Overcoming the Tyranny of the Majority in Social Information Filtering Systems

Kazuhiro Kitagawa, Fumihiro Kato, Nobuo Saito
Keio University
Contact: kaz@sfc.keio.ac.jp

Keywords: social information filtering, reputation, statistical analysis, probabilistic reasoning, tyranny of the minority

Introduction and problem statement

Social information filtering [1] is a typical application of the Semantic Web [2]. It constructs a Web of trust based on reputation. The Semantic Web is a means of creating a large information space that contain users' profiles, including bookmarks and historical behaviors, in shareable machine understandable format. These profiles are considered to be good information sources about users for the purposes of information filtering. In other words, social information filtering systems can create trust information for the Semantic Web.

Social filtering is a process in which documents are selected to satisfy a relatively stable and specific information need. Research and services associated with the automatic mediation of information housed on networks have emerged in the past decade alongside the expansion of the Web and e-commerce. The surge of interest in social information filtering has contributed to a flood of information on the Web.

The most popular and successful social information filtering method is collaborative filtering [3]. This method is widely deployed in commercial recommender systems and search engines. The primary strategy of collaborative filtering is to maximize the average user's satisfaction. This method may lead to a "tyranny of the majority" whereby a small number of users with preferences that vary significantly from the norm never achieve satisfaction. This is a well-known problem of the collaborative filtering.

The other possible strategy for social filtering is to maximize the equitable distribution of satisfaction among users. This goal may lead to instances where the preferences of a small

minority ("tyranny of the minority") override the preferences of a large majority. A social filtering system must pay attention to these potential conflicts.

Users in a group may demonstrate multi-faceted characteristics. Forming a small group according to a single characteristic might solve the conflicts mentioned above by removing all characteristics of the group except for the one under particular consideration. In the past, some social filtering systems took account of explicitly given user information and the category of information being addressed. However, this process must incorporate implicit information. Users' preferences may change due to their immediate environment and dividing information into fixed classes is an arbitrary exercise.

Our social filtering research is unique in that we explore the issues surrounding how specific preferences can enable the identification of smaller groups comprised of users whose specific preferences are better aligned. We are interested in exploring how detailed user factors might be deployed to establish the specific interests of smaller groups rather than taking account of overall preferences. We believe commonality of preferences among small affinity groups is an important source of information about users.

We are now developing a prototype filtering system that takes account of the preferences of affinity groups that share bookmarks, preferences, and historical behavior with temporal information recorded in RDF [4] format. We will begin by presenting our research contributions to the Semantic Web and then discuss the architecture of the prototype system. Finally, we will address implementation issues.

Contributions to the Semantic Web

This work introduces the following research contributions:

1. A new approach to social information filtering that avoids a tyranny of the Web majority.
2. Improving quality of social filtering system
3. This will be an alternative solution to construct the semantic web trust layer.
4. Application of this method is unlimited, and can be applied not only to search engines and recommender systems, but also to any reputation management tool, such as FOAF [5].

Design Architecture

The Semantic Web is a type of huge distributed database that contains various kinds of information expressed according to a single information model, RDF. This database provides a good source of information about users for social filtering purposes. Information about each user's tastes, historical behavior, and online associates are available in RDF format for this social information filtering system.

First, the filtering system gathers data to establish a huge multi-faceted group to which the user belongs through the FOAF network. At this stage, the system doesn't know the availability of shared users' data. FOAF helps to organize the database to be used by the system over the network.

Filtering software attached to the user agent collects user information, automatically recording bookmarks, historical behavior such as past access history with temporal information, and read or ignored information from the search engine results. The information is expressed as only one of two values, like or dislike. The record of negative behavior is a very good source of information [6] as are Web bookmarks

Finally, the collected profile information is processed by statistical data analysis and probabilistic reasoning to achieve social filtering

Analysis of user information and reasoning

We have adopted a multi-stage evaluation methodology. The system first attempts to identify people who share general preferences by analyzing all the user information. Then, the system evaluates the results obtained in the previous stage by identifying the preferences of small affinity groups in order to avoid local minimum solutions, and tyranny of the Web majority. These two stages are repeated until a sufficiently small affinity group is achieved. Evaluation methods during these stages are guided by statistical analysis.

Finally, the system filters the information by formulating a reputation for the affinity group

created in previous stages. Whether probabilistic reasoning, fuzzy inference, or statistical methods will be used to determine the information that a given user may need is determined on the basis of user circumstances.

Implementation

This prototype system is implemented as plug-in software to the user agent and written in C. All data to be used for the filtering system are presented in RDF, using common Semantic Web data format. These RDF data are gathered from various users over the network guided by FOAF information.

Acknowledgements

Many thanks should go to Rachel Yould, Keio University for her significant contribution to this research. Mr. Nobuhisa Shiraishi, NEC spared for his time to improve system design and implementation.

References

1. U. Shardanaand, P. Maes, "Social information filtering: Algorithms for Automating Word of Mouth," Proc. of ACM CHI'95.
2. T. Berners-Lee, J. Handler, O. Lassila, "The Semantic Web," Scientific America, May 2001.
3. P. Resnick et al., "GroupLens: An Open Architecture for Collaborative Filtering of Netnews," Prof. of ACM CWCS'94
4. O. Lassila, R. Swick, "Resource Description Framework (RDF): Model and Syntax Specification," htt://www.w3.org/TR/1999/REC-rdf-syntax-19990222
5. The foaf project, "Friend of a Friend," http://www.foaf-project.org/
6. M.Morita, Y.Shinoda, "Information filtering based on user behavior analysis and best match text retrieval," Proc. of 7th Int. ACM-SIGIR Conf. on research and development in information retrieval, Springer-Verlag, 1994