

RECOMMENDING AND EVALUATING CHOICES IN A VIRTUAL COMMUNITY OF USE

Will Hill, Larry Stead, Mark Rosenstein and George Furnas

Bellcore, 445 South Street, Morristown, NJ 07962-1910

willhill@bellcore.com, lstead@bellcore.com, mbr@bellcore.com, gwf@bellcore.com

<http://community.bellcore.com>

ABSTRACT

When making a choice in the absence of decisive first-hand knowledge, choosing as other like-minded, similarly-situated people have successfully chosen in the past is a good strategy --- in effect, using other people as filters and guides: filters to strain out potentially bad choices and guides to point out potentially good choices. Current human-computer interfaces largely ignore the power of the social strategy. For most choices within an interface, new users are left to fend for themselves and if necessary, to pursue help outside of the interface. We present a general history-of-use method that automates a social method for informing choice and report on how it fares in the context of a fielded test case: the selection of videos from a large set. The positive results show that communal history-of-use data can serve as a powerful resource for use in interfaces.

Keywords: Human-computer interaction, interaction history, computer-supported cooperative work, organizational computing, browsing, set-top interfaces, resource discovery, video on demand.

INTRODUCTION

With vast stores of multimedia events and objects to choose from, future users of the national information infrastructure will be overwhelmed with choices and human-computer interface designers will be called upon to address the problem. The aim of this research is to evaluate the power of a particular form of virtual community to help users find things they will like with minimal search effort.

Taking *video selection* as an initial test domain, the technique compares a viewer's personal ratings of videos with those of hundreds of others to find people with similar preferences

and then recommends unseen videos that these similar people have viewed and liked. The technique outperforms by far a standard source of movie recommendations: nationally recognized movie critics.

Virtual community, not virtual reality nor intelligent agents

The term *community* means "a group of people who share characteristics and interact". The term *virtual* means "in essence or effect only". Thus, by *virtual community* we mean "a group of people who share characteristics and interact in essence or effect only". In other words, people in a Virtual Community influence each other *as though* they interacted but they do not interact. Thus we ask: "Is it possible to arrange for people to share some of the personalized informational benefits of community involvement without the associated communications costs?" Such costs might include for example, the time costs of developing a personal relationship, costs to privacy, costs of synchronous face-to-face communications.

We wish to contrast our idea of virtual community with two popular themes in human interface work: virtual reality and intelligent agents. First we draw the contrast with virtual reality.

Popular future visions of networked computing and infrastructure marry perceptual immersion in virtual reality to high-bandwidth telecommunications. They seek a photo-realistic and real-time "cyber-face to cyber-face" social environment [10]. This immersive vision expects total involvement from participants. The result is what might be called a virtual reality community with its central issues of visual, auditory and temporal fidelity. By virtual community we do not mean virtual reality community. The pitfalls of seeking higher and higher fidelity to face-to-face communication have been well discussed in Brothers et al. [2]. Virtual community is about attempting to realize some of the

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 ACM. To copy otherwise, or to republish, requires a fee and/or specific permission.

CHI '95, Denver, Colorado, USA

© 1995 ACM 0-89791-694-8/95/0005...\$3.50

benefits of community without the associated communications costs.

A second popular vision of networked computing and infrastructure paints scenarios which include a large role for "intelligent agents". The idea is that of semi-autonomous programs somehow endowed with intelligence great enough to impress us with their ability to interpret our needs and their work on our behalf. Our notion of virtual community includes no central role for intelligent agents other than the human participants in the virtual community.

Relation of current work to previous research

Malone et al. [7] propose three types of information filtering activities: cognitive, economic and social. Cognitive activities filter information based on content. Economic filtering activities filter information based on estimated search cost and benefits of use. Social activities filter information based on individual judgments of quality communicated through personal relationships. This paper concentrates upon the computer-assisted mediation of Malone's third type: social filtering activities. However, a basic thesis of this work is that personal relationships are not necessary to social filtering. In fact, social filtering and personal relationships can be teased apart and put back together in interesting new ways. For instance, the communication of quality judgments can occur through less personal, and even impersonal relationships as well as personal relationships. Obviously, people want a satisfying mix of both personal and impersonal relationships.

