Issues in Understanding Collaborative Filtering

Christopher Lueg
AI-Lab, Department of Computer Science
University of Zurich
Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
Tel. +41-1-63 54577 Fax +41-1-63 56809
lueg@ifi.unizh.ch

Abstract

In this paper, we discuss open issues in collaborative filtering, such as the nature of recommendations and the difference between generated recommendations and personal recommendations. We outline some strengths of socially embedded collaborative filtering processes, such as active collaborative filtering, and present passive collaborative filtering of Usenet news as a collaborative filtering approach that does not impose additional burden to the debit of its users.

1 Introduction

Collaborative filtering [4, 14], also referred to as recommending [15] or social filtering [17], represents a novel approach to information filtering that does not rely on the content or shape of objects as it is the case in content-based filtering. Instead, filtering relies on meta-data *about* objects, such as CDs, movies, book, or webpages. Data can either be collected automatically, i.e., data is inferred from the user's interaction with the filtering system, or data has to be voluntarily provided by the users of the filtering system. GroupLens[1] is a popular example for a hybrid approach combining ratings provided by users, data inferred from user behavior, e.g., the time spent reading articles as indicator for interest, and content-based data extracted from the objects under investigation, such as the proportion of spelling errors and included text in documents [16].

A related research direction in collaborative filtering is called active collaborative filtering [11]. It is based on encouraging people to share information with others rather than collecting ratings and modeling user interests in order to compile recommendations as in traditional collaborative filtering approaches.

Our interest in collaborative filtering originates from trying to understand information needs and the emergence of interests. Related research includes situated information filtering [10] which is a behavior-based approach to the filtering of Usenet newsgroups that does not rely on modeling user interests. Collaborative filtering offers additional ways to filter information without relying on explicit models of interests. Socially embedded filtering processes, such as active collaborative filtering [11] and passive collaborative filtering (see below), typically avoid two major obstacles in the context of traditional collaborative filtering. First, socially embedded approaches do not abstract away the social context of recommendations. Abstracting and aggregating recommendations changes the social nature of recommendations [8]. Second, abstraction always involves the so-called abstraction gap [10].

We proceed as follows. First, we discuss issues in understanding the nature of recommendations. Then, we discuss passive collaborative filtering as an collaboration approach that avoids major obstacles.

2 The Nature of Recommendations

A prerequisite to understanding collaborative filtering as a social process is understanding the nature of automatically generated (and mostly de-personalized) recommendations as well as the nature of personal recommendations. What are the differences and what do these two extremes have in common? What kind of recommendation is appropriate in a particular setting?

Our concerns with traditional collaborative filtering are based on the understanding that there is no such thing as context-independent information. The value of information always depends on its origin and the situational context of its usage. In this sense, our assumptions are divergent from the assumptions underlying much work in collaborative filtering (e.g., [7] and less radical [19]).

We consider automatically generated recommendations as similar -in a sense- to automatically generated newspapers. As pointed out by John Seely Brown and Paul Duguid [1], a newspaper does not just report news, it makes it. The news items included gain social status and warrants that comes from the combination of editorial selection, location on the page, and wide distribution. Their claim is that the personally tailored, genuinely unique electronic newspaper offers neither physical nor social continuity. Each individual output would be no more than that, individual, with little or no indication of its social significance. Transferred to recommendations, this may indicate that automatically generated recommendations

[1] `http://www.grouplens.org/`

lack social context in a similar way as automatically generated newspapers do.

Our perspective is somehow supported by empirical evidence collected during our experiments with active webpage filtering [9]. One of the results of an informal inquiry at the local computer science department with approximately 50 returned questionnaires was that most participants prefer personal recommendations sent by friends or colleagues. Only a minor part of the participants would investigate recommendations sent by people who are unknown to them.

Experiences with passive collaborative filtering in the Usenet domain (see below) also suggest that the receiver as well as the originator of information are important dimensions in contributing as well as consuming information.

