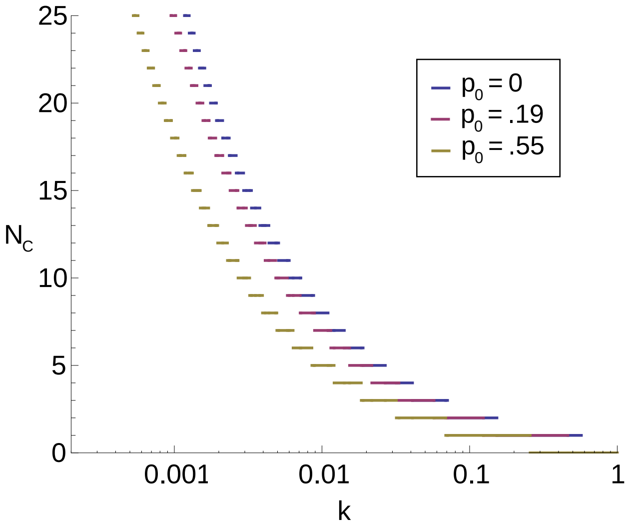

Figure 1: Comparison Agent: k, the coefficient representing the cost an agent places on reviewing each result, is plotted against N_C, the optimal number of results to review. An empirically obtained power law exponent of $b = 1.9$ is used.

We plot N_C against k in Figure 1 and see that $N_C(k)$ in the Comparison Agent is a power law step function. Assuming the probability there are no satisfying results is approximately the percentage of abandoned queries, which is between 19% and 55% depending on locale and platform (Li, Huffman, and Tokuda 2009), we can plot these as various p_0 values. As expected, the more valuable a satisfying result is to the agent (smaller k), the more results the agent reviews.

Threshold Agent

An alternate agent is one that determines whether an individual result will be satisfying before clicking. We name this agent the Threshold Agent since in practice, it would look at the result snippet and decide whether it surpasses some quality threshold. This new ability supercedes the comparison ability, which is no longer necessary for finding a satisfying result. Unlike the Comparison Agent which always reviews N results and then stops, the Threshold Agent has three options at each step: click the current result, review another result, or abandon the search.

At each step, the agent reviews an additional result if the current result is unsatisfying and it has reviewed less than N results. Unlike the Comparison Agent, the optimal N for the Threshold Agent, N_T, is not necessarily the actual number of results reviewed since the Threshold Agent stops once it has found a satisfying result. In other words, the Threshold Agent reviews $n = min(\sigma, N)$ results where σ is the position of the first satisfying result. If a satisfying result is not found among the first N results, the Threshold Agent abandons the search. To prove that this strategy is optimal, we show that the expected utility is a unimodal function. We then express the relationship between k and N_T.

The expected benefit is the probability of finding a satisfying result among N results. The benefit does not change from the Comparison Agent so (3) expands to,

$$\mathbb{E}B(N) = p_1 + p_2 + \ldots + p_N = \frac{H_b(N)}{Z}$$

where Z is the normalizing constant $\frac{\zeta(b)}{1-p_0}$, and $H_b(N)$ is the generalized harmonic number of order N of b. Similarly, the expected cost can be expressed[3],

$$\begin{aligned}\mathbb{E}C(N) &= k(p_1{+}2p_2{+}\ldots{+}Np_N){+}kN(p_0{+}p_{N+1}{+}\ldots)\\ &= k\sum_{i=1}^{N} i\cdot p_i + kN(p_0 + \sum_{j=N+1}^{\infty} p_j)\\ &= \frac{kH_{b-1}(N)}{Z} + \frac{kN(Zp_0+\zeta(b)-H_b(N))}{Z}. \qquad (6)\end{aligned}$$

This represents the final cost at the stopping point when the agent has reviewed N results or has found a satisfying result early.

Using equation (1), the expected utility becomes

$$\begin{aligned}&\mathbb{E}U(N) = \mathbb{E}B(N) - \mathbb{E}C(N)\\ &= \frac{1}{Z}\left(H_b(N) - kH_{b-1}(N) - kN(Zp_0 + \zeta(b) - H_b(N))\right).\end{aligned}$$

Like with the Comparison Agent, we will show that $\mathbb{E}U(N)$ is a unimodal function to prove that the optimal user strategy is to review up to N_T results, maximizing $\mathbb{E}U(N)$.

