

GroupLens: An Open Architecture for Collaborative Filtering of Netnews

Paul Resnick*, Neophytos Iacovou**, Mitesh Suchak*, Peter Bergstrom**, John Riedl**

* MIT Center for Coordination Science
Room E53-325
50 Memorial Drive
Cambridge, MA 02139
617-253-8694
Email: presnick@mit.edu

** University of Minnesota
Department of Computer Science
Minneapolis, Minnesota 55455
(612) 624-7372
Email: riedl@cs.umn.edu

ABSTRACT

Collaborative filters help people make choices based on the opinions of other people. GroupLens is a system for collaborative filtering of netnews, to help people find articles they will like in the huge stream of available articles. News reader clients display predicted scores and make it easy for users to rate articles after they read them. Rating servers, called Better Bit Bureaus, gather and disseminate the ratings. The rating servers predict scores based on the heuristic that people who agreed in the past will probably agree again. Users can protect their privacy by entering ratings under pseudonyms, without reducing the effectiveness of the score prediction. The entire architecture is open: alternative software for news clients and Better Bit Bureaus can be developed independently and can interoperate with the components we have developed.

KEYWORDS: Collaborative filtering, information filtering, electronic bulletin boards, social filtering, Usenet, netnews, user model, selective dissemination of information.

INTRODUCTION

Computer networks allow the formation of interest groups that cross geographical barriers. Bulletin boards have been an important mechanism for that. Rather than addressing an article directly to a known set of people, the writer posts it in a newsgroup, a public place available to anyone interested in the topic. The Usenet netnews system creates the illusion of a single bulletin board available anywhere in the world. It propagates articles so that, with some delays, an article posted from anywhere in the world is available to everyone else.

Recent counts indicate that there are more than 8000

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association of Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.

CSCW 94- 10/94 Chapel Hill, NC, USA
© 1994 ACM 0-89791-689-1/94/0010..\$3.50

newsgroups, with an average traffic of more than 100 MB per day¹. The newsgroups carry announcements, questions, and discussions. In a discussion, often called a thread, one article induces replies from several others, each of which may also induce replies. The January 24, 1994 estimates of netnews participation indicate that more than 140,000 people posted articles in the previous two weeks. There are many more "lurkers" who read but do not post articles. Clearly, a lot of people are getting value from these bulletin boards.

In fact, netnews' rapid broadcast nature and widespread readership has reshaped the way the computing community works. System administrators depend on netnews to keep in touch with the latest development work, the latest security holes, and the latest bug fixes. Researchers depend on netnews as a way of keeping up-to-date on new research directions and important results in between conferences. Many others use netnews just to keep in touch with other people around the world, to learn about new books, new recipes, new music, and what life in other cities is like. Over the years netnews has become a principal medium for sharing among computer users.

Even so, the experience of using netnews is not completely satisfying. Almost everyone complains that the signal to noise ratio is too low. Writers cannot easily tell whether their comments are valued, except by the vocal few who post responses. Some seem not to care about reader interest, only about their own right to write. Moreover, tastes differ, so that no one article will appeal to all the readers of a newsgroup. Each reader ends up sifting through many news articles to find a few valuable ones. Often, readers find the process too frustrating and stop reading netnews altogether.

Netnews provides two mechanisms that help readers limit their attention to articles likely to interest them. First, the division of the bulletin board into newsgroups allows

¹ See the newsgroup news.lists for these and other Usenet statistics

⊕ readers to focus on a few topics. When the number of postings in a newsgroup gets too large, it is often split into two or more newsgroups with identifiable subtopics. Second, some newsgroups are moderated. Attempted postings to these newsgroups are automatically forwarded to the moderator, who decides whether or not they belong in the newsgroup. Usenet propagates only those articles that receive the moderator's stamp of approval.

⊖ In addition, software packages for reading netnews (hereafter referred to as news clients) provide other mechanisms that ease readers' burdens. First, most news clients display a summary of the author and subject line for each message in a newsgroup. The user then indicates which articles she would like to read. Second, most news clients display all of the articles in a particular discussion thread together. Some initially show only the first article in each thread, allowing users to quickly peruse the current discussion topics. Third, some news clients provide "kill files." A kill file identifies text strings that are not interesting to a particular user. If a user puts the subject line of an article into the kill file, no further articles on that subject will be displayed. If a user puts the author's name into a kill file, no further articles from that author will be displayed. Finally, some news readers provide string search facilities. If the user is particularly interested in articles that mention "collaborative filtering," the news client can find them.

GroupLens provides a new mechanism to help focus attention on interesting articles. It draws on a deceptively simple idea: people who agreed in their subjective evaluation of past articles are likely to agree again in the future. After reading articles, users assign them numeric ratings. GroupLens uses the ratings in two ways. First, it correlates the ratings in order to determine which users' ratings are most similar to each other. Second, it predicts how well users will like new articles, based on ratings from similar users. The heart of GroupLens is an open architecture that includes news clients for entry of ratings and display of predictions, and rating servers for distribution of ratings and delivery of predictions.

Related Work

The general problems of information overload and low signal to noise ratio have received considerable attention in the research literature. We use the term information filtering generically to refer both to finding desired information (filtering in) and eliminating that which is undesirable (filtering out), but related work also appears under the labels of information retrieval and selective dissemination of information [2]. In addition, research on agents [12, 13], user modeling [1, 9], knowbots [8], and mediators [21] has explored semi-autonomous computer programs that perform information filtering on behalf of a user.

Malone et al. [13] describe three categories of filtering techniques, cognitive, social, and economic, based on the information sources the techniques draw on in order to predict a user's reaction to an article. The three categories provide a useful road map to the literature.

Cognitive, or content-based filtering techniques select documents based on the text in them. For example, the kill files and string search features provided by news clients perform content filtering. Even the division of netnews into newsgroups is a primitive example, since a reader restricts his attention to those articles with a particular text string in their "newsgroup:" field.

Other content-based filtering techniques could potentially be used as well. The profile of which texts to include or kill could be more complex than a collection of character strings. For example, strings could be combined with the Boolean operators AND, OR, and NOT. Alternatively, the profile could consist of weight vectors, with the weights expressing the relative importance of each of a set of terms [4, 5, 16].

Some content filtering techniques update the profiles automatically based on feedback about whether the user likes the articles that the current profile selects. Information retrieval research refers to this process as relevance feedback [17]. The techniques for updating can draw on Bayesian probability [2], genetic algorithms [18], or other machine learning techniques.

Social filtering techniques select articles based on relationships between people and on their subjective judgments. Placing an author's name in a kill file is a crude example. More sophisticated techniques might also filter out articles from people who previously co-authored papers with the objectionable person.