However, the success of more or less de-personalized movie-recommendations in newspapers as well as the success of companies utilizing anonymizing recommender systems, such as the online bookstore Amazon[2], seems to indicate that de-personalized or even anonymous recommendations are sufficient in many settings. However, the success factors are largely unknown. It is possible that other factors in the broader context of the recommendations influence the success of recommendations, such as the summaries of movies that are typically jointly supplied with recommendations of movies in newspapers.

3 Collaborative Filtering & Groupware

We investigated in detail the introduction of active collaborative filtering of webpages to a medium-size organization [9]. In addition, we are participating in the SELECT[3] project which is a European research project in information filtering and collaborative filtering. SELECT is funded by the European Community. Part of our work in the broader context of this project is understanding collaborative filtering in its social dimension, i.e., prerequisites that are necessary to successfully introduce a recommender system to an organization as well as ways to raise interest in collaborative filtering among the members of the organization. Also, we are interested in assessment tools for the potential benefit of collaborative filtering for individuals as well as for organizations.

One of the major problems with collaborative filtering seems to be that it typically imposes additional burden to the debit of the users supplying recommendations in one form or another. The so-called cold-start problem has frequently been discussed in the literature but is yet unresolved. The problem is well-known from experiences with the frequent failure of early groupware applications: "who does the work and who gets the benefit?" [5]. Yet it is unclear in the collaborative filtering context whether a top-down benefit-oriented approach as envisioned by [3] or bottom-up motivation-based approaches, such as [11] and [9], are more promising when introducing collaborative filtering to organizations. Examples of successful motivation-based approaches in the groupware context have been reported in the literature [6].

[2] http://www.amazon.com/

[3] http://cmc.dsv.su.se/select/

4 Passive Collaborative Filtering

In order to investigate the potential benefit of exploiting social relationships for information filtering purposes we have developed a passive collaborative filtering approach for the Usenet domain. Especially within high-volume newsgroups it is a major problem to find the interesting discussions among the so-called noise, i.e., discussions that are off-topic or less interesting. Finding interesting discussions is even more complicated since topic drift [13] is a peculiarity of Usenet discussions, i.e., the "official" title, is often divergent from the topic that is actually discussed.

Passive collaborative filtering is based on providing users with data about the newsreading of other Usenet users. The approach is based on the observation that experienced Usenet participants use not only the subject of discussions but also the occurrence of contributions by certain other participants as indicators for potentially interesting discussions. Primitive author-based monitoring is possible with most existing newsreaders that offer some kind of filtering functionality. Author-based monitoring, however, works only if the selected persons indeed contribute *new* articles to a discussion. It remains invisible if these persons passively follow a discussion, i.e., reading all the articles of a discussion without contributing new ones.

Passive collaborative filtering extends author-based monitoring by visualizing the discussions that user-selected persons *passively* follow. Empirical evidence suggests that information about these discussions, i.e., discussions that have been found worth reading, helps users find interesting discussions. Considering this additional information then makes searching for interesting discussions a collaborative activity [2] instead of an individual activity.

Users of passive collaborative filtering are provided with visualizations of the newsreading data collected from selected persons acting as collaborators. Such collaborators are selected by the user and are typically socially respected persons participating in similar discussions as the user. The collaborators have to agree to provide the data since passive collaborative filtering builds on private usage data that is not accessible to the public. Usage data is automatically collected and does not impose additional burden to the debit of those providing the data as it is the case with rating-based approaches to collaborative filtering.

Besides the ease of use, passive collaborative filtering allows the user to consider only data provided by selected persons. This choice is typically not possible with traditional collaborative filtering approaches.

Our current implementation of passive collaborative filtering involved the modification of the newsserver, the newsreader, and the communication protocol between them. Modifying the InterNetNews (INN) 2.2 newsserver involved only the newsreader demon (NNRPD). The tracking function of the demon is now able to track individual users at particular sites. In addition, we extended the command set of the demon to deliver usage data of particular users on request. This is basically a transparent application-specific extension to the Network News Transport Protocol (NNTP) similar to the rating-exchange extension discussed in [12].