We have been particularly interested in how social filtering activities can be simultaneously streamlined and enriched through the careful design of communication media. The social relationships in which filtering of information occurs can be streamlined by making them less personal and enriched by making them more personal. For example, adding or removing the communications costs of synchronous face-to-face encounter, anonymity, and choosing a more personal medium such voice or a less personal medium such as text are all means of influencing the personal aspects of communication. Social filtering can be simultaneously streamlined and enriched by making some aspects of a relationship less personal while making other aspects of the relationship more personal.

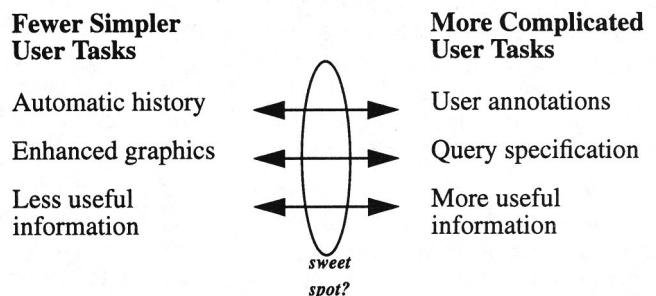
In the realm of computer-assisted mediation of social filtering, a few HCI experiments sparsely dot the space of possible designs. Goldberg's Tapestry system [3] is a site-oriented email system encouraging the entry of free text annotations with which on-site users can later filter messages. Annotations are rich in high quality information and their successful uses are valuable. However, despite hopes to the contrary, the twin tasks of writing annotations to enter filtering data and specifying queries to use filtering data require significant user effort. Domains where the invested efforts pay off readily are few, but they do exist. In the case of annotations where the method of entering filtering information for the benefit of others has significant user costs,

Grudin's question [4] "Who does the work and who gets the benefit?" becomes noticeably relevant.

Reacting against the trend of interface designers loading additional tasks on users in order to help them find things, the history-enriched digital objects approach (HEDO) [5][6][11] attempts to explore a region of the interface design space that minimizes additional user tasks. Through a combination of automatic interaction history and graphics, depictions of communal history within interface objects hint at their use while user effort is minimized. HEDO techniques record the statistics of menu-selections, the count of spreadsheet cell recalculations and time spent reading documents (e.g., email, reports, source-code,) in a line-by-line manner summing over sections and whole documents. Displays are simple shadings on menus, spreadsheets and document scroll bars. Because the HEDO data are less informative than annotations, they tend to be less useful, but they cost less to gather and use. There is evidently a trade-off here.

One way to think about the trade-off is considering the two approaches to social filtering mentioned so far as two ends of spectrum. On one end of the spectrum we have social filtering interfaces that expect more work from the user and give more value. On the other end of the spectrum we have interfaces that expect no additional work from the user but provide less value. Our thought is that perhaps somewhere in the middle of this spectrum between the two end alternatives, there might lie special niches that offer relatively more filtering value for relatively less filtering work. Such locations on the spectrum, if they existed, we could call design "sweet spots". Figure 1 depicts the spectrum and places a "sweet spot" in the middle.

Figure 1 *Spectrum of Social Filtering Interface Tasks*



We have in mind the ideal of a community of users routinely entering personal ratings of their interest concerning digital objects in the simplest form possible: a single keypress or gesture. These evaluations are pooled and analyzed automatically in service of the community of use. Members of this community, at their pleasure, receive recommendations of new or unfamiliar digital objects that they are likely to find interesting.

Recommendations might, for instance, take the form of recommendation-enhanced browse-products that tattoo symbols of predicted interest upon object navigation and control points. Later on, Figure 4 shows such a Mosaic Browsing

interface with recommendation enhanced hypermedia links and menus.

Of course the question is: does this kind of virtual community work? The answer as we will show is "yes" for videos and probably yes for many other forms of consumer level information items: books (categorized by author), video games, gaming scenarios, music, magazines and restaurants.

Concerning the use of ratings, Allen [1] reported unencouraging results on one of the first investigations (known to us) into personal ratings for HCI-type user-modeling. Recently, Resnick et al. [9] have designed a social filtering architecture based upon personal ratings and demonstrated its application to work-group filtering of Netnews. In a study of eight users reading 8000 Netnews messages, Morita and Shinoda [8] observed strong positive correlations between time spent reading messages and personal interest ratings of those messages. Their work suggests it might be possible for time-on-task measures to stand in for ratings, further reducing user tasks.