Lemma 2. *The Threshold Agent's expected utility, $\mathbb{E}U(N)$, is a unimodal function.*

Proof. Unimodality can be demonstrated first by showing

$$\mathbb{E}U(N) > \mathbb{E}U(N+1) \qquad (7)$$

implies

$$\mathbb{E}U(N+1) > \mathbb{E}U(N+2). \qquad (8)$$

Inequality (7) can be simplified,

$$k > \frac{1}{(N+1)^b(Zp_0+\zeta(b)-H_b(N))}. \qquad (9)$$

Since $\frac{N+1}{N+i} < \frac{N+2}{N+i+1}$ for $\forall i \geq 2$,

$$(N{+}1)^b(Zp_0{+}\zeta(b){-}H_b(N)) < (N{+}2)^b(Zp_0{+}\zeta(b){-}H_b(N{+}1)).$$

From this it follows,

$$k > \frac{1}{(N+2)^b(Zp_0+\zeta(b)-H_b(N+1))}$$

which is an expansion of the inequality in (8).

The final step is to note that $U(1)$ is a finite number and $U(N) \to -\infty$ when $N \to \infty$, and thus $U(N)$ starts decreasing at some point. □

Therefore, the inequalities in (4) must also hold for some N_T, the maximum number of results to review in the optimal strategy. This can be rewritten,

$$\begin{aligned}&\frac{1}{(N_T+1)^b\,(Zp_0+\zeta(b)-H_b(N_T))}\\ &\qquad < k \leq \frac{1}{N_T^b\,(Zp_0+\zeta(b)-H_b(N_T-1))}. \qquad (10)\end{aligned}$$

[3]Note that $\sum_{i=1}^{N} p_i = \frac{H_b(N)}{Z}$.

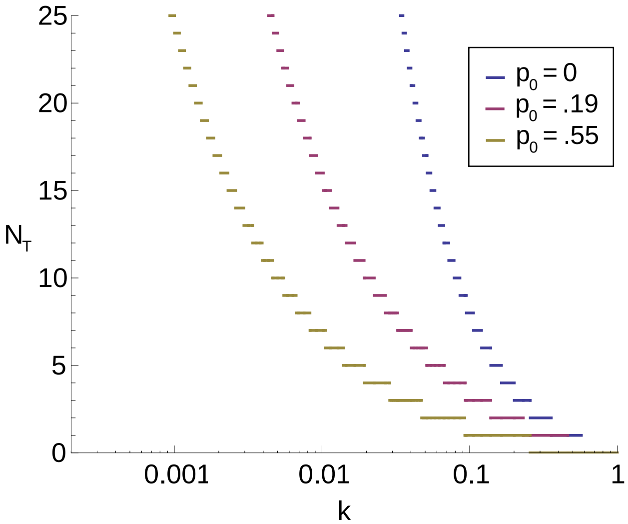

Figure 2: Threshold Agent: k, the coefficient representing the cost an agent places on reviewing each result, is plotted against N_T, the optimal maximum number of results to review. A power law exponent of $b = 1.9$ is used.

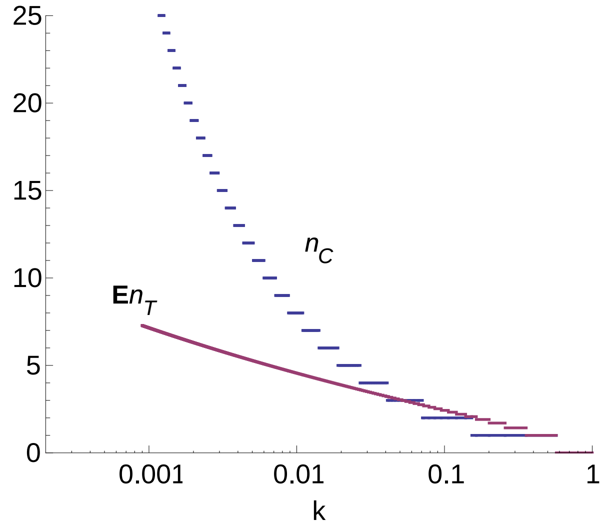

Figure 3: k, the coefficient representing the cost an agent places on reviewing each result, is plotted against $\mathbb{E}n_T$ and n_C, the expected number of results reviewed for the Threshold and Comparison agents, for $b = 1.9$, $p_0 = 0$.

A rational Threshold Agent with cost k maximizes expected utility by reviewing at most N_T results. Unfortunately, N_T cannot be represented in a closed form from inequality (10). However, we can plot the relationship between N_T and k in Figure 2.

