

Web Link and Traffic Analysis for Recommendation

Brian D. Davison
Department of Computer Science and Engineering
Lehigh University
davison@acm.org

The Web has two pervasive sources of activity that can be useful for recommendation systems. One is the traffic on the Web, reflected in individual user histories as well as proxy and origin server usage logs. The second is in hypertext link authoring. Both provide information that can often be treated as a value judgment. A user who clicks on a link, in effect, believes that there is some content of value at the destination. The significance of link creation is thus even stronger, as it reflects the judgment of the author (presumably) after viewing the target content and assessing it to be of sufficient value to spend the time to make the connection available to readers.

We have examined the relationship between the description of a page and its contents [4], and explored the value of link analysis for search engine result ranking and community discovery [8], as well as tested for nepotistic links that may influence such analysis [3].

Perhaps unlike most researchers in recommender systems, our interest is not only recommendation for human usage, but also recommendation for systems. Adaptive systems can then use those recommendations (from whatever source) to improve performance invisibly (unlike, for example Letizia [10] or WebWatcher [9] which alter the Web interface that is presented to the user). We are also concerned with appropriate evaluation of such systems.

While in the past we have examined techniques to build user models for user interfaces (such as the UNIX shell [7]), in recent years we have concentrated on issues related to modeling Web users for pre-loading of Web content to improve user-perceived performance [11]. We have proposed adaptive Web prefetching [1] which incorporates prediction from multiple sources, including:

- individual user histories,
- workgroup histories from proxy cache hints (a form of collaborative filtering),
- global user histories from origin server hints (again collaborative filtering),
- predictions from analysis of recently viewed content [6], and
- predictions from other available context, such as bookmarks, e-mail and usenet news.

Evaluation of Web pre-loading systems can be difficult. To that end, we have proposed [2] and implemented one

approach to evaluate prefetching proxy caches simultaneously, and built and validated [5] a detailed network and caching simulator to estimate user latencies for testing and comparing alternate history-based prefetching approaches.

References

- [1] B. D. Davison. Adaptive Web Prefetching. In *Proc. of the 2nd Workshop on Adaptive Systems and User Modeling on the WWW*, pp. 105–106, Toronto, May 1999.
- [2] B. D. Davison. Simultaneous Proxy Evaluation. In *Proc. of the 4th Int'l Web Caching Workshop (WCW99)*, pp. 170–178, San Diego, Mar. 1999.
- [3] B. D. Davison. Recognizing Nepotistic Links on the Web. In *Proc. of Workshop on Artificial Intelligence for Web Search*, pp. 23–28. AAAI Tech. Rep. WS-00-01, July 2000.
- [4] B. D. Davison. Topical Locality in the Web. In *Proc. of the 23rd Annual Int'l Conf. on Research and Development in Information Retrieval (SIGIR 2000)*, Athens, July 2000.
- [5] B. D. Davison. HTTP simulator validation using real measurements: A case study. In *Proc. of MASCOTS*, Cincinnati, OH, Aug. 2001.
- [6] B. D. Davison. Predicting Web actions from HTML content. Under submission, 2001.
- [7] B. D. Davison and H. Hirsh. Predicting sequences of user actions. In *Proc. of Workshop Predicting the Future: AI Approaches to Time-Series Problems*, pp. 5–12, Madison, WI, AAAI Tech. Report WS-98-07, July 1998.
- [8] B. D. Davison, A. Gerasoulis, K. Kleisouris, Y. Lu, H. Seo, W. Wang, and B. Wu. DiscoWeb: Applying Link Analysis to Web Search. In Poster proc. of the 8th Int'l World Wide Web Conf., pp. 148–149, Toronto, May 1999.
- [9] T. Joachims, D. Freitag, and T. Mitchell. WebWatcher: A Tour Guide for the World Wide Web. In *Proc. of the 15th Int'l Joint Conf. on Artificial Intelligence*, pp. 770–775. Morgan Kaufmann, Aug. 1997.
- [10] H. Lieberman. Letizia: An Agent That Assists Web Browsing. In *Proc. of the Int'l Joint Conf. on Artificial Intelligence (IJCAI'95)*, pp. 924–929, Montreal, Aug. 1995.
- [11] V. N. Padmanabhan and J. C. Mogul. Using predictive prefetching to improve World Wide Web latency. *Computer Communication Review*, 26(3):22–36, July 1996.