Interface Design Goals

In the process of achieving our overall goal of making personal evaluations do significant interface work for a virtual community, our approach illustrates a number of supportive community-oriented design goals:

- *Recommendations and evaluations should simultaneously ease and encourage rather than replace social processes.* Virtual community should make it easy to participate without heavy personal involvement while leaving in hooks for people to pursue more personal relationships if they wish.
- *Recommendations should be information rich, e.g. make apparent their origins.*
- *Recommendations and evaluations should be for sets of people not just individuals.* Of course a set of users who will receive a recommendation or evaluation may contain only a single individual, but the methods should not be limited to that case. This multi-person case of recommending is often important, for example, when two or more people want to choose a video to watch together.
- *Recommendations and evaluations should be from people not a black box machine or so-called "agent".* A particularly useful set of reference users to use in this capacity is the set of people similar to the target users. Equally interesting sets are colleagues or friends of the person or persons for whom the recommendations are computed since existing social context enriches the exchange.
- *Recommendations and evaluations should tell how much confidence to place in them, in other words they should include indications of how accurate they are.*
- *Data entry suffices for data query.* In our case, using a virtual community database to get recommendations requires merely submitting ratings to the community database. In other words, the way to match to others' preferences is to submit your own preferences which in turn may be

matched by others at a later time. So use of the database grows the database. Since data entry also queries the database, the work expected of users is cut in half.

Our design also embodies two research tactics.

- *The recommendation and evaluation methods should be generic in the sense of not relying on specific properties or features of items.* This rules out whole classes of approaches, particularly those based on feature analysis where the power comes from regularities among properties specific to the domain of items. This insures a certain generality to the techniques not available in feature-based approaches.
- *The recommendations and evaluations should get their strength from sheer amount of data.* Thus, if the database grows, the quality of recommendations grows. The communal history-of-user approach makes this possible. From the point of view of any given user, a larger database contains at least as many and probably more people with similar preferences than a smaller (i.e., subset) database. Thus overall performance, which is based how similar others' tastes are to the user, improves as a virtual community database grows.

The Research Questions

In order to understand the power of recommending and evaluating choices in a virtual community, we posed three basic questions:

- How well does virtual community perform?
- How reliable are the ratings data?
- How should one measure quality of recommending/evaluating algorithms for incorporation into user interfaces?
- According to the measures, what are the best algorithms for recommending items? For evaluating specific items? What are the trade-offs?

The second and third of these questions deserve further comment. The second question is straight-forward and standard statistical methods apply for answering it. On the third question, no standard measures have emerged as a consensus. At present, we consider two measures: (1) In a split-data test, how well do item ratings predicted by the recommending/evaluating system correlate with actual ratings submitted by users? (2) How do users evaluate the results they see from the algorithms? We report on these measures in the Results section.

METHOD: AN INTERNET CONCEPT TRIAL

Our method was to seed a virtual community in the Internet and to do all the work necessary to exchange high quality recommendations among participants. People participated (and still participate) through an email interface at videos@bellcore.com. From October 1993 through May 1994 we collected data on how the virtual community functions, how people like it, and how well it performs for participants.

How Virtual Community Technology Works

The virtual community support provided by at videos@bellcore.com consists of a generic object-oriented database to store and access preference efficiently and give out recommendations and evaluations. It is generic in the sense that one can construct various domains of items: videos, restaurants, books, document pages, and places to visit. In particular, at the time of our analysis, videos@bellcore.com included a data set of 55,000+ ratings of 1750 movies by 291 users. It includes recommending algorithms whose predictions improve as the data grow, and the number of movies, users and ratings and continues to grow daily.

Organization of the Database

The database is organized as set of interrelated instances of object classes. The objects are:

- *person*, which in the videos@bellcore.com case is an Internet subscriber who voluntarily participates by email.
- *item*, which in the videos@bellcore.com case is one of 1750 movies
- *rating*, which is a triple of item, person, and score
- *score*, which is either an integer representing a person's rated preference for an item or a symbol representing the person's relation to this item: *not-interested-in-item*, *must-see-item*, *no-previous-use-of-item*, *item-pending-as-recommendation* or *item-evaluated-for-user*
- *correlation*, which in the case of videos@bellcore.com is a triple of person1, person2, and correlation value, a floating point number between -1.0 and 1.0

The database contains 17 modules. A single high level database interface consisting of the following functions suffices to control it in most circumstances: *load-database*, *save-database*, *add-user*, *erase-user*, *add-item*, *erase-item*, *add-ratings*, *recommend-items*, *evaluate-items*.