Discussion

Optimal strategies of rational agents do not represent actual behaviors of users. Nevertheless, a theoretical model provides insights into the factors that influence users, like the invisible hand of rationality that guides economic behavior. Here we compare the two agents and relate them to users.

The Threshold Agent can better optimize the review of search results to determine whether a result is satisfying. The ability of the Threshold Agent to stop when it sees a satisfying result makes it more efficient but less discriminative. The advantage of the Comparison Agent is it selects the perceived best among several satisfying results. As mentioned previously, the Comparison Agent always reviews N_C results in an optimal strategy, i.e. $n_C = N_C$ where n is the number of results reviewed, but the Threshold Agent does not because it can stop before reviewing N_T results. In Figures 1 and 2, for example values of $p_0 = 0.55$ and $k = 0.01$, the optimal strategy for the Comparison Agent is to review 5 results while the Threshold Agent reviews at most 7 results. Figure 3 plots the Threshold Agent's $\mathbb{E}n_T$ against the Comparison Agent's n_C for $p_0 = 0$. Overall, the expected number of results the Threshold Agent reviews is quite low—as k reaches 0, $\mathbb{E}n_T$ remains below 10 because of the high likelihood of a satisfying result occuring early. This agrees with findings from empirical studies that users rarely view more than the first page of results (Joachims et al. 2007). An empirical distribution of n obtained from eye-tracking observations has shown that user behavior closely resembles the power law like the Comparison Agent[4]. While this Figure 3 shows the Threshold Agent reviews less results than the Comparison Agent on average, i.e. $\mathbb{E}n_T < n_C$ for most k, Figures 1 and 2 show $N_T > N_C$. This means that the Threshold Agent typically reviews less results, but will occasionally review more results than the Comparison Agent during some trials. This may be surprising, but can be attributed to the Comparison Agent not knowing when it has passed a satisfying result and hence accumulates more costs.

Comparing the relationships between N and k for the Comparison Agent (Theorem 1) and Threshold Agent in inequality (10), we find that for small values of k (and hence large values of N_C, N_T), $Zp_0 + \zeta(b) - H_b(N_T - 1) \approx Zp_0$. Therefore, for the Threshold Agent, $k \approx \frac{1}{N_T^b Z p_0}$. From this, we arrive at $\frac{N_C}{N_T} \approx \sqrt[b]{p_0}$. Conversely, for large values of k (small values of N_C, N_T), $Zp_0 + \zeta(b) - H_b(N_T - 1) \approx Zp_0 + \zeta(b)$ and using the equality, $Zp_0 + \zeta(b) = \frac{\zeta(b)}{1-p_0} p_0 + \zeta(b) = \frac{\zeta(b)}{1-p_0} = Z$, we arrive at $N_C \approx N_T$.

Theorem 2. *$\frac{N_C}{N_T} \approx \sqrt[b]{p_0}$ for small k, $N_C \approx N_T$ for large k.*

Theorem 2 tells us that when the cost of reviewing a result is relatively high, both rational agents continue reviewing to the same point. In contrast, when the cost is relatively low, the Comparison Agent has a lower stopping point than the Threshold Agent.

As expected, a higher probability of no satisfying results, p_0, causes the agents to review less results. For some values of k, an agent that has not yet found a satisfying result abandons the search even when $p_0 = 0$, i.e. there is a guaranteed satisfying result in a lower position, because

[4] View distributions, obtained via correspondence with (Cutrell and Guan 2007), followed a power law ($R^2 = 0.98$).

the cost of additional review is higher than the probability the next result is satisfying. This abandonment is similar to observed user behavior (Li, Huffman, and Tokuda 2009), where search users either reformulate their query or give up after reviewing a certain number of results. It also agrees with modeled browsing behavior (Huberman et al. 1998), where the user abandons surfing if the potential value for the destination page is low. What this means is that good results that are fetched but not highly ranked will be reviewed by few agents and users because of the high cost of reviewing and a low inherent value of a satisfying result.

The three parameters in our model support generalization to different search engines and search instances. The p_0 and b parameters define how likely the search engine will return a satisfying result and the degree of the power law distribution of these satisfying results. The k parameter adjusts the cost to the searcher for reviewing one result relative to the benefit of finding a satisfying result. For users of a web search engine, this represents the effort required to review a result and importance of a query. While a specific value of k may be difficult to measure, its value can be guided by user surveys or adjusted based on the search environment. For example, if a search engine targets a mobile device and knows that its users value a satisfying result half as much, it can infer how many less results to show on the screen; if a study shows that users spend more time reading result snippets for informational queries than navigational queries, e.g. (Cutrell and Guan 2007), measuring the relative effort between the query types can guide the placement of search assistance tools in the results, such as query reformulation suggestions or encouragement to continue reviewing results.

