Cover Page

Title of submission: The Visual Decision Maker – a recommendation system for collocated users

Category of submission: Sketch

Name and full contact address (surface, fax, email) of the individual responsible for submitting and receiving inquiries about the submission: Steven M. Drucker, Microsoft Research, 1 Microsoft Way, Redmond, WA. 98052, USA, +1 425.703.5850, sdrucker@microsoft.com

The Visual Decision Maker – A recommendation system for collocated users

Steven M. Drucker
Microsoft Research
1 Microsoft Way
Redmond, WA 98052 USA
sdrucker@microsoft.com

Asta Roseway
Microsoft
1 Microsoft Way
Redmond, WA 98052 USA
astar@microsoft.com

Tim Regan
Microsoft Research, UK
Roger Needham Building
7 J J Thomson Ave
Cambridge, UK CB3 OFB
timregan@microsoft.com

Markus Lofstrom
1 Microsoft Way
Redmond, WA 98052 USA
markus@lofstrom.com

Abstract
We present the Visual Decision Maker (VDM), an application that gives movie recommendations to groups of people sitting together. The VDM provides a TV like user experience: a stream of movie stills flows towards the center of the screen, and users press buttons on remote controls to vote on the currently selected movie. A collaborative filtering engine provides recommendations for each user and for the group as a whole based on the votes. Three principles guided our design of the VDM: shared focus, dynamic pacing, and encouraging conversations. In this paper we present the results of a four month public installation and a lab study showing how these design choices affected people's usage and people's experience of the VDM. Our results show that shared focus is important for users to feel that the group's tastes are represented in the recommendations.

Keywords
Movies, collaborative filtering, user modeling, shoulder to shoulder, single display groupware, co-located collaboration, multi-person interfaces, multiple input devices, flow, awareness.

Project/problem statement
There are many times when we need to make joint decisions for a variety experiences; we might be going out for dinner with a group of friends, planning a trip

Challenge:

- Create a compelling group recommendation tool
- Make user preference model building interesting to users

Strategy:

- Exploit shoulder-to-shoulder interaction style
- Make visually compelling
- Stimulate conversation amongst participants

with a spouse, choosing interior design with the family, or simply choosing a movie to watch with a group of friends for the weekend. We set out to design a system that could facilitate the decision making process by modeling users' joint preferences elicited by visual images. One of the biggest problems in making recommendations to users is in gathering user preferences for building accurate user models. We wanted to make the process of gathering user preferences more enjoyable by stimulating conversation amongst participants and integrate the recommendations into a seamless and enjoyable experience.

The initial domain that we chose to examine was movie recommendation systems for a number of reasons: this was a common task amongst a wide variety of users; user models were easily built from user input using existing techniques (collaborative filtering); and data (in the form of still images, movie information and collaborative preferences) was readily available. We wanted to explore the interface issues, and examine the validity of our techniques in this domain before trying the same techniques in other domains. In particular, we wanted to examine ways of facilitating interaction in 'shoulder to shoulder' situations (multiple people looking at a single display) since this situation is used quite effectively in viewing television, or playing video games, but has not been exploited effectively for other purposes.

Background

The team included one graphics and one UI researcher, a graphic designer, and a software development engineer and the project was initiated in the spring of 2002. The project included a 4 month public installation of the system in a cafeteria in the fall.

While the design constraints of our system were different (notably for joint users of a single display), there have been many recommendation systems built in the past, see [Brees98] for an overview of collaborative filtering recommendation systems. Most of these papers were focused on single users, but there have been some work including the Flytrap [Crosse02] system and PolyLens [O'Conn02] that have focused on group recommendations. The Flytrap system monitored user's music preferences on their own personal machines and used RFID tags to track user's movements in to rooms and selected appropriate music based on the occupants. The PolyLens system is a modification of the MovieLens web based film recommendation system. It allows users to form groups and display recommendations of the group as a whole, but each user interacts with their own display at potentially distant locations. We used the data and algorithms from the MovieLens project for our recommendations.

Challenges

One of the primary theoretical problems for recommendation systems is the collection of preferences [Shard95]. It was our goal to make this collection period itself a stimulating experience. We'll discuss our approach to addressing this in the next section.