The Email Interface

Internet participants send a message containing "subject: ratings" to videos@bellcore.com. The system replies with an alphabetical list of 500 videos for the user to evaluate on a scale of 1-10 for the titles they have seen. Rating 1 is low and 10 is high. Users may also rate an unseen movie as "must-see" or "not-interested" as appropriate. Surprisingly, early usability tests showed that it was reasonable to expect self-selected Internet users to rate movies on an alphabetical list of 500 movies. However we do not expect this to be a feature of a deployed system. In order to reduce item/item bias, for every participant 250 of the 500 movies listed are selected randomly. To increase rating hits and to gather a standard set of data for purposes of fair comparison, for every participant the remaining 250 titles are a fixed set of popular movies.

When users return their movie ratings to videos@bellcore.com, an EMACS client process parses the incoming message, and passes ratings data inside a request for a recommendations-text to the server database process. The server process performs *add-user*, *add-ratings* and *recommend-items*. In the

initial phase of adding ratings for a new user, ratings are added not only in the 1-10, "must-see" and "not-interested" categories, but also in the "unseen" category for titles that the user could have rated but did not. These unseen movies are the first pool from which to compute recommendations.

When a user is new, the database first looks for correlations between the new user's ratings and ratings from a random subsample of known users. We use the random subsample to limit the number of correlations computed to be $O(n)$ rather than $O(n^2)$ in the number of participants. One-tenth of the new user's ratings are held out from the analysis for later quality testing purposes. The most similar users found are used as variables in a multiple-regression equation to predict the new user's ratings. The generated equation is then evaluated by predicting the held out one-tenth of the new user's ratings and then correlating these predictions with the actual ratings.

Once the predication equation exists, it is quite fast to evaluate every unseen movie, sort them by highest prediction and skim off the top to recommend. When recommended, movies are marked in the database as "pending-as-suggestion". A recommendation text is generated and passed back to the EMACS front-end client process where it is mailed back to the user or users.

The Internet email interface is currently a subject-line command interface and there are many commands for specialized actions. Further details are available by sending mail to videos@bellcore.com.

Here is sample reply from the system. Names have been changed to protect anonymity:

Suggested Videos for: John A. Jamus.

Your must-see list with predicted ratings:

- 7.0 "Alien (1979)"
- 6.5 "Blade Runner"
- 6.2 "Close Encounters Of The Third Kind (1977)"

Your video categories with average ratings:

- 6.7 "Action/Adventure"
- 6.5 "Science Fiction/Fantasy"
- 6.3 "Children/Family"
- 6.0 "Mystery/Suspense"
- 5.9 "Comedy"
- 5.8 "Drama"

The viewing patterns of 243 viewers were consulted. Patterns of 7 viewers were found to be most similar.

Correlation with target viewer:

- 0.59 viewer-130 (unlisted@merl.com)
- 0.55 bullert,jane r (bullert@cc.bellcore.com)
- 0.51 jan_arst (jan_arst@khld.decnet.philips.nl)
- 0.46 Ken Cross (moose@denali.EE.CORNELL.EDU)
- 0.42 rskt (rskt@cc.bellcore.com)
- 0.41 kkgg (kkgg@Athena.MIT.EDU)
- 0.41 bnn (bnn@cc.bellcore.com)

By category, their joint ratings recommend:

Action/Adventure:

"Excalibur" 8.0, 4 viewers
 "Apocalypse Now" 7.2, 4 viewers
 "Platoon" 8.3, 3 viewers

Science Fiction/Fantasy:

"Total Recall" 7.2, 5 viewers

Children/Family:

"Wizard Of Oz, The" 8.5, 4 viewers
 "Mary Poppins" 7.7, 3 viewers

Mystery/Suspense:

"Silence Of The Lambs, The" 9.3, 3 viewers

Comedy:

"National Lampoon's Animal House" 7.5, 4 viewers
 "Driving Miss Daisy" 7.5, 4 viewers
 "Hannah and Her Sisters" 8.0, 3 viewers

Drama:

"It's A Wonderful Life" 8.0, 5 viewers
 "Dead Poets Society" 7.0, 5 viewers
 "Rain Man" 7.5, 4 viewers

Correlation of predicted ratings with your actual ratings is: 0.64 This number measures ability to evaluate movies accurately for you. 0.15 means low ability. 0.85 means very good ability. 0.50 means fair ability.

