

Adaptive Web Sites: User Studies and Simulation

Doug Warner
RightNow Technologies
Bozeman, MT

doug@rightnow.com

Stephen D. Durbin
RightNow Technologies
Bozeman, MT

sdurbin@rightnow.com

J. Neal Richter
RightNow Technologies
Bozeman, MT

nealr@rightnow.com

Zuzana Gedeon
RightNow Technologies
Bozeman, MT

zgedeon@rightnow.com

ABSTRACT

Adaptive web sites have been proposed to enhance ease of navigation and information retrieval. A variety of approaches are described in the literature, but consideration of interface presentation issues and realistic user studies are generally lacking. We report here a large-scale study of sites with dynamic information collections and user interests, where adaptation is based on an Ant Colony Optimization technique. We find that most users were able to locate information effectively without needing to perform explicit searches. The behavior of users who did search was similar to that on Internet search engines. Simulations based on site and user models give insight into the adaptive behavior and correspond to observations.

Categories and Subject Descriptors

I.2.6 [Artificial Intelligence]: Learning; I.6.4 [Simulation and Modeling]: Model Validation and Analysis

General Terms

Algorithms, Measurement, Experimentation, Human Factors

Keywords

Adaptive Web Site, Ant Colony Optimization

1. INTRODUCTION

Web sites that adapt to improve users' experience—by making it easier to locate relevant information—can be much more effective and easier to maintain than static sites. The AI community has responded to the challenge to develop methods for such adaptive web sites [6] with a variety of approaches [5, 7]. We report here on a different approach based on Ant Colony Optimization (ACO) concepts [2], in which site visitors rather than software agents act as the “ants.” We thus use the knowledge and interests of human searchers to influence the results of subsequent searches and potentially allow searching and browsing of identical document sets on different web sites to adapt to local user needs.

2. DESCRIPTION OF THE SYSTEM

Our study concerns an ACO search system as incorporated in the RightNow Technologies system for Internet customer service previously described in [3]. We are concerned here with the end-user pages which provide access, through various search interfaces, to documents relevant to self-service inquiries.

The two primary ACO methods in the RightNow system include

Copyright is held by the author/owner(s).
WWW 2006, May 23–26, 2006, Edinburgh, Scotland.
ACM 1-59593-323-9/06/0005.

providing an initial list of the currently most popular documents in the system, and suggesting related documents for each document. Each of these methods incorporates an algorithm for aging the ratings in the manner of ACO pheromone decay, and each also includes methods for bootstrapping initial performance (see [3] for more details). In the ACO approach, a pheromone scent is added to each selected document to induce subsequent visitors to view it. This document popularity, derived from both implicit and explicit measures, is used to present an ordered list of top-ranked documents on the initial search page, before any search query has been entered. Similarly, related documents are identified by building a directional link between documents viewed sequentially in a user session. These values are used for lists of potentially interesting documents presented to later users.

3. USER BEHAVIOR

For this analysis we collected database records and web logs from user sessions at six sites during two time periods, December 20-21, 2004 (‘-’ in the table) and December 26-27, 2004 (‘+’ in the table). Sites had from 637 to 9603 documents. Pre- and post-Christmas dates were chosen to allow comparison of any holiday rush effect. Summary results are shown in Table 1.

Table 1. Summary statistics for sites studied

	Session	Docs View	%Sess Search	Srchs	Sess Acts
Technical -	1073	1.54	14.66	2.40	2.39
Technical +	800	1.59	12.42	2.53	2.41
Game1 -	145633	1.48	20.08	2.41	2.18
Game1 +	207831	1.58	23.36	2.59	2.40
Retail -	1363	2.02	18.20	2.42	3.59
Retail +	5233	2.03	23.65	2.47	4.48
Software -	420	2.06	22.94	2.93	4.52
Software +	203	2.31	20.00	3.49	4.18
USGovt -	84270	1.36	15.49	2.54	2.24
USGovt +	139183	1.40	15.25	2.63	2.34
Game2 -	24770	2.44	30.55	3.37	3.64
Game2 +	93800	1.70	33.21	2.90	3.07

The small percentage of sessions with search queries (% Sess Search) represents an important distinction between the system considered here and Web search engines [4]. This is possible, at least in part, because the ACO approach presents likely documents immediately. Clear evidence of user economization of effort is the rapid drop-off in viewing of results pages beyond the first. The observed exponential drop-off is similar to that reported in Jansen and Spink [4]. Interestingly, we found that users who performed searches were no different in this regard.

4. SIMULATION OF AN ACO WEB SITE

Since the user data presented above comes from live, commercial sites, it is not possible to experimentally manipulate the system functionality. In order to explore variations in system features without affecting human users, and to gain insight into ways that the ACO framework of an informational web site adapts by aggregating navigational behavior, we developed the simulation described in this section. Our approach is to start with simplified, plausible models that allow us to discern the main effects that should apply to generic ACO-style sites, not just RightNow sites.