Collaborative filtering, based on the subjective evaluations of other readers, is an even more promising form of social filtering. Human readers do not share computers' difficulties with synonymy, polysemy, and context when judging the relevance of text. Moreover, people can judge texts on other dimensions such as quality, authoritativeness, or respectfulness. A moderated newsgroup employs a primitive form of collaborative filtering, choosing articles for all potential readers based on evaluations by a single person, the moderator.

The Tapestry system [6] makes more sophisticated use of subjective evaluations. Though it was not designed to work specifically with netnews, it allows filtering of all incoming information streams, including netnews. Many people can post evaluations, not just a single moderator, and readers can choose which evaluators to pay attention to. The evaluations can contain text, not just binary accept/reject recommendations. Moreover, filters can combine content-based criteria and subjective evaluations. For example, a reader could request articles containing the word "CSCW" that Joe has evaluated and where the evaluation contains the word, "excellent".

Our work is similar in spirit to Tapestry but extends it in two ways. First, Tapestry is a monolithic system designed to share evaluations within a single site. We share ratings between sites and our architecture is open to the creation of

new news clients and rating servers that would use the evaluations in different ways. Second, Tapestry does not include any aggregate queries. The rating servers we have implemented aggregate ratings from several evaluators, based on correlation of their past ratings. A reader need not know in advance whose evaluations to use and in fact need not even know whose evaluations are actually used. In GroupLens, ratings entered under a pseudonym are just as useful as those that are signed.

Maltz has developed a system that aggregates all ratings of each netnews article, determining a single score for each [14]. By contrast, GroupLens customizes score prediction to each user, thus accommodating differing interests and tastes. In return for its reduced functionality, Maltz's scheme scales better than ours, because rating servers can exchange summaries of several users' ratings of an article, rather than individual ratings.

The subjective evaluations used in collaborative filtering may be implicit rather than explicit. Read Wear and Edit Wear [7] guide users based on other users' interactions with an artifact. The GroupLens news clients monitor how long users spend reading each article but our rating servers do not yet use that information when predicting scores.

Economic filtering techniques select articles based on the costs and benefits of producing and reading them. For example, Malone argues that mass mailings have a low production cost per addressee and should therefore be given lower priority. Applying this idea to netnews, a news client might filter out articles that had been cross-posted to several newsgroups. More radical schemes could provide payments (in real money or reputation points) to readers to consider articles and payments to producers based on how much the readers liked the articles.

Stodolsky has proposed a scheme that combines social and economic filtering techniques [19]. He proposes on-line publications where the publication decision ultimately rests with the author. During a preliminary publication period, other readers may post ratings of the article. The author may then withdraw the article, to avoid the cost to his reputation of publishing an article that is disliked.

Outline

The GROUPLENS section of the paper describes the GroupLens architecture and its evolution. The ONGOING EXPERIMENTATION section describes a larger scale test of the architecture that is in preparation. The SOCIAL IMPLICATIONS section addresses social changes in the use of Netnews that may be precipitated by GroupLens.

GROUPLENS

GroupLens is a distributed system for gathering, disseminating, and using ratings from some users to predict other users' interest in articles. It includes news reading clients for both Macintosh and Unix computers, as well as "Better Bit Bureaus," servers that gather ratings and make predictions. Both the overall architecture and particular

components have evolved through iterative design and pilot testing to meet the following goals: Criteria

Openness: There are currently dozens of news clients in common use, each with a strong following among its user community. Any or all of these clients can be adapted to participate in GroupLens. GroupLens also allows for the creation of alternative Better Bit Bureaus that use ratings in different ways to predict user interest in news articles.

Ease of Use: Ratings are easy to form and communicate, and predictions are easy to recognize and interpret. This minimizes the additional burden that collaborative filtering places on users.

Compatibility: The architecture is compatible with existing news mechanisms. Compatibility reduces user overhead in taking advantage of the new tool, and simplifies its introduction into netnews.

Scalability: As the number of users grows, the quality of predictions should improve and the speed not deteriorate. One potential limit to growth will be transport and storage of the ratings, if GroupLens grows very large.

Privacy: Some users would prefer not to have others know what kinds of articles they read and what kinds they like. The Better Bit Bureaus in GroupLens can make effective use of ratings even if they are provided under a pseudonym.

Overview

Usenet consists of Internet sites as well as UUCP sites. Typically a site will declare a machine to act as its news server. Users at each site invoke news clients on their computers and connect to the news server in order to retrieve news articles. Users can also write new articles and post them to the news server through their news clients.

When a user posts an article, it travels from the news client where the article is composed to the local news server and from there to news servers at nearby sites. After leaving the originating site, an article propagates throughout Usenet, hopping from site to site. Since there is no centralized coordination of the distribution process, an article may arrive at a site via more than one route. Because articles have globally unique identifiers, however, and are never altered once they are posted, any site can recognize a duplicate copy of an article and avoid passing it on. Lotus Notes uses a similar distribution process [10]. The netnews architecture is summarized in Figure 1.

GroupLens adds one new type of entity to the netnews architecture, Better Bit Bureaus, as shown in Figure 2. The Better Bit Bureaus provide scores that predict how much the user will like articles, and gather ratings from news clients after the user reads the articles. The Better Bit Bureaus also use special newsgroups to share ratings with each other, to allow collaborative filtering among users at different sites. The remainder of this section traces the processes of rating creation, distribution, and use and describes how they meet

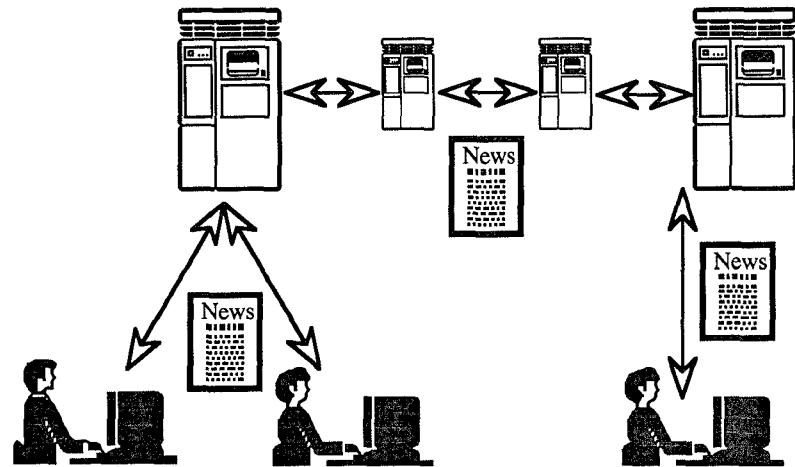


Figure 1: The netnews architecture. News articles hop from news server to news server. A news client connects to the news server at its site and presents articles to users.