Instructions are also given for exercising various options in the community. For example, one can order up joint recommendations for more than one person and from a particular set of community members. This second example shows both capabilities at once. Jane and Jim want a joint recommendation of what movie to watch together. They also want recommendations only from Mary and Dick rather than the community at large. Again names have been changed.

*Suggested Videos for: Jane Robins, Jim Robins, together.**Your video categories with average ratings:*

6.9 "Drama"
 6.7 "Mystery/Suspense"
 5.8 "Comedy"
 5.7 "Children/Family"
 5.5 "Action/Adventure"
 5.0 "Horror"
 4.5 "Musical"

The viewing patterns of 2 viewers were consulted. Correlations with target viewers:

0.21 Mary Divens (no email address)
 0.18 Dick Divens (didi@westend.com)

*By category, their joint ratings recommend:**Drama:*

"Crossroads (1986)" 9.0, 2 viewers
 "To Kill A Mockingbird" 10.0, 1 viewer
 "Milagro Beanfield War, The" 8.0, 1 viewer

Mystery/Suspense:

"Body Double" 8.0, 2 viewers

"Coma" 7.5, 2 viewers

Comedy:

"After Hours" 9.0, 2 viewers
 "Adventures of Sherlock Holmes Smarter Brother, The" 9.0, 1 viewer

Children/Family:

"Beauty And The Beast (1991)" 8.0, 1 viewer

Action/Adventure:

"Lawrence of Arabia" 9.0, 1 viewer

Horror:

"Rear Window" 8.0, 1 viewer

Musical:

"West Side Story" 9.0, 1 viewer

THE COMMUNAL HISTORY-OF-USE ALGORITHMS

We have algorithms for two purposes, recommending items and evaluating items. Having tried a few versions of each, we report on the best we have discovered so far. We do not have evidence that these are the best algorithms possible, only that they are good. The algorithms we use for recommending have the following abstract functional form:

recommend-items (n-items, from-items, target-users, reference-users, methods, method-combination, database)

The function name is **Recommend-items** and n-items is how many items to recommend. From-items is the set of items from which to recommend a subset. Target-users is the set of users for which the recommendation is computed. In the case of a recommendation for one user, it is a singleton set. Reference-users is the set of users whose preferences will serve as the basis of the recommendation. For now, we normally compute the reference-users set by finding users whose preferences on the set of from-items correlate most positively with the preferences of the target-users. However, as the second example reply illustrates, community participants may specify a particular set of reference-users (friends, colleagues, whom-ever) to consider when computing recommendations. The earlier second sample reply showed this. This interesting alternative addresses the idea of supporting rather than replacing social processes. Methods is an optional argument that specifies one or more ways in which information about items, target-users, and reference-users will be analyzed. The Method-combination argument tells how to combine the results of the methods and the database argument specifies the particular data to use. The optional arguments default to standard values.

The function to return an evaluation of a proposed choice-looks like this:

evaluate-items (items, target-users, reference-users, methods, method-combination, database)

Here items refers to the set of items to be evaluated for the target-users on the basis of looking at reference-users. The methods, method-combination and database arguments can be used to specify alternative methods for evaluation, how to combine the results and data to use.