We also encountered a number of practical problems along the way in installing the system in a public place. Having several people all interacting with the display simultaneously presented some difficulties: a touch-screen kiosk is better suited to a single user. We wanted an experience where people could sit around, have a discussion, and interact with the system. Eventually, multiple IR remote controls were used, but

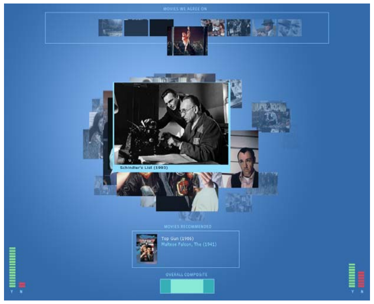


Figure 1: The VDM interface, configured for 2 users. See color plate 1 for an enlargement.

Implementation Details:

- Uses DirectX 9 graphics library
- Uses C# and managed code extensions to DX9 for rapid prototyping
- Requires graphics card of NVidia GeForce 2.0 or better (32 Mb texture memory)
- Requires PC of 1.5 GHz or better
- Uses SQL server backend for logging
- Uses SONY RM-VL900 programmable remote controls

since the installation was in a public place, these remote controls needed to be chained in place.

We also didn't want to constrain the system so that people needed to 'log in' and 'log off'. This meant the system needed to notice when participants stopped responding and potentially reset the user model that had been built up over time for users. In addition, there was no convenient way to identify return users so that previous models of their behavior could be used for the system.

Solution

As mentioned previoulsy, it was our goal to make the collection of preferences stimulating and integrated into the overall recommendation experience. We did this partially by using some of the ideas proposed in The Jack Principles [Gottl02], always maintaining pacing, awareness of the participants' actions and rapid feedback to individual responses. The basic idea was to create a flowing stream of images directed to the front of the display. Only one image was selected at any one time and all the users of the system had a chance to vote on this single image. Any individual user's interaction was extremely simple: indicate that the user liked the movie, thumbs down to indicate dislike, or pass to indicate either neutral preference or no knowledge of the movie. Still images were used instead of text since different users have highly variable reading rates and we did not want to have extensive amounts of text being shown to the users. Still images also did a good job evoking memories of the entire movie for those users that had previously viewed a movie. The still image was supplemented by the name and year of the movie for better recognition.

See fig 1, for the interface as configured for 2 users. This was the initial interface and one used for laboratory studies. The interface that was used in the installation was adapted to support up to 4 users.

Although a number open issues still remain about details of the interface (as will be discussed in section C), the overall response to the installation was extremely encouraging and definitely achieved the ends to which we set out – assisting groups of people in making decisions about movies.

When we first set out to build our intended design, we confronted a number of challenges. These ranged from how to effectively create shared focus amongst the participants, to ways of representing and displaying movies, handling input and providing feedback for multiple, simultaneous users. The following sections overviews some of these issues.

Shared Focus

The first issue was whether to present the users with a number of different movies simultaneously versus one movie at a time. Many different movies would allow subjects to move more quickly through a large number of movies since you could effectively present 8 or 12 movies at once and allow users to just select amongst the ones that they know and like. However then, each user would need to have their own selection mechanism and keep that distinct from the others. In addition, each user would not be as aware of the selections of the other users, thus stimulating conversation, an important part in the decision making process. Instead we adopted a serial presentation approach, where one image was shown to the entire group and everyone has a chance to vote on the movie.

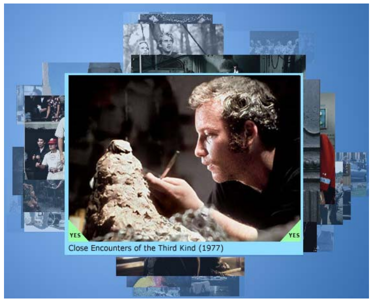


Figure 2: Close-up of the focused image with responses from two users.

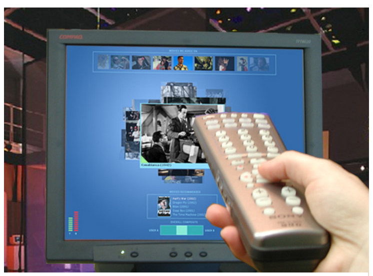


Figure 3: Users interacted with the VDM via 5 keys on a remote control: thumbs up, thumbs down, pass, pause, and request recommendations.

Figure 4: Close-up of the top center of the screen showing movies that users agreed upon.