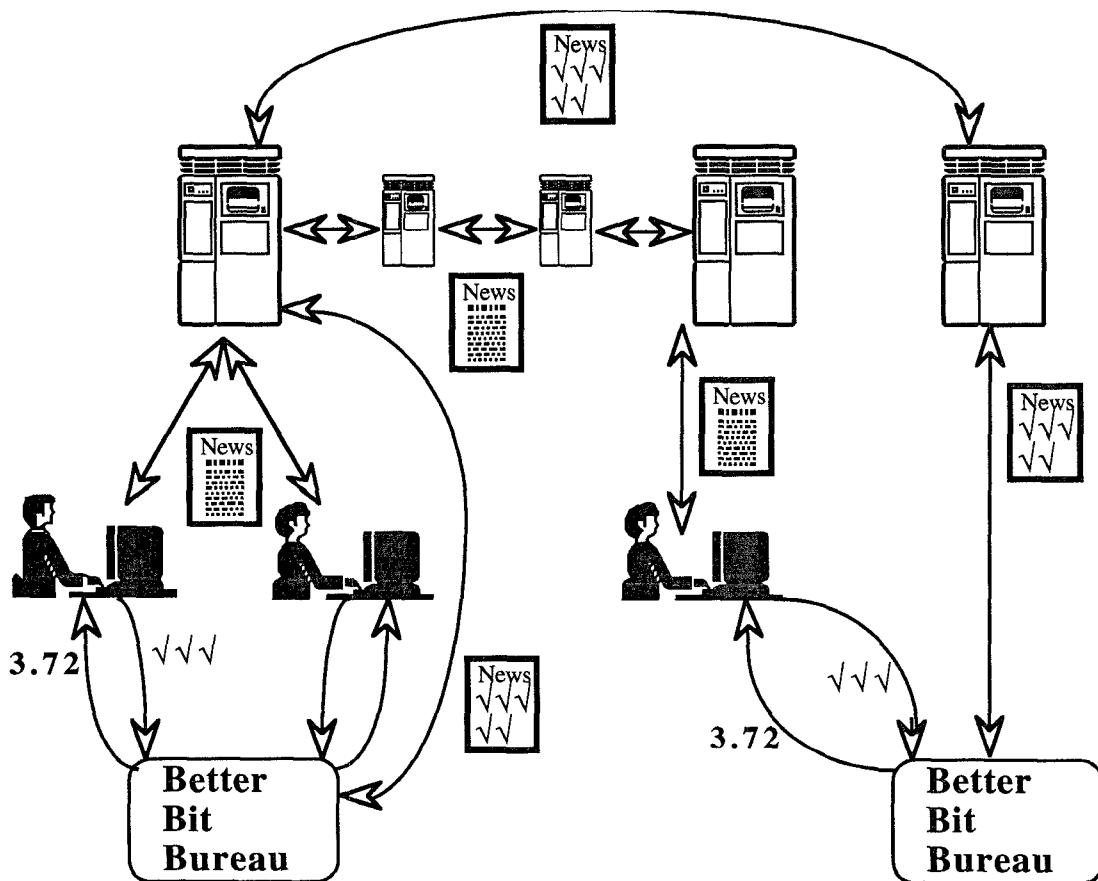


Figure 2: The GroupLens architecture. Better Bit Bureaus collect ratings from clients, communicate them by way of news servers, and use them to generate numeric score predictions that they send to clients. Clients connect to a local news server, and can connect to a Better Bit Bureau that uses the same or a different news server.

the design goals of openness, ease of use, compatibility, scalability, and privacy.

Entering Ratings

In GroupLens, a rating is a number from 1 to 5, optionally supplemented by the number of seconds which the user spent reading the article. Users are encouraged to assign ratings based on how much they liked the article, with 5 highest and 1 lowest. The user chooses a pseudonym to associate with her ratings that may be different from the name she uses for posting news articles. This preserves the ability to detect that two ratings came from the same person, while preventing detection of exactly who that person is.

The GroupLens choice of the form and meaning of ratings is only one possibility in a rich design space. There are many possible dimensions along which to rate articles: interest in subject, quality of writing, authoritativeness of the author, etc. Rather than a single composite rating, separate ratings on several dimensions could be solicited from readers. Free text ratings could be entered rather than numbers. Readers could be asked to predict how well they think other readers will like an article rather than report how much they themselves liked it. Ratings could be restricted only to positive, or only to negative evaluations. The degree of privacy could also be varied, from completely anonymous to authenticated signatures.

In fact, an earlier implementation of a Macintosh news client [20] employed ratings with quite a different form than

the current GroupLens architecture. Users entered only endorsements, positive ratings, on the assumption that since the signal to noise ratio in netnews is so low it is only important to point out the good articles. Readers endorsed articles that they thought others in a known small group would like. Finally, readers signed endorsements with their real names, allowing other people to select all the articles endorsed by a particular friend.

A pilot test of that earlier endorsement mechanism at a Schlumberger research lab indicated that a group of seven people may not be large enough to get the full available benefit of collaborative filtering. As we contemplated a much larger group size, we believed that some users would be less willing to sign their ratings and that it would become increasingly difficult for users to know what articles others in the group would like.

The pilot test also reinforced the importance of making it as easy as possible to enter endorsements. To make an endorsement, a user had to select from a pull-down menu, wait for a window to open up, optionally enter text in the window, and then close it. While the whole process took only a matter of seconds if the user entered no text, it was still significantly longer than it normally takes to go on to the next article.