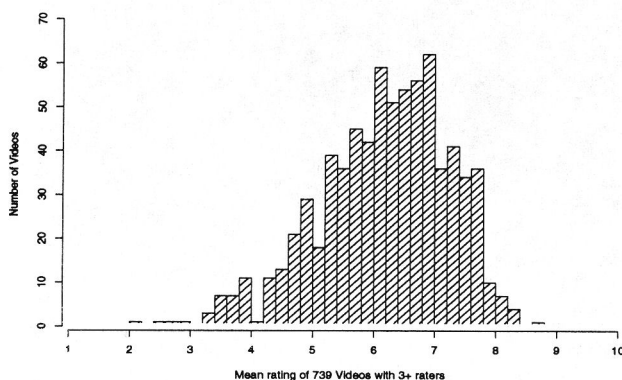
RESULTS

The Data

Currently the database consists of 291 participants in the community, 55,000 ratings on a 1-to-10 scale, another 2100 "must-see" or "not-interested" ratings, 64,000 "unseen" and 1200 "pending-as-suggestion" ratings. Of the 1750 movies in the database, 1306 have at least one rating and 739 have at least 3 ratings. 208 movies have more than 100 ratings, and 2 movies have more than 200 ratings. Users rate an average of 183 movies each with a standard deviation of 99. More than 220 of 291 total participants rated more than 100 movies. The database is small, but large enough to conservatively but accurately estimate a number of performance parameters.

For the 739 movies that have three or more ratings. Figure 2 shows the distribution of movies by their mean rating. Notice the slight bias toward positive ratings.

Figure 2 *Distribution of Video Mean Ratings*



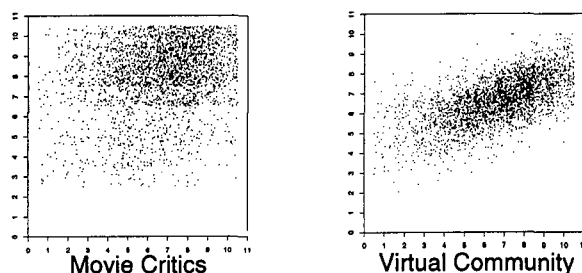
Reliability

Six weeks after they initially tried videos@bellcore.com for the first time by submitting ratings and receiving recommendations, 100 early users were asked to re-rate exactly the same list of movie titles as they had rated the first time. 22 volunteers replied with a second set of ratings. Three outliers were removed from the reliability analysis since they correlated perfectly and were evidently copies of the original ratings rather than second independent sets of ratings. For the remaining 19 users, on movies rated on both occasions, the Pearson r correlation between first-time and second-time ratings six weeks apart was 0.83. This number gives a rough estimate how reliable a source of information the ratings are.

Cross-validated Correlation Study

We held out 10% of every participant's movie ratings to provide a cross-validation test of accuracy. The cross-validated correlation of predicted ratings and actual ratings estimates how well our recommendation method is working. Figure 3 shows that our current best similar viewers algorithm correlates at 0.62 with user ratings. This is a strong positive correlation which means the recommendations are good. How good? We may expect three out of every four recommendations will be rated very highly by a potential viewer. We compared the quality of our virtual community recommendation method to a standard method of getting recommendations, that is, following the advice of movie critics. The ratings of movies by two nationally-known movie critics were entered. Their ratings correlate much more weakly at only the 0.22 level with viewer ratings. Thus the virtual community method is dramatically more accurate, as Figure 3 also shows.

Figure 3 *Two Scatterplots of Actual Ratings by Predicted Ratings. Plot on left shows movie critics as predictor ($r=0.22$). Plot on right shows virtual community as predictor ($r=0.62$) (all values are jittered for the purpose of visual presentation, 3269 predictions each for 291 users)*



User Feedback

Email responses from videos@bellcore.com include a request for open-ended feedback. Out of 51 voluntary responses, 32 were positive, 14 negative and 5 neutral. Here are some sample quotes:

- *Pretty good! Of 8 films you suggested I might enjoy, I agree with 4, disagree with 3, and am uncertain about 1.*
- *The recommendations are so-so. I agree with a few, but the slant of most of them is a suspense, high tension drama. This is not the only thing I watch.*
- *Nice work.*
- *I have seen eight of the ten flicks you recommended and thought they were pretty good; the other two I would definitely like to see.*
- *Your predicted films that I would like were so-so.*
- *Nothing was really off the mark, but I think I prefer a good comedy movie over anything else...*
- *The suggestions hit close to the mark.*
- *None of them were favorites and a few (like werewolf) were real bombs.*
- *It seemed pretty accurate.*
- *I think the recommendations make sense.*

- I received 5 recommendations. Four looked decent but one was a movie I had already seen and rated low.
- Most were things I had seen and had liked, or that seem like things I would enjoy
- I wasn't impressed. The list is not what I am interested in seeing.
- I think the recommendations were reasonably valid based on the input I gave.
- Not bad considering how quickly I filled out my ratings questionnaire.
- I'm not convinced that I want to see them.
- Some of them were right on (like "Die Hard," and "This is Spinal Tap," I liked both of those) while others were way off (I really don't like the Pink Panther movies and I didn't like Born on the Fourth of July at all.
- Intriguing! Of the 10 movies recommended, 8 have been on my "want to see" list for a long time.