Conclusions

We model the task of reviewing search results using two rational agents. The goal is to maximize the probability of finding a satisfying result given some cost to review results. The general problem of how to review results has been studied empirically before; our work complements the user-based approaches of cognitive modeling, click data analysis, and eye-tracking studies. We compare optimal strategies under two different reviewing behaviors: comparing between viewed results vs. picking the first satisfying result. Our solution finds the optimal number of results that agents employing these strategies review to maximize expected utility. We show the relationship between N_C, N_T and k, the importance of finding a satisfying result relative to time and effort. We plan to continue this ongoing work by applying it to study re-reviewing of search results. Specifically, what actions does a rational agent take when it clicks but finds an unsatisfying destination page? The findings would help guide design of search interfaces and give us a better understanding of the process of reviewing search results.

Acknowledgements

We thank Jonathan Cross, Xueming Huang, and Alan Ritter for their feedback on earlier drafts. Anna Kazeykina was supported by RFBR (no. 08-07-00158) and the Scientific and Pedagogical Staff of Innovative Russia (1.2.1, NK-15P).

References

Ariely, D. 2009. *Predictably Irrational: The Hidden Forces That Shape Our Decisions*. HarperCollins Publishers Ltd.

Aula, A.; Majaranta, P.; and Räihä, K.-J. 2005. Eye-tracking reveals the personal styles for search result evaluation. In *Proceedings of INTERACT '05*, 1058–1061.

Aumann, Y.; Hazon, N.; Kraus, S.; and Sarne, D. 2008. Physical search problems applying economic search models. In *Proceedings of AAAI '08*, 9–16.

Brumby, D. P., and Howes, A. 2008. Strategies for guiding interactive search: An empirical investigation into the consequences of label relevance for assessment and selection. *Human-Computer Interaction* 23(1):1–46.

Cutrell, E., and Guan, Z. 2007. What are you looking for?: an eye-tracking study of information usage in web search. In *Proceedings of CHI '07*, 407–416.

Faloutsos, M.; Faloutsos, P.; and Faloutsos, C. 1999. On power-law relationships of the internet topology. In *Proceedings of SIGCOMM '99*, 251–262.

Freeman, P. 1983. The Secretary Problem and Its Extensions: A Review. *Int. Stat. Rev.* 51(2):189–206.

Fu, W.-T., and Pirolli, P. 2007. Snif-act: a cognitive model of user navigation on the world wide web. *Human-Computer Interaction* 22(4):355–412.

Huberman, B. A.; Pirolli, P. L.; Pitkow, J. E.; and Lukose, R. M. 1998. Strong Regularities in World Wide Web Surfing. *Science* 280(5360):95–97.

Joachims, T.; Granka, L.; Pan, B.; Hembrooke, H.; Radlinski, F.; and Gay, G. 2007. Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search. *ACM Trans. Inf. Syst.* 25(2):7.

Klöckner, K.; Wirschum, N.; and Jameson, A. 2004. Depth- and breadth-first processing of search result lists. In *Proceedings of CHI '04 extended abstracts*, 1539–1539.

Li, J.; Huffman, S.; and Tokuda, A. 2009. Good abandonment in mobile and pc internet search. In *Proceedings of SIGIR '09*, 43–50.

Miller, C. S., and Remington, R. W. 2004. Modeling information navigation: implications for information architecture. *Human-Computer Interaction* 19(3):225–271.

Pass, G.; Chowdhury, A.; and Torgeson, C. 2006. A picture of search. In *Proceedings of InfoScale '06*.

Varian, H. R. 1999. The economics of search. In *SIGIR '99 invited plenary*.

Wang, K.; Gloy, N.; and Li, X. 2010. Inferring search behaviors using partially observable markov (pom) model. In *Proceedings of WSDM '10*, 211–220.

Wang, K.; Walker, T.; and Zheng, Z. 2009. Pskip: estimating relevance ranking quality from web search clickthrough data. In *Proceedings of KDD '09*, 1355–1364.

Weitzman, M. L. 1979. Optimal search for the best alternative. *Econometrica* 47(3):641–54.

Young, R. M. 1988. *Rational analysis of exploratory choice*. Oxford, UK: Oxford University Press. 76–83.