We have taken care in the GroupLens system to make entry of ratings as easy as possible. We have modified three news clients, Emacs Gnus and NN for UNIX machines and NewsWatcher for Macintoshes. In each case, entry of a

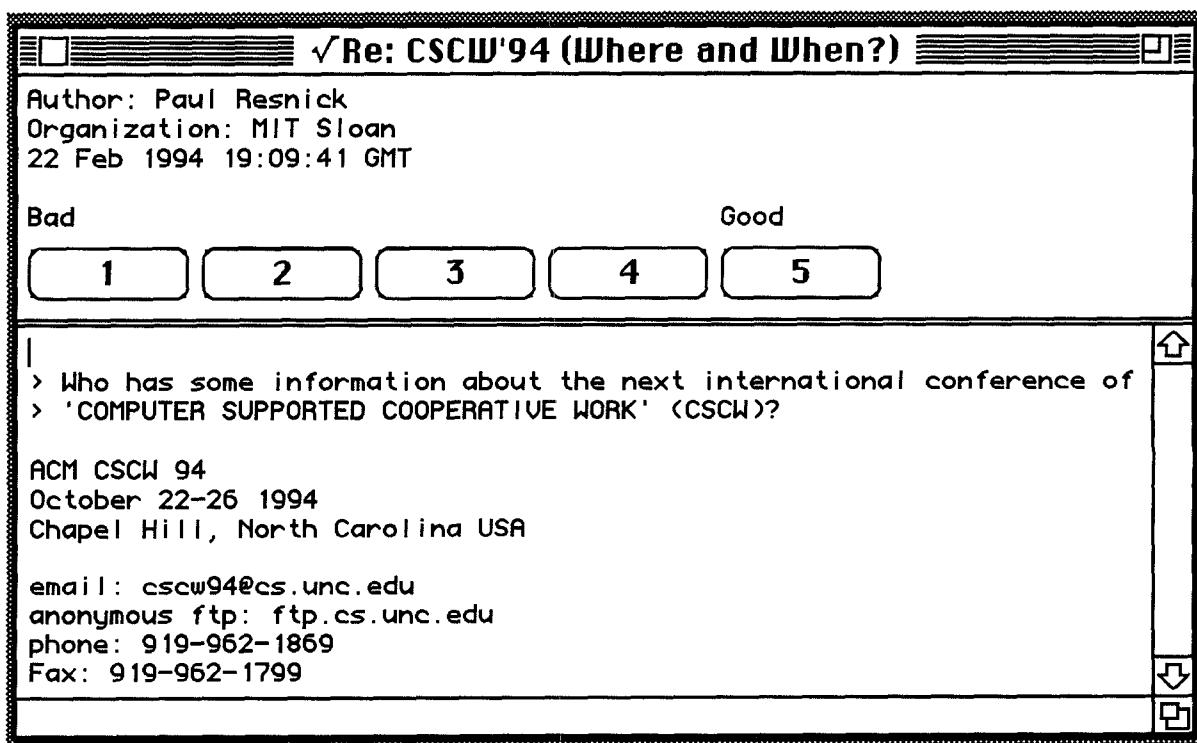


Figure 3. Reading an article with the modified NewsWatcher client. The user can click on one of the five ratings buttons with the mouse, or type a number from 1 to 5 on the keyboard.

rating fits into the overall paradigm of the news client. For example, in the modified NewsWatcher, the numbers 1 to 5 appear as selectable buttons any time a user reads an article (Figure 3), and the user can also type a number as a keyboard shortcut for those buttons. In Gnus, no buttons are displayed, but readers still type the ratings directly. With NN, readers first type the letter 'v' (to enter into "rating mode") and then the rating.

The GroupLens architecture requires only that ratings be reported on a 1 to 5 scale, not that they be displayed by news clients on that scale. To make the rating scale easy for students to understand, the NN and Gnus clients accept letter grades rather than numbers. When reporting the ratings to the Better Bit Bureau, they translate 'a' to 5, 'b' to 4 and so on. Other news clients could allow more gradations of ratings (e.g., 1 to 100) and report them as fractions between 1 and 5.

Distributing Ratings

GroupLens does not interfere with the Usenet propagation scheme at all. On the contrary, it relies upon it heavily. The Better Bit Bureau packages one or more ratings into a news article, following the format in Figure 4, and posts it to a news server. This allows GroupLens to take advantage of the Usenet propagation scheme. Over the years Usenet has demonstrated its ability to propagate articles to every other Usenet site, even as the number of news servers has grown dramatically. Rating servers could exchange ratings directly, through internet or UUCP links, but they would have to reimplement many of the propagation features already found in Usenet.

The message format we have defined allows several ratings to be batched in a single article. Each rating is just one line of text, while each Usenet netnews article requires several lines of headers. Thus, packaging several ratings in one article can save a considerable amount of overhead. Our Better Bit Bureaus (BBBs) batch at the session level (i.e., all ratings entered by a user during a reading session go into one ratings article). Other batching policies, such as all ratings from a site over the last hour, could be implemented.

Ratings are posted in newsgroups dedicated solely to ratings articles. One natural configuration is to set up a parallel

"ratings transport" newsgroup for each "normal" Usenet group. One deficiency of this approach is that if a rating article contains several ratings, it may have to be cross-posted to many ratings newsgroups. Another deficiency is that it requires news servers to carry a large number of new newsgroups devoted solely to ratings, which may increase administrative overhead. Currently, our BBBs post all ratings in a single newsgroup.

To facilitate the initial spread of GroupLens, users can participate even if their local news servers do not carry the ratings newsgroup and even if their local site administrators have not set up Better Bit Bureaus. The GroupLens architecture permits this by allowing users to connect to a remote BBB. The left side of Figure 2 illustrates a local BBB that posts ratings articles to the same news server that the clients connect to. The right side of Figure 2 illustrates a client connecting to a remote BBB that propagates ratings articles through a different news server.

Predicting Scores

The Better Bit Bureaus (BBBs) predict how much readers will like articles. While content filters would make predictions based on the presence or absence of words in the articles, the BBBs in GroupLens use the opinions of other people who have already rated the articles. If no one has read an article, the BBBs are unable to make predictions about it.

When ratings for an article are available, they are unlikely to be uniform, due to differences of opinion and goals among the raters. A BBB combines the different ratings to produce a predicted score. Moreover, additional readers are likely to have different opinions about the article. A BBB thus might use the same ratings to predict different scores for different readers, by changing the relative weight given to the ratings. |

When predictions are on the same scale as ratings, prediction can be modeled as matrix filling, where the columns are people, the rows are articles, and the cells contain the ratings that people have posted, as shown in Figure 5. Many of the cells of the matrix are empty, because readers have not yet examined those articles or have elected not to rate them. A BBB predicts scores for missing cells before the readers examine the corresponding articles.

From: MIT GroupLens Better Bit Bureau			
Subject: Ratings; please ignore			
Message-ID: <771185369@gilder.mit.edu>			
Groups_Rated: news.adin.policy, news.groups			
Raters: [Pseudo1]			
<MATT.94May19124319@physics5.berkeley.edu>		[Pseudo1] 1 12	news.adin.policy
<fred_scq2FF6.Mtt@netcom.com>		[Pseudo1] 2 7	news.groups

Figure 4: A sample ratings article. Each line in the body of the article contains a rating of one article by one person. The five fields on each line are the id of the article, the pseudonym of the rater, a rating, the number of seconds the reader spent examining the article before rating it, and the newsgroups the article is in. The time count is optional. Additional keyword identified fields can also be included at the end of line.

message #	Ken	Lee	Meg	Nan
1	1	4	2	2
2	5	2	4	4
3			3	
4	2	5		5
5	4	1		1
6	?	2	5	?