Open ended feedback from users also indicated interest in establishing direct social contacts within their virtual community. Users can participate in either an anonymous or signed fashion. Interestingly, only four users exercised the anonymity option. Wishing to extend the social possibilities of the virtual community, two users asked if they could set "single and available" flags in the community indicating they wanted to use the community as a means of dating. One user found a long lost friend from junior high school. Another wrote that he took the high correlation between his movie tastes and those of someone he was dating as evidence for a long future relationship.

The Upper Limit

One of the standard uses of reliability measures is to put a bound on prediction performance. The basic idea is since a person's rating is noisy (i.e., has a random component in addition to their more underlying true feeling about the movie) it will never be possible to predict their rating perfectly. Standard statistical theory says that the best one can do is the square root of the observed test-retest reliability correlation. (This is essentially because predicting what the user said once from what they said to the same question last time has noise in at both ends, squaring its effect. The correlation with the truth, if some technique could magically extract it, would have the noise in only once, and hence is bounded only by the square root of the observed reliability). The point to note here is that the observed reliability of 0.83 means that in theory one might be able to get a technique that predicts preference with a correlation of 0.91. The performance of techniques presented here, though much better than that of existing techniques, is still much below this ideal limit. Substantial improvements may be possible.

FUTURE DIRECTIONS

We see a potential for deployment to customers of national information access who will be faced with thousands of possible choices for information and entertainment, in addition to videos.

Virtual Community recommending in the Mosaic Interface to the World Wide Web.

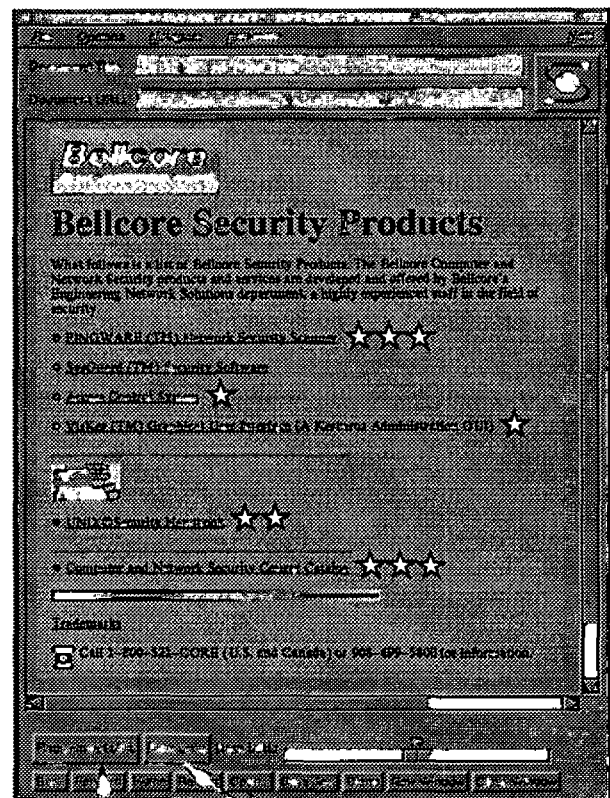
We have instantiated a version of our server where items are World Wide Web URLs (universal resource locators) in place

of videos. Figure 4 displays a modified Mosaic browser interface that accepts ratings of WWW pages on a slider widget (near bottom) and reports them to an appropriate virtual community server. When a user clicks on the **Recommend URL** button (near bottom), the browser contacts the virtual community server to get recommended URLs and then fetches the recommended page. It also displays next to every hypertext link, one-half to four stars which represent the virtual community's predicted value of chasing down the hypertext link.

Community Headroom

One direction in which we plan to push the research is toward more individual and social aspects. In particular we are interested in distributed peer-to-peer versions rather than the centralized client/server version that we have now. A wireless deployment of a peer-to-peer version could include wearable PCS devices, pairs of which will, when in close physical proximity, exchange ratings data for local virtual community computation.