Movie representation

Choosing the proper way to represent a movie was also debated. While a movie title along with the release date and other information (actors, director, etc.) could be used, we found that textual information was absorbed at different rates by the users. We also found that visual images evoked a much stronger emotional response then textual descriptions. However, many users could not identify a movie simply from a still image, so the image was supplemented by just the textual information of movie title and year. We also tried using movie posters as the representation for the movie, but some users found the wide variations in typeface, imagery and overall style somewhat jarring. People did not comment on this when still images from the movies were used instead. We used images of over 900 movies taken from an interactive movie CD-ROM [Cinemania – 1997] which had already chosen images that were more representational than just a single random image taken from the film. The problem of choosing the best image to represent an entire film is still an open one which we did not address in this work.

Image Presentation

Once the decision was made to serially present the movies, we had a number of options available. We could leave each image up for a set period of time and then flip to the next one, but users found the rapid switch from one image to the next confusing. Fading could also work, but we tried an approach used effectively in the You Don't Know Jack Game [Gott02]. Images were streamed from a virtually distant point to be the front of screen. The images were essentially on rails, moving forward at a set pace. Users could see the selected image in front of the other background images that would later be the focus for selection. The rate at which the images were streamed was based on how quickly users were responding to the selected image. If all the users had already responded to an image, it was taken off the stream, and the one behind it became the center of focus. The rate of forward movement was then made slightly quicker so that each image was focused upon for approximately the same amount of time. If one user did not respond, the system would wait for 60 seconds before moving on to the next image. Users could pause the process at any time so that discussions about particular movies could ensue. The order of the images was from more common movies to less common ones so that people would not spend too long using the system without being familiar with any of the movies. While the initial ordering of the top 30 movies was not adjusted over the duration of installation, the subsequent list of movies was random.

Multiple User Handling

Since the system was built for a café installation, there was never an explicit 'log on' or 'log off' process. We instrumented the remote controls to detect when they were held by a user to allow for each user to have a chance to vote on a movie, however this instrumentation turned out to be unreliable so we used the metric that if someone has not voted on any image for 5 successive times, the system assumed that no one was present for that remote control. Since we had no way of tracking whether different users, with different tastes were now voting, the system reset the user model after assuming that a user wasn't present. After a few minutes when the system detected no users, the system went into 'attract' mode where directions for using the system were presented along with a more rapid stream of images.

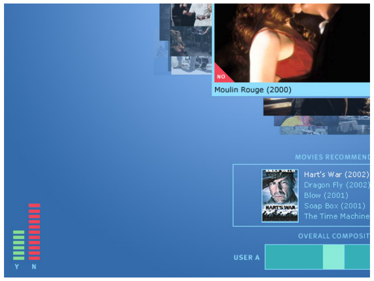


Figure 5: Other feedback areas on the screen: tallies showing number of thumbs up, and thumbs down votes, current list of recommendations, and the degree of overlap between 2 users.

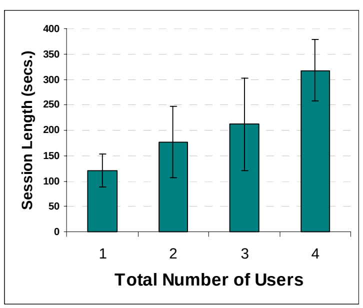


Figure 6: Average session length based on number of people in a group.

Feedback

As soon as a user responded with a positive or negative vote on an image, the system marked a corner of the image as voted: a red triangle with NO, a green triangle with YES, or a light blue triangle with PASS, was used. Movies that people agreed about where moved to an area on the screen and the last 8 movies where users agreed where displayed. In addition, the system kept a tally on the number of positive and negative votes for each user at corners of the screen as well as a measurement of overlap in agreement between users. Immediate feedback was important so that users could see their own and each other's responses. Sound was also used to indicate agreement or disagreement.

Recommendations

The top 5 recommendations (with a thumbnail for the top recommendation) were shown as the users interacted with the system. This was based on a joint model of preferences for all the current users of the system. We used the algorithms described in [O'Conn02] for computing the best recommendation. A special button on each of the remotes took users to a recommendation screen, with a list of the top 20 movie recommendations. This could be brought up (and dismissed) at any time by any user. If the recommendation screen is left up for more than 2 minutes, the system would revert back to the attract mode.

Evaluating the experience

The experience was evaluated both in the café installation and in a separate laboratory setting to investigate some issues that arose during public usage. We instrumented the VDM to collect data over the entire period of the installation. One measure for evaluating the relative enjoyment of the experience is to examine "session length" which we derive as the period of time between when at least one user starts to vote on images until 2 minutes go by without a vote. Figure 5 shows that the more users, the longer the session, even though the users would use the system for about the same amount of time before asking for recommendations. This data can be interpreted as the greater the number of users, the greater the potential entertainment value of the experience and conversation during the gathering of recommendation preferences. This is also born out in the density of response times based on the number of users. Groups of 4 tended to move less rapidly through the images with reportedly more time spent on discussion.