On the client side, we modified the spynews newsreader to incorporate usage data. spynews [10] is a newsreader that supports situated information filtering by monitoring the user's newsreading behavior and reordering discussions according to the attention the user paid to them. The spynews functionality is implemented as augmentation to the NNTP-based Knews[4] newsreader.

Passive collaborative filtering works like this: when the user enters a discussion group, the newsreader requests the usage data of the collaborators that the user has decided to consider for this particular newsgroup. The discussions that have been read by the collaborators are then augmented with little tags according to their usage. These collaboration-tags act as indicators for interesting discussions since one or more collaborators found these discussions worth reading. Good collaborators will read mostly interesting discussions and avoid or quickly abandon reading less interesting discussions. However, interests differ among people and people may read even uninteresting discussions for some reason. This is why collaboration-tags are indicators that have to be interpreted by the user.

First user experiences are promising and the additional information seems to be valuable since it provides help without any additional burden to the debit of the those supplying the helping information. However, there are many open questions that are currently under investigation. For example, it is unclear, how long data about usage in the past should be incorporated. This is especially important if it took several articles to find out that a discussion is uninteresting. Experiences suggest to consider only the last 24 to 72 hours (depending on the newsreading frequency of the collaborators).

Another open question is the best method to compute the collaborator-tag. We are investigating various absolute and relative measurements, such a s the number or the percentage of articles of a discussion that have been read. Besides other computational questions, such as privacy (NNTP is an insecure protocol) and scalability, this research directly addresses open social issues in collaborative filtering, such as reputation and trust, since passive collaborative filtering explicitly builds on the social relations among Usenet participants. These relations are typically abstracted away in traditional collaborative filtering approaches.

Experiences with passive collaborative filtering also suggest that many people are not willing to provide data concerning sensible domains except it is guaranteed that only personally known persons have access to the data. In our passive collaborative filtering approach, we have resolved the issue by relying on personal responsibilities. Requesting usage data requires first of all authentication to the newsserver as participant of the passive collaborative filtering experiment. In addition, a special password may be required to access the data of a particular person. This password has to be requested from the supplier of the data. This quite simple procedure ensures that only particular persons have access to the usage data of particular persons.

Currently, we are augmenting passive collaborative filtering in order to facilitate broader usage. A combination of virtual newsgroups as proposed by [18] and a tool (probably web-based) that allows for adding and removal of collaborators then permits the usage of standard off-the-shelf newsreaders. The virtual newsgroups then contain those discussions that are read by collaborators. The additional tool is necessary since standard newsreaders do not support filtering-related communication with the newsserver.

5 Conclusions

We have shown that there are still a lot of open issues in *using* collaborative filtering systems besides purely technical questions. At the workshop, we would like to contribute our perspective and our experiences. In turn, we would like to learn from the experiences of others in order to better understand the potential successes and failures of recommender systems.

6 Acknowledgments

This research was funded in part by the Swiss Federal Office for Education and Science under contract BBW No. 98.0065.

References

[1] John Seely Brown and Paul Duguid. Keeping it simple: Investigating resources in the periphery. In Terry Winograd, editor, *Bringing Design to Software*, pages 129–145. Addison-Wesley Publishing Company, New York, NY, 1996.

[2] Kate Ehrlich and Debra Cash. Turning information into knowledge: Information finding as col-

[4] `http://www.matematik.su.se/~kjj/`

laborative activity. In John L. Schnase, John J. Leggett, Richard K. Furuta, and Ted Metcalfe, editors, *Proceedings of the First Annual Conference on the Theory and Practice of Digital Libraries*, pages 119–125, June 1994.

[3] Natalie S. Glance, Damián Arregui, and Manfred Dardenne. Knowledge pump: Supporting the flow and use of knowledge. In U. Borghoff and R. Pareschi, editors, *Information technology for management*. Springer-Verlag, 1998.