Figure 4 Recommendation-enhanced Browsing in a specialized Mosaic Interface. Note that next to each hyper-link are one to four stars that predict the user's interest in following that link. At the bottom of the window are a recommend button and rating slider to rate the current page.



Recommend URL

Rate Links

CONCLUSIONS

Choice under uncertainty is an opportunity to benefit from other more knowledgeable people. How to support such social filtering with computation has been the topic of this paper. We have demonstrated a *virtual community* method that allows human-computer interfaces to harness the power of a social strategy involving minimal additional work with good utility. We have reported on how it fares in the context of a fielded test case: the selection of videos from a large set. In the case of videos, virtual community recommendations are measurably successful and can be used to recommend or evaluate videos for participants. Virtual communities may also sprout up around other domains such as music, books and catalog products. Targeting both groups and individuals for recommendations and evaluations, it performs well on stringent tests and will continue to improve as the virtual community database grows. When presenting choices in the interface and when a virtual community of users exists to inform those choices, there is no reason to leave users without recommended courses of action. The positive result we have reported suggests that others may want to investigate the power that communal history-of-use data can bring to interfaces.

ACKNOWLEDGMENTS

This work revives a line of research initially pursued jointly with Tim McCandless and David Wroblewski in the late eighties. Their continued interest, suggestions, advice and encouragement concerning the current work have been invaluable. On many occasions, Jim Hollan, Tom Landauer and Bob Allen asked questions and made suggestions that influenced the direction of this work over the past few months. We would also like to thank Ben Bederson, Diane Duffy, Dan Lin, Alan Mcintosh and Kent Wittenburg for many helpful conversations and suggestions. Finally, without the participation of the Internet community evolving around videos@bellcore.com, the lessons of this paper would have been impossible to learn. Special thanks to all who made suggestions and found bugs.

REFERENCES

1. Allen, R.B. (1990) User models: Theory, method and practice, *International Journal of Man-Machine Studies*, 32, 511-543.
2. Brothers, L., Hollan, J., Nielsen, J., Stornetta, S., Abney, S., Furnas, G., and Littman, M., Supporting informal communication via ephemeral interest groups. *Proc. ACM CSCW'92 Conf. Computer-Supported Cooperative Work* (Toronto, Canada, 1P4 November 1992), 84-90.
3. Goldberg, D., Nichols, D., Oki, B.M. and Terry, D. (1992) Using Collaborative Filtering to Weave an Information Tapestry. *Communications of the ACM*, 35, 12, pp. 51-60.
4. Grudin, J., Social Evaluation of the User Interface: Who Does the Work and Who Gets the BENEFIT?, Proceedings of IFIP INTERACT'87: Human-Computer Interaction, 1987, 805-811.
5. Hill, W. C., Hollan, J. D., Wroblewski, D., and McCandless, T. (1992) Edit Wear and Read Wear. In: *Proceedings of ACM Conference on Human Factors in Computing Systems, CHI'92*. ACM Press, New York City, New York, pp.3-9.
6. Hill, W.C., Hollan, J.D. (1994) History-Enriched Digital Objects: Prototypes and Policy Issues, *The Information Society*, 10, pp. 139-145.
7. Malone, T.W., Grant, K.R., Turbak, F.A., Brobst, S.A. and Cohen, M.D. (1987) Intelligent Information Sharing Systems. *Communications of the ACM*, 30, 5, pp. 390-402.
8. Morita, M., Shinoda, Y. (1994) Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval, *Proceedings of the 17th Annual International SIGIR Conference on Research and Development*, pp. 272-281.
9. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J. (1994) GroupLens: An Open Architecture for Collaborative Filtering of Netnews. *Center for Coordination Science, MIT Sloan School of Management Report WP #3666-94*.
10. Rheingold, H., (1993) *The virtual community: homesteading on the electronic frontier*, Reading Mass: Addison-Wesley.
11. Wroblewski, D., McCandless, T., Hill, W. (1994) Advertisements, Proxies and Wear: Three Methods for Feedback in Interactive Systems, in *Dialogue and Instruction*, Beun, R., Baker, M., and Reiner, M. editors. Springer-Verlag (forthcoming).