Figure 5: a sample matrix of ratings.

All the scoring methods we have implemented are based on the heuristic that people who agreed in the past are likely to agree again, at least on articles in the same newsgroup. This heuristic will mislead on occasion, but preferences for most kinds of articles are likely to be fairly stable over time.

(1)

To implement this heuristic, our BBBs first correlate ratings on previous articles to determine weights to assign to each of the other people when making predictions for one of them. Then, they use the weights to combine the ratings that are available for the current article. We have investigated several techniques for correlating past behavior and using the resultant weights, based on reinforcement learning [12], multivariate regression, and pairwise correlation coefficients that minimize linear error or squared error.

We illustrate one of the correlation and prediction techniques by computing Ken's predicted score on article 6, the last row of the matrix. First, we compute correlation coefficients [15], weights between -1 and 1 that indicate how much Ken tended to agree with each of the others on those articles that they both rated. For example, Ken's correlation coefficient with Lee is computed as:

$$r_{KL} = \frac{\text{Cov}(K, L)}{\sigma_K \sigma_L}$$

$$= \frac{\sum_i (K_i - \bar{K})(L_i - \bar{L})}{\sqrt{\sum_i (K_i - \bar{K})^2} \sqrt{\sum_i (L_i - \bar{L})^2}}$$

$$= \frac{-2 - 2 - 2 - 2}{\sqrt{10} \sqrt{10}} = -0.8$$

Pearson's
coefficient
Correlation

-1 : disagree

0 : agree

0 : not correlated

In the formula above, \bar{K} is the average of Ken's ratings. All the summations and averages in the formula are computed only over those articles that Ken and Lee both rated. We have conveniently arranged for \bar{K} and \bar{L} to be 3 in this example, but that need not be true in practice.

Similarly, Ken's correlation coefficient with Meg is +1 and with Nan is 0. That is, Ken tends to disagree with Lee ($r_{KL} = -0.8$) and agree with Meg ($r_{KM} = +1$). His ratings are not correlated with Nan's.

To predict Ken's score on the last article in the matrix, take a weighted average of all the ratings on article 6 according to the following formula:

$$\text{prediction of } K_6^{\text{pred}} = \bar{K} + \frac{\sum_{\text{raters}} r_{KJ} \times \text{Ken's rating}}{\sum_{\text{raters}} |r_{KJ}|} = \frac{3 + \frac{2r_{KM} - r_{KL}}{|r_{KM}| + |r_{KL}|}}{|1| + |-0.8|} = 4.56$$

article 6
Meg Lee
avg of J

This is a reasonable prediction for Ken, since the article received a high rating from someone who agreed with him in the past and a low rating from someone who disagreed. Carrying through similar calculations for Nan yields a lower prediction of 3.75. Since Nan had partial agreement with Lee in the past, Lee's low rating for the article partially cancels out the high ratings that Meg gave it.

The score prediction system is robust with respect to certain differences of interpretation of the rating scale. If two users are perfectly correlated, but one user gives only scores between 3 and 5 and the other only scores between 1 and 3, a 5 score from the first user will result in a prediction of 3 for the second. If two users would be perfectly correlated, but the first mistakenly thinks 1 is a good score and 5 is bad, the two will be negatively correlated and a 1 score from the first will result in a prediction of 5 for the second. This leads to a clear explanation to the user of how to assign ratings: assign the rating you wish GroupLens had predicted for this article.

Allen's study of five subjects' preferences for newswire articles [1] found very small correlations between subjects, thus calling into question our basic assumption that people who agreed in the past are likely to agree again. It may be, however, that a larger sample of subjects would have yielded some pairs with larger overlaps in their ratings. More importantly, it may be that pairs of people will share interests in some topics but not others. Two people may agree in their evaluations of technical articles, but not jokes. Our BBBs keep separate rating matrices for each newsgroup.

One hopes that the accuracy of the predictions improve as the BBB has more past ratings to use in computing correlations. Four people at the University of Minnesota participated in a pilot test of an earlier version, using a slightly different scoring function. While all four participants reported that the predicted scores eventually matched their interests fairly closely, they did observe that there was a start-up interval before the predictions were very useful. Further experiments and analysis are necessary to determine just how long the start-up interval is likely to be for each new user.

It seems likely that better scoring mechanisms can be developed. In addition to better matrix filling techniques, it may be helpful to use both others' ratings and the contents

proposition of hybrid technique.

of articles in making predictions. It may also be helpful to take into account the time people spent reading articles before rating them, information collected but not used by our BBBs.

implicit feedback proposition

Fortunately, the GroupLens architecture is open: anyone can implement an alternative BBB so long as it posts ratings articles in the format described above and communicates with clients the same way that our BBBs do. We hope that the development of alternative BBBs will become an active area for future research. As we describe below, our next pilot test should yield rating sets that we will make available to others who wish to evaluate alternative scoring algorithms.

Using Ratings

It is up to the news client how best to use the scores generated by a BBB. Some may filter out those articles with scores below a threshold. Some may sort the articles based on the scores.) Others may simply display the scores, numerically or graphically) In keeping with the ease of use design goal, developers should modify each news client in a manner consistent with that client's overall design.

One trend in news clients is to display a summary of the unread articles in a newsgroup. Each line of the summary contains information about one article, typically the author, the subject line and the length. A user browses the summary and requests display of the full text of those articles that seem interesting. All three of the news clients we modified use this display technique.

The three modified clients we implemented make slightly different uses of the scores in the summary display. The modified NN client displays articles in the same order a regular NN client does, namely the order in which the

articles arrived at the news server. It merely adds an additional column containing the predicted scores. In the first version of this client, the scores were displayed numerically.

The modified Gnus client uses the predicted scores to alter the order of presentation of articles in the summary. Gnus clusters articles by thread. The modified Gnus client sorts the threads based on the maximum predicted score over the articles in the thread. Within each thread, however, articles are still displayed in chronological order, to preserve the flow of discussion. As in the modified NN, the scores are displayed in an additional column in the summary.

The Minnesota pilot test included users of both the Gnus and NN clients. As expected, participants tended to believe that the sorting and display mechanisms of their own news reader were best, but all were glad to see the score predictions incorporated into that standard format.