We assume a very basic web site that provides access to the information in a collection of documents. A simple user interface (UI) presents the documents as a list of links, ordered by descending document popularity. Each link provides a title to inform the user about its contents, and hence its likelihood of being a "goal" document. For any but the smallest collections, this list is normally spread over a number of pages, with 10-20 documents per page. In later considerations, the model is extended to include means to enter a search query with results ordered by match strength (possibly on multiple pages).

The document collection structure can be described by a document-document similarity matrix S , in which similarity is highest within a subtopic, and lowest if the respective documents are from different major topics. For the simulation, the document similarities were taken in random ranges reflecting the topic/subtopic structure.

The user actions modeled are viewing an individual document by clicking its link, advancing to the next page of the document list, or quitting the session. Selection of an action is governed by random numbers. We assume that each user is characterized by values for the following four attributes: topic interest, degree of focus on this interest, ease of satisfaction, and tolerance for item scanning and page turns. We also allow that a user might choose to view a document that turns out not to be as relevant as expected, due to limited information scent in a document title. If a search box is available, the user may elect to enter a search query representing his or her information need.

We restrict ourselves here to selected results relating to the following two questions: (1) Can simple ranking by document popularity reduce user effort and increase satisfaction? (2) Can document similarity (S) be reconstructed from the document-document links induced by the user navigation behavior?

Not surprisingly, the simulation shows that, as a site adapts, an average user is likelier to find the desired information, and find it sooner. Less intuitively, investigation of the role of user tolerance highlighted the importance of relatively tolerant users, who are more likely to reach and thereby reinforce potentially popular documents not highly ranked initially. We also found more rapid adaptation with stronger non-uniformity of user interest.

Addressing the second question, we found that the degree to which the simulated navigation-induced similarity matrix matches the assumed underlying document similarity matrix depends on the model parameters for the degree of focus of users, their methodical vs. random selection style, and their relative propensities to browse or search. The effect of searching, which reduces the length and increases the relevance of the document list that a user faces, is similar to having more patient users with lower drop-off rates. The navigation-induced similarity reflects fairly well the actual similarity, representing both topic and

subtopic structure. The same was found for the actual, working USGovt site.

5. DISCUSSION AND CONCLUSIONS

We found user behavior to be similar on web search engine page and ACO site search pages; except where architectural differences increase the effectiveness of the ACO approach. Even with the changes in visitation experienced between the two dates, the summary statistics reported remain fairly consistent. Despite the difference in the nature of the web sites (corporate sites using the RightNow application vs. a full web search engine), the number of searches performed in a search session, 2.7, was similar to those reported in [4]. For both types of sites, the drop-off in results pages viewed was roughly exponential, though decreasing more rapidly for the ACO sites than for the web search engine. Additional effects of web spiders and unusual spikes in usage within a single site will be discussed in the poster.

Simulated users were consistent with observed behavior in terms of numbers of documents viewed. However, we found that page drop-off rates as high as those observed provide severe difficulties for a simplistic popularity algorithm, because few users reach later pages. Usage of search effectively increases users' patience. Still, a more adaptive approach seems indicated, perhaps via page-dependent normalization. Strategies such as placing new items on the first page where they can be assessed by most users, may also be necessary. A similar issue was discussed in [1].

A major feature of the RightNow system is that it does not require an initial search before presenting an ordered list of documents. The relatively low percentage of sessions containing searches, about 20%, suggests that it is not necessary to enter a search query to have a successful experience with the site. This is also supported by independent user surveys.

6. REFERENCES

- [1] Cho, J., and Roy, S. Impact of Search Engines on Page Popularity. In Proceedings of the World-Wide Web Conference (WWW 2004), pp. 20-29.
- [2] Dorigo, M., Di Caro, G., and Gambardella, L. M. Ant algorithms for discrete optimization. *Artificial Life*, 5(2), 1999, pp. 137-172.
- [3] Durbin, S., Warner, D., Richter, J. N., and Gedeon, Z. Information Self-Service with a Knowledge Base That Learns, *AI Magazine*, 23(4), 2002, pp. 41-49.
- [4] Jansen, B. J., and Spink, A. An Analysis of Web Documents Retrieved and Viewed, Proceedings of the International Conference on Internet Computing, IC '03, Arabia, H.R., and Mun, Y., Eds., 2003, pp. 65-69.
- [5] Koutri, M., Avouris, N., and Daskalaki, S. A survey of web usage mining techniques for building web-based adaptive hypermedia systems, *Adaptable and Adaptive Hypermedia Systems*, Chen, S. Y., and Magoulas, G. D., Eds, IRM Press, Hershey, PA, 2005, pp. 125-149.
- [6] Perkowski, M., and Etzioni, O. Adaptive Web sites: An AI challenge, Proc. Fifteenth International Joint Conference on Artificial Intelligence, IJCAI 97, 1997, pp. 16-23.
- [7] Perkowski, M., and Etzioni, O. Towards adaptive Web sites: Conceptual framework and case study, *Artificial Intelligence*, 118(1-2), 2000, pp. 245-275.