We also provided a mechanism for user feedback based on handwritten or web based forms. This feedback revealed general enjoyment of the system (and verified that groups tended to converse about the movies during use) but revealed two areas of concern. One was the accuracy of the recommendations and another was speed. We decided to address both of these issues in a laboratory experimental setting.

In the laboratory, we examined paired participants and varied the interface in 3 different ways: one interface was the same as the café installation; another interface changed the wait time from 60 seconds between when one person responded and the other person had to respond to 1 second; the final interface was to allow each user to go through the same set, but on separate screens at their own pace. We examined the different interfaces to determine 3 questions: which interface fostered the most conversation; which interface had the

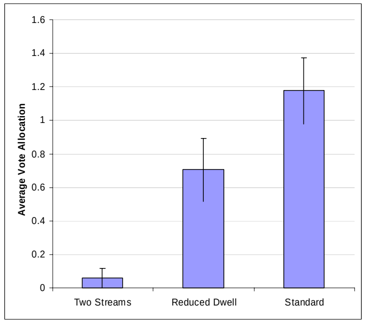


Figure 7: Amount of conversation per interface.

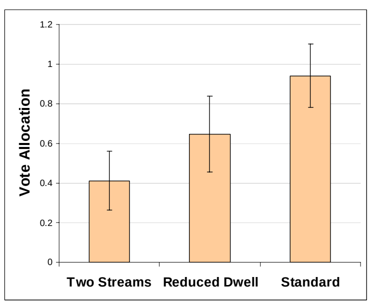


Figure 8: Group recommendation quality per interface.

most preferred pacing; and which system produced the best recommendations. The experiment was run on 34 participants (17 male and 17 female) which included 4 married couples, 10 sets of friends, and the rest colleagues or house-mates. Results showed that the users significantly preferred the joint experiences over the separate experiences and significantly preferred being able to move at their own pace rather than a rapid pace dictated by the system. Perhaps most surprising was that the users significantly preferred the recommendations produced by the joint experience over the separate experience ***even though*** the recommendations were produced using exactly the same algorithms. This was explained by the fact that the users were aware of each other's responses and thus understood why the system made the recommendations that it did. Figures 7 and 8 show breakdowns of the amount of conversation per interface and the perceived quality of recommendation per interface.