Several users, however, noticed that it was somewhat difficult to visually scan the predictions to find the high ones. A revised version of the NN client (Figure 6) rounds off to the nearest integer and reports that as a letter grade (A-E), a scale familiar to students at U.S. Universities.

The modified NewsWatcher client displays the predicted scores as bar graphs rather than numbers (Figure 7), making it easier to visually scan for articles with high scores (longer bars). Otherwise, it follows the conventions of the original NewsWatcher client. Articles are grouped into threads and the summary display initially shows header lines only for the first article in each thread. Users can twist down the triangle associated with a thread to see the header lines for the rest of the articles.

NN client :: comp.multimedia				NN client :: 358 of 2229 151 FEHLERDETEKTION			
a.Alois Bock	11	>*** 7 RASH STATEMENTS ***					
b.Bernhard Schwall	9	Driver for ATI Graphics Ultra Pro/Plus					
c.Kuny Terry	20	Question: Video Input Boards					
d.Francois Zarroca	8	C SB16 mod-editor ???					
e.Patrick Corbett	9	B REALLY good encyclopedia on CD_ROM?					
f.Lesley Davidow	26	A >					
g.Isa Helderma	9	A >>					
h.Dave Skwarczek	32	Cyberfest.594					
i.hkaplan@woods	9	Hypercard????					
j.eruffing@bcrvm1	5	B FTP Sites for JPG, GIF, TIF, BMP, PCX, TGA					
k.Aarts ing. R.M.	22	B MM-standard what is the latest?					
l.Kees de Groot	31	B Manipulating Spatial Objects and Relations					
m.Steven Koster	24	A Line Audio in to Quadra 700?					
n.Isa Helderma	19	Need help with MM Director QuickTime Lingo commands					
15:28 -- SELECT -- help: -- 15 -- level 3 --							

Figure 6: The modified NN client. The third column displays the number of lines in the article. The fourth column displays the score predictions as letter grades, translated from the numeric predictions that the Better Bit Bureau makes (5=A, 4=B, etc.). When no one has evaluated an article, no prediction is made.

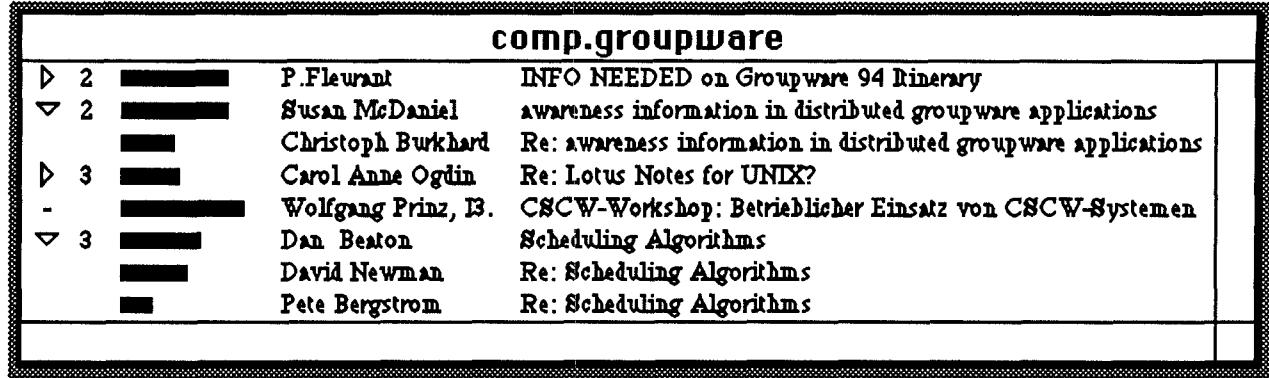


Figure 7: The modified NewsWatcher client displays predicted scores as bar graphs. Disclaimer: the scores were randomly generated for demonstration purposes. In practice, we would expect articles by Pete Bergstrom (one of the authors of this paper) to have much higher predicted scores.

Scale Issues

Further research is needed to understand how performance will change as the scale increases. In the case of GroupLens, there are several relevant performance measures: prediction quality, user time, Better Bit Bureau compute time and disk storage, and network traffic.

The first measure is the quality of score predictions. We expect prediction quality to increase as the number of users increases, since more data will be available to the prediction algorithm.

Another measure is how long users have to wait to post ratings and receive predictions. In an earlier version of GroupLens, the functions of the BBB were incorporated in the news client itself. One major advantage of the separate BBB is that it can pre-fetch ratings and pre-compute predictions rather than computing them when the user starts the news client. Thus, user time should remain roughly constant as GroupLens grows, even if it takes more CPU time to compute scores.

For many possible prediction formulas CPU time will grow even faster than linearly with increases in the number of users. To reduce CPU time, BBBs could use only a part of the ratings matrix, trading off compute time against quality of predictions.

Even though each rating is short, each news article might be read and rated by many raters, so the total volume of ratings could exceed the volume of news. To minimize storage requirements, BBBs may employ algorithms that use and discard ratings as they arrive, rather than storing them.

Three basic techniques could reduce network traffic: reduce the size of the ratings, reduce the number of ratings, and reduce the number of places where each rating is sent. Our BBBs batch several ratings in a single article, a first step toward reducing the amount of storage per rating, but

further compression is possible. The number of ratings could be reduced by limiting the total number of ratings per article or the number of ratings from users with similar profiles.

The separation of the BBBs from the news clients in the GroupLens architecture reduces the number of destinations for each rating: each news client receives only score predictions rather than all the individual ratings that contribute to those predictions.

The number of destinations for each rating could be further reduced by sending ratings to some BBBs but not others. For example, BBBs could be clustered, based on geography or interest, and exchange ratings only within clusters. The size of each cluster must be small enough to limit the amount of ratings information distributed, but large enough to provide an effective peer group. The table below estimates daily network traffic for various cluster sizes assuming each user rates 100 articles per day and each rating requires approximately 100 bytes. For comparison purposes, the current netnews traffic is around 100MB per day.

<u>Cluster size</u>	<u>Daily ratings traffic</u>
100 users	1 MB
50,000 users	100 MB
1,000,000 users	10 GB

Summary of GroupLens Architecture

The heart of GroupLens is an open architecture for distributing ratings. The architecture specifies the format of ratings produced in batches by BBBs, the propagation of the ratings by Usenet, and the interface for delivering predictions and ratings between news clients and BBBs. Otherwise, the architecture is completely open. BBBs and

news clients can be freely substituted, providing an environment for experimentation in predicting ratings and in user interfaces for collecting ratings and presenting predictions.