[4] David Goldberg, David Nichols, Brian M. Oki, and Douglas Terry. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12):61–69, December 1992.

[5] Jonathan Grudin. Why CSCW applications fail: Problems in the design and evaluation of organizational interfaces. In *Proceedings of the International Conference on Computer Supported Collaborative Work (CSCW'88)*, pages 85–93, September 1988.

[6] Jonathan Grudin and Leysia Palen. Why groupware succeeds: Discretion or mandate? In *Proceedings of the Fourth European Conference on Computer Supported Collaborative Work (ECSCW'95)*, pages 263–278, September 1995.

[7] Will Hill, Larry Stead, Mark Rosenstein, and George Furnas. Recommending and evaluating choices in a virtual community of use. In Irvin R. Katz, Robert Mack, Linn Marks, Mary Beth Rosson, and Jakob Nielsen, editors, *Proceedings of the Annual ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'95)*, pages 194–201. ACM Press, May 1995.

[8] Christopher Lueg. Social filtering and social reality. In *Proceedings of the 5th DELOS Workshop on Filtering and Collaborative Filtering*, pages 77–81. European Research Consortium for Informatics and Mathematics (ERCIM), November 1997. ERCIM-98-W001.

[9] Christopher Lueg. Considering collaborative filtering as groupware: Experiences and lessons learned. In *Proceedings of the Second International Conference on Practical Applications of Knowledge Management (PAKM'98)*, pages 16.1–16.6, Oktober 1998.

[10] Christopher Lueg. Supporting situated actions in high volume conversational data situations. In Clare-Marie Karat, Arnold Lund, Joëlle Coutaz, and John Karat, editors, *Proceedings of the Annual ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'98)*, pages 471–479. ACM Press, April 1998.

[11] David Maltz and Kate Ehrlich. Pointing the way: Active collaborative filtering. In Irvin R. Katz, Robert Mack, Linn Marks, Mary Beth Rosson, and Jakob Nielsen, editors, *Proceedings of the Annual ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'95)*, pages 202–209. ACM Press, May 1995.

[12] David A. Maltz. Distributing information for collaborative filtering on Usenet net news. Master's thesis, Massachusetts Institute of Technology, May 1994.

[13] Larry N. Osborne. Topic development in USENET newsgroups. *Journal of the American Society for Information Science (JASIS)*, 49(11):1010–1016, 1998.

[14] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. GroupLens: An open architecture for collaborative filtering of netnews. In Richard Furuta and Christine Neuwirth, editors, *Proceedings of the International Conference on Computer Supported Cooperative Work (CSCW'94)*, pages 175–186. ACM Press, October 1994.

[15] Paul Resnick and Hal R. Varian. Recommender systems. *Communications of the ACM*, 40(3):56–58, March 1997.

[16] Badrul M. Sarwar, Joseph A. Konstan, Al Borchers, John Herlocker, Brad Miller, and John Riedl. Using filtering agents to improve prediction quality in the GroupLens research collaborative filtering system. In *Proceedings of the International Conference on Computer Supported Cooperative Work (CSCW'98)*, pages 345–354. ACM Press, November 1998.

[17] Upendra Shardanand and Pattie Maes. Social information filtering: Algorithms for automating "word of the mouth". In Irvin Katz et al., editor, *Proceedings of the Annual ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'95)*, pages 210–217. ACM Press, May 1995.

[18] Curt Stevens. *Knowledge-Based Assistance for Accessing Large, Poorly Structured Information Spaces*. PhD thesis, Graduate School of the University of Colorado, USA, 1992.

[19] Loren Terveen, William C. Hill, Brian Amento, David McDonald, and Josh Creter. Building task-specific interfaces to high volume conversational data. In *Proceedings of the Annual ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'97) Conference Proceedings*, pages 226–233. ACM Press, 1997.