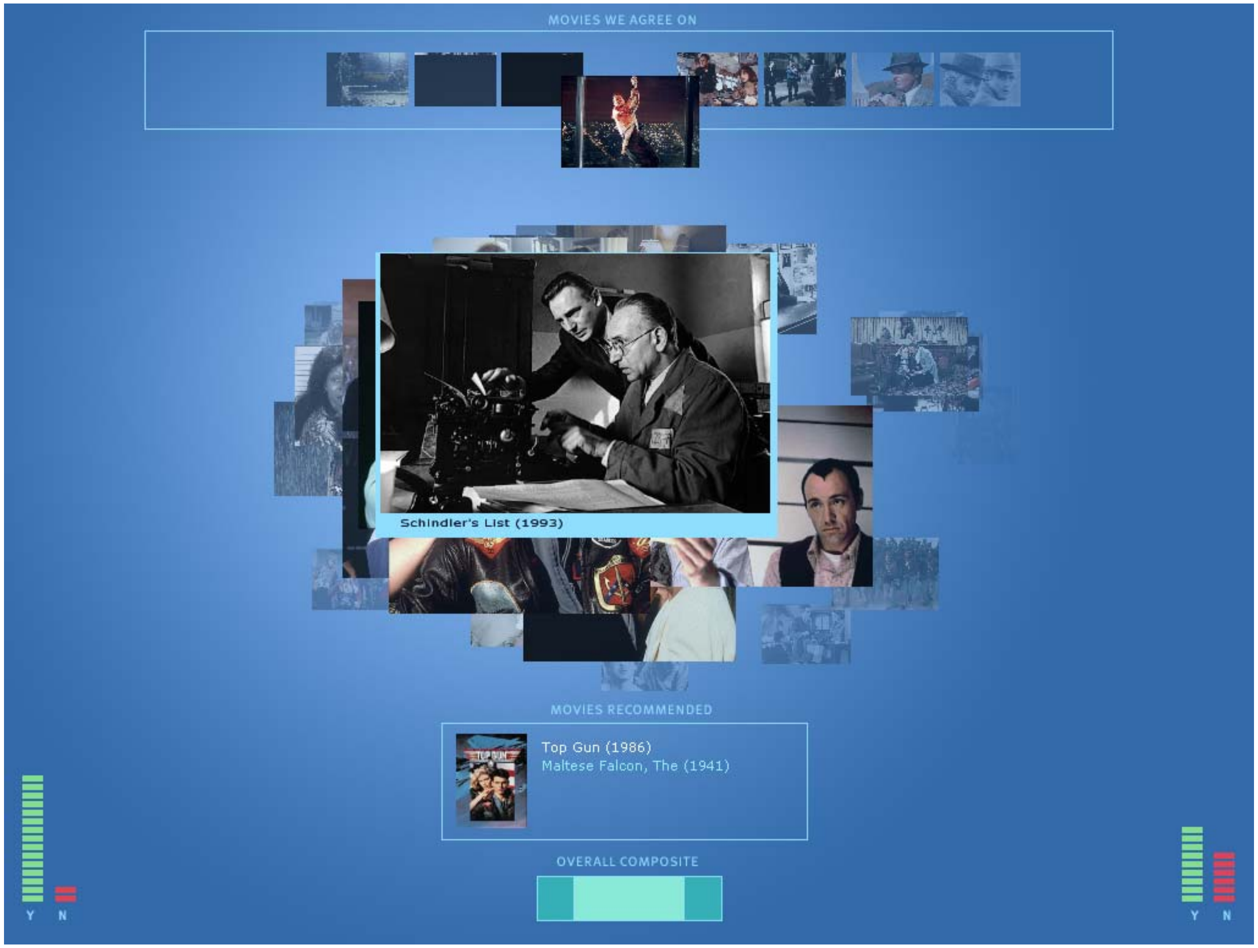
Conclusion and Next Steps

The results of the installation and the experiment were encouraging. The interface designed helped address one of the significant drawbacks of recommendation systems based on user models – that is the collection of preferences from users – by making the experience a more compelling and entertaining one. The joint interface elicited conversations between the participants and perhaps most encouraging, the participants preferred the recommendations made by the joint system over recommendations made based on separately gathered preferences.

Some of the problems encountered were that the system had no way of recognizing users from session to session thus preference data could not be stored across sessions. This could be fixed by requiring a login process, but we did not want to limit spontaneous use of the system. Better ways of dealing with users starting and stopping to interact with the system would also be helpful. The capacitive based sensing system that we tried was promising, but was too unreliable for use in the café installation. The current user model is based on collaborative filtering, but other models for viewer behavior could also be examined with potentially better recommendations resulting.
We also would like to apply the system to other domains rather than just movie recommendations. Trip planning, restaurant recommendations, and interior decorating are all potential domains to apply the system.

Bibliography

1. Breese, J. S., Heckerman, D., and Kadie, C. (1998). "Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In Proceedings of Fourteenth Conference on Uncertainty" in Artificial Intelligence, Madison, Wisconsin. Morgan Kaufmann
2. Crossen A., Budzik J., and Hammond K. (2002). Flytrap: Intelligent Group Music Recommendation. *Proceedings of IUI2002*.
3. Gottlieb, H. (2002), "The Jack Principles of the Interactive Conversation Interface", Jellyvision Inc, http://www.jellyvision.com/deal/jackprinc.htm
4. O'Conner M., Cosley D., Konstan J.A., and Riedl J. (2001). PolyLens: A Recommender System for Groups of Users. *Proceedings of ECSCW2001*.
5. Shardanand, U and P. Maes. Social information filtering: algorithms for automating 'word of mouth'. In proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems, pages 210-217, 1995.



Color Plate 1: The Visual Decision Maker (VDM) interface. A stream of movie images is presented to the users. As the users vote on whether they liked or disliked (or have no opinion) on the movies, the system builds a user model of the preferences of the users. The system then proceeds to make recommendations to the group. Feedback is given in a number of ways. The number of positive and negative votes for each user is shown in the corners; the overall amount of agreement is shown in the bottom middle. 2 Recommendations are shown in the lower center part of the screen. Movies that the users have agreed on are shown in the top center of the screen.

Color plate 2: Two users using remote controls to interact with system. An instrumented version of the VDM was installed in a café for 4 months, usage logs and surveys were obtained, and a laboratory experimental study for the efficacy of the interface was performed.