ONGOING EXPERIMENTATION

Both of the previous pilot tests, at Schlumberger and the University of Minnesota, involved only local sharing of ratings. These tests led to improvements in both the overall architecture and the user interfaces of news clients, as discussed already. The next step is a larger scale, distributed test, that we plan to carry out this summer. We have established a newsgroup on the news servers at MIT and Minnesota and two (slightly different) Better Bit Bureaus that communicate ratings through that newsgroup.

The test is not designed to demonstrate that people prefer to read netnews with our collaborative filters than without them. We believe that such an evaluation should wait for at least one more iterative design cycle. Rather, the goals are to identify any unexpected scaling issues that may arise and to gather a data set that will be useful in evaluating alternative score prediction algorithms.

The primary benchmark of any algorithm's effectiveness will be its ability to predict values that have been deleted from a rating matrix. At first glance, it might seem that any large set of ratings would be useful in creating such a benchmark. Upon closer inspection, however, complete ratings matrices are much more valuable than sparse ones. For example, suppose that users read and rate only a small number of articles, based on score predictions they receive from BBB X. If users read different articles, this generates a sparse matrix of ratings. Now suppose that we wish to compare X to an alternative, Y, that predicts different scores for the users. We can compare Y's and X's predictions on those articles that users read, but the sample is biased. Perhaps with Y's scores, the users would have read other articles and liked them.

To allow unbiased comparisons, we are asking each of the participants in the next pilot test to read and rate all the articles in a training set. The training set will contain a number of articles from each of the newsgroups that will be included in the test. Since users will contribute ratings under a pseudonym, we will be able to share the ratings in this training set with other researchers. In addition, we will retain the full texts of the articles in the training set. That will enable evaluation of BBBs that perform content filtering, or a combination of content filtering and collaborative filtering, as well as those that use only other users' ratings.

SOCIAL IMPLICATIONS

Collaborative filtering may introduce many social changes in the already rapidly evolving Netnews community. For example, the utility of moderated newsgroups may decline. New social patterns will have to develop to encourage socially beneficial behaviors, such as reviewing articles that have already received a few low ratings. Finally, if

GroupLens is effective at creating peer groups with shared interests, will those peer groups be permeable or will the global village fracture into tribes?

Changes to Netnews Behaviors

GroupLens has the potential to change Netnews as we now know it. For one thing the quality of articles individual users choose to read should increase. More significantly, as more and more users rely on GroupLens the total number of low-quality articles on Usenet may decrease significantly. Since few people will read such articles, the incentive to post them will decrease. GroupLens may also supplant or supplement other established Netnews behaviors.

Moderated Newsgroups

GroupLens may reduce the need for moderated newsgroups. The advantages of GroupLens over the existing approach are that "moderators" can be groups of people as well as individuals, and that each user can rely on a different moderator rather than having a single moderator for the entire group.

Some newsgroups might choose to use both a moderator and GroupLens. The moderator of a newsgroup will make the initial pass through the article submissions. Peer ratings would then allow further filtering.

Newsgroup Splits

Currently, newsgroups start off with broad topics and split into narrower topics as traffic increases. For example, the newsgroup rec.sport.football eventually split into the subgroups australian, canadian, rugby, pro, college, fantasy, misc, and one for each team in the NFL. These splits are a form of content filtering, initiated and managed by the users.

GroupLens users may find that many such splits are less important, and in some cases undesirable. Over the course of time users will find themselves reading only the subset of the newsgroup they are most interested in, as they correlate with a peer group with similar interests. Splits of interest between groups of users will appear naturally, with no additional user or administrative effort. Allowing the splits to happen through GroupLens rather than through explicit content filtering allows more cross-pollination of general interest articles. For instance, interesting articles posted by Bills fans about an upcoming football game against the Cowboys would also reach Cowboys fans with GroupLens, but would not if the articles were posted in the more specialized newsgroup rec.sport.football.bills.

Kill-Files

Kill files are a content filtering mechanism implemented in some news clients. Many users who strongly dislike particular subjects or particular authors, however, do not use kill files because they find the mechanism complicated and cumbersome. GroupLens might be an easier means to the same end. A user's peer group will give such articles low ratings, so only a few users will have to read them.

Incentives

Individuals put additional effort, albeit a modest amount, into providing ratings through GroupLens. These ratings provide benefit to other users who can use them to select interesting articles. It's a two-way street: everyone can be both a producer and a consumer of ratings.

When someone reads and rates an article, there is an incentive to provide honest ratings, because dishonest ratings will cause the BBB to make poor future predictions for that user. On the other hand, there is no incentive to rate articles at all. On the contrary, there is an incentive to wait for others' ratings rather than read and rate an article oneself. A certain amount of altruism or guilt may cause most people to "do their share" of rating, but fewer than the socially optimal number of ratings are likely to be produced.

The four-person Minnesota pilot test included a high-volume newsgroup, *rec.arts.movies*. The volume of articles was so high that each participant was unwilling to read a one-quarter share of the total daily volume. The newsgroup was quickly dropped from the test. It may be that a larger user population would generate ratings even for a high-volume list such as *rec.arts.movies*, but it is harder to draw on a "do-your-share" mentality when collaborating with larger groups of people.

There are other, more subtle incentive problems that can arise as well. For example, there is an asymmetry between the effects of positive and negative ratings. If the first few readers rate an article too highly, others will read the article and give it lower ratings. On the other hand, if the first few ratings of an article are negative, others who would have rated it highly may never look at it because of the initial negative rating.

To avoid this, it may be necessary to provide external incentives to some people to read and rate articles that have initially low ratings. The external incentives could be money, fame, or simply access to others' ratings: those who did not contribute their share of ratings might be denied access to the Better Bit Bureau's predictions.

Global Villages

Present newsgroups, like newspapers and local television shows before them, provide a shared history for their community of readers. With GroupLens, users may choose to read articles only from a small group with whom they share many common interests. Over time this could lead to a fracture of the global village into many small tribes, each forming a virtual community but nonetheless isolated from each other.

Some kind of fracture is inevitable and even desirable, because no user can keep up with the overwhelming volume of news produced each day. The question is whether the subgroups will be closed or permeable. One argument for prognosticating permeability is that many groups will form for a short time and then disband [3].

Another is that many users will participate in several subgroups, providing a mechanism for the best ideas to cross boundaries of interest groups.

CONCLUSION

Shared evaluations are useful in all sorts of activities. We ask friends, colleagues, and professional reviewers for their opinions about books, movies, journal articles, cars, schools, and neighborhoods. Clearly, some form of shared evaluations should also help in filtering electronic information streams such as netnews. It is not yet clear exactly what form those evaluations should take, how they should be collected and disseminated, and how they should be used in selecting articles to read.

