

Machine Learning in the World Wide Web

Tom M. Mitchell

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213, USA
Tom.Mitchell@cmu.edu

The flood of information available over the world wide web provides new opportunities for machine learning research. In software systems that find information for users, there is a role for learning the interests of individual users, and for learning how to locate different types of information. We describe two applications of machine learning to information finding, then suggest several directions for future research.

One task where learning can be useful is netnews reading. Although there are over 8000 electronic newsgroups, nobody reads more than a few. A newsreader that learned its user's interests could therefore help the user by locating appropriate articles within the thousands of unread newsgroups. NewsWeeder [Lang 1994] is one system with this goal. It employs a minimum description length approach to learn user ratings based on words occurring within the article.

A second task where learning is appropriate is helping user's locate desired information while browsing the world wide web. WebWatcher [Armstrong, et al. 95] is one system with this goal. It accompanies users as they browse the web, suggesting which hyperlinks to follow, and learning to give better advice based on the observed search trajectories of many users.

This talk will describe both of these systems, then consider a variety of basic machine learning research issues they raise. These include learning over extremely high dimensional spaces (i.e., text) with mostly irrelevant features, learning search control from observation and experimentation, combining supervised and unsupervised training data, and combining data obtained from many interrelated learning tasks (e.g., when learning the interests of many correlated users).

References

[Lang, 1994] Lang, K., NewsWeeder, 1994. Web URL <http://anther.learning.cs.-cmu.edu/ifthome.html>.

[Armstrong, et al., 1995] Armstrong, R., Freitag, D., Joachims, T., and Mitchell, T., WebWatcher: A learning apprentice for the world wide web, in *Proceedings of the 1995 AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments*, March 1995. Also see Web URL <http://www.-cs.cmu.edu:8001/afs/cs.cmu.edu/project/theo-6/web-agent/www/project-home.-html>