GroupLens is one promising approach. A single number gives a composite rating of an article on all dimensions relevant to a particular reader. We have modified three news reading clients to enable easy entry of such numeric ratings. We have also modified the way that the clients display subject lines to include predicted scores based on others' ratings.

Naturally, there will be differences of opinion among readers about particular articles, due to varying interests or quality assessments. To accommodate differences of opinion, not all readers will place equal trust in particular evaluators. The algorithms we have implemented automatically determine how much weight to place on each evaluation, based on the degree of correlation between past opinions of the reader and evaluator. This has the beneficial side effects that readers need not know initially whose evaluations to trust and the evaluators' opinions can become trusted even if the evaluators choose to remain anonymous.

The GroupLens architecture allows new users to connect and new rating servers to come on line, without global coordination. A new user need only use a modified news client and have a connection to a rating server. The user need not convince the administrator of her netnews server to modify the news server, run any additional software, or even to carry any additional newsgroups. A new rating server needs only to get access to a news server that carries the ratings newsgroups.

Moreover, the architecture is open. Anyone who wishes to can modify a news client to allow entry of evaluations or to use predicted scores, so long as the client follows the protocol we have established for communicating with the rating server. Anyone who wishes to improve on the score predictions that our rating servers make can do so. There may be better ways to correlate past evaluations. There may also be ways to use the evaluations in conjunction with content filtering. For example, when correlating past evaluations, the scoring algorithm might consider evaluations only of past articles that are somehow similar to the current one. Our next pilot test should yield a data set that can be used for evaluating alternative prediction methods.

Only further testing can reveal whether GroupLens gathers the right kind of evaluations and uses them in ways that people like. If the simple numeric evaluations turn out to be sufficient, the architecture will scale up to large numbers of rating servers and users. If not, then data from our tests will help develop and evaluate other mechanisms for sharing and using evaluations.

Right now, people read news articles and react to them, but those reactions are wasted. GroupLens is a first step toward mining this hidden resource.

ACKNOWLEDGMENTS

Shumpei Kumon's keynote address at CSCW 92 [11] inspired our investigation of the practical application of reputations to social filtering. Thanks to Lorin Hitt and Carl Feynman for helpful discussions about how to predict scores based on past correlations. Peter Foltz and Sue Dumais generously provided a test rating set generated from one of their experiments on content filtering [5]. Thanks also to Chris Avery, Joe Adler, Yannis Bakos, Erik Brynjolfsson, David Goldberg, Bill MacGregor, Tom Malone, David Maltz, Vahid Mashayekhi, Lisa Spears, Doug Terry, Mark Uhrmacher, and Zbigniew Wieckowski.

REFERENCES

1. Allen, R.B. User Models: Theory, Method, and Practice. *International Journal of Man-Machine Studies*, 32, (1990), pp. 511-543.
2. Belkin, N.J. and Croft, B.W. Information Filtering and Information Retrieval: Two Sides of the Same Coin? *CACM*, 35, 12 (1992), pp. 29-38.
3. Brothers, L., Hollan, J., Nielsen, J., Stornetta, S., Abney, S., Furnas, G. and Littman, M. Supporting Informal Communication via Ephemeral Interest Groups. In *Proceedings of CSCW 92* (1992, New York: ACM), pp. 84-90.
4. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K. and Harshman, R. Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science*, 41, 6 (1990), pp. 391-407.
5. Foltz, P.W. and Dumais, S.T. Personalized Information Delivery: An Analysis of Information Filtering Methods. *Communications of the ACM*, 35, 12 (1992), pp. 51-60.
6. Goldberg, D., Nichols, D., Oki, B.M. and Terry, D. Using Collaborative Filtering to Weave an Information Tapestry. *Communications of the ACM*, 35, 12 (1992), pp. 61-70.
7. Hill, W.C., Hollan, J.D., Wroblewski, D. and McCandless, T. Edit Wear and Read Wear. In *Proceedings of CHI 92 Conference on Human Factors in Computing Systems* (1992, New York: ACM), pp. 3-9.
8. Kahn, R.E. and Cerf, V.G. *The Digital Library Project, Volume 1: The World of Knowbots. An Open Architecture for a Digital Library System and a Plan for Its Development*. CNRI, 1895 Preston White Drive, Suite 100, Reston, VA 22091 Tech Report (March, 1988).
9. Karlgren, J. *Newsgroup Clustering Based on User Behavior-- A Recommendation Algebra*. Swedish Institute of Computer Science #SICS-T--94/04-SE (March, 1994).
10. Kawell, L.J., Beckhardt, S., Halvorsen, T. and Ozzie, R. Replicated Document Management in a Group Communication System. In *Proceedings of CSCW 88* (1988, New York: ACM).
11. Kumon, S. From Wealth to Wisdom: A Change in the Social Paradigm. In *Proceedings of CSCW 92* (1992, New York: ACM), pp. 3.
12. Maes, P. and Kozierok, R. Learning Interface Agents. In *Proceedings of AAAI 93* (1993, San Mateo, CA: American Association for Artificial Intelligence).
13. Malone, T.W., Grant, K.R., Turbak, F.A., Brobst, S.A. and Cohen, M.D. Intelligent Information Sharing Systems. *Communications of the ACM*, 30, 5 (1987), pp. 390-402.
14. Maltz, D.A. *Distributing Information for Collaborative Filtering on Usenet Net News*. MIT Department of EECS MS Thesis (May, 1994).
15. Pindyck, R.S. and Rubinfeld, D.L. *Econometric Models and Economic Forecasts*. MacGraw-Hill, New York, 1991.
16. Salton, G. and Buckley, C. Term-Weighting Approaches in Automatic Text Retrieval. *Information Processing and Management*, 24, 5 (1988), pp. 513-523.
17. Salton, G. and Buckley, C. Improving Retrieval Performance by Relevance Feedback. *Journal of the American Society for Information Science*, 41, 4 (1990), pp. 288-297.
18. Sheth, B. *A Learning Approach to Personalized Information Filtering*. MIT Department of EECS MS Thesis (February, 1994).
19. Stodolsky, D.S. *Invitational Journals Based Upon Peer Consensus*. Roskilde University Centre, Institute of Geography, Socioeconomic Analysis, and Computer Science. ISSN 0109-9779-29 #No. 29/ 1990 (, 1990).
20. Suchak, M.A. *GoodNews: A Collaborative Filter for Network News*. MIT Department of EECS MS Thesis (February, 1994).
21. Wiederhold, G. Mediators in the Architecture of Future Information Systems. *IEEE Computer*, March, (1992), pp. 38-49.