Netflix Recommendations for Groups

Steve Berry [1], Steven Fazzio [2], Yongyi Zhou[3], Bethany Scott [4], Luis Francisco-Revilla [5]
School of Information, University of Texas at Austin
1616 Guadalupe #5.202
Austin, TX 78701-1213
[1] sberry@unideal.net, [2] fazzios@gmail.com, [3] zhouyongyi@gmail.com, [4] killer.tofu7@gmail.com,
[5] revilla@ischool.utexas.edu,

ABSTRACT

Digital TV has significantly increased the amount of programming available to end users, making it difficult to find content related to their interests. In response, recommendation systems have been built to improve the users' ability to find content that is desirable to them. Netflix, one of the most full-featured online movie services in the United States, has created a well-known recommender system for long-form video content that provides highly accurate viewing suggestions to individual users. Movie watching, however, is frequently a group activity or even a social event that brings together multiple family members and/or friends for a social viewing experience. Netflix's current streaming video system doesn't assist groups of users in deciding what to watch. In order to explore the potential benefits of a recommender system for groups of users, we conducted a formative study aimed at identifying the general watching habits of Netflix users, to establish the basis for a prototype that would support this type of interaction. Based on the study results, we created a group recommender prototype that combines the advantages of the existing Netflix recommender system and user profile systems. Our prototype was constructed by utilizing the Netflix REST API to provide real recommendations that can be customized for a group of users. We conducted a focus group with Netflix users in order to evaluate the prototype and to collect insights into possible enhancements. The evaluation revealed that participants value the concept of group recommendations and would like to use the proposed system.

Keywords

Digital Television, Program Recommendation, Joint Recommendations, Collaborative Filtering, User Profile.

INTRODUCTION

Netflix has created arguably the best recommender system for long-form video content, providing highly accurate suggestions to users. With the introduction of streaming video and "watch instantly" content, there is a new context for multi-user or group recommendations. Netflix's current allowance for multiple user profiles within an account provides the basis to implement a combined recommendation system in the future, which would allow users to receive suggestions tailored to the specific friends or family members who wish to watch movies together.

Netflix's user profile function allows each household member to have a separate queue of movies, ratings, and recommendations. The account holder also allocates how many DVDs to be sent from each queue, saving time spent arranging different family members' movies manually. For example, for a family with a plan allowing 4 DVDs to be sent to them at a time, the account holder can specify to have one movie from his list, one movie from his spouse's list, and two movies from the children's list sent out at any one time. This allows for different rates of watching and returning movies without the user having to gather and collate family members' choices into a single queue.

In June 2008, Netflix announced that it was discontinuing the profile functionality. The change would consolidate the rental history of separate profiles, making future recommendations inaccurate. Any ratings and reviews created by profiles would be deleted (Leban, 2008). Ultimately, Netflix decided not to delete the profile function. The responses generated during the time when profiles were in danger were highly revealing. By coding and analyzing these responses, we gained insight into the different reasons Netflix subscribers use and value their profiles.

While many families used profiles in the way described above, taking advantage of the parental controls and customized recommendations, other users described their use of profiles for different genres, like movies and television shows, with the different queues used for organization and to ensure that only one DVD of each type was sent at a time (Jeremy, 2008). Other users cited the limit on number of movies that can be added to a queue, leading one to comment, "the reason I have two profiles is because I maxed out my first queue" (Pogue, 2008). Users from each group expressed their perceived inability to use Netflix effectively without profiles, demonstrating its high value to them as a product feature.

This is the space reserved for copyright notices.

ASIST 2010, October 22–27, 2010, Pittsburgh, PA, USA.
Copyright notice continues right here.

User profile functionality is important to our project because the ability for different members of a household to rate movies and receive recommendations separately provides the basis for our proposed group recommendation system. Since family members can maintain profiles customized with their individual preferences, we are able to combine the movie recommendations built on these preferences in order to suggest Netflix programming that will appeal to everyone in the household.

The following section of the paper provides background information on personalization in recommender systems, outlining current methods for providing recommendations to groups of people. The third section describes the findings of a survey focused on exploring the current viewing habits of Netflix users. Section four, describes the creation of the prototype and explains the design challenges. In section five, presents the findings of a face-to-face focus group conducted Netflix users who have multiple viewers in their households and report on the discussion of how our proposed system could benefit general Netflix users in real life. Finally, section six explains our conclusions, including some recommended improvements to the prototype and suggestions for further research in this area.

RELATED WORK

This literature review is divided into four related sections. The first three sections discuss three categories of personalization strategies in the field of recommender systems. The fourth section discusses recent techniques that have been used with existing group recommender systems.

With the abundance of television channels and increasing amount of digital content available, viewers can benefit from recommendations that guide them to the programs that they will most enjoy. However, it is not always clear how user preferences can be identified from among the seemingly endless number of programming choices. As Adomavicius & Tuzhilin (2005) classifies the approaches to generating recommendations into three basic categories: content-based, collaborative, and hybrid recommenders.

Content-Based Recommenders

Content-based recommender systems attempt to find items similar to ones the user has ranked highly in the past. The user is given suggestions of items similar to the ones that he or she has rated as preferred. Problems with these systems include limited diversity in the recommendations, particularly for new users with little or no past rating behavior. The intensive time and labor required to define each item in the system can also be an issue.

Rashid et al. (2002) discuss ways to get around the "New User" problem inherent in recommender systems. The goal is to begin providing meaningful recommendations based on the fewest possible rated items, so that the user sees effective recommendations after spending the shortest possible amount of time rating items. Items that everyone likes do not provide much useful information about an individual user's tastes. Rashid et al. suggest identifying items with a high "entropy" factor – items usually given either very high or very low ratings – as a way to find important distinctions in the user's preferences.

Collaborative Recommenders

Collaborative systems recommend items that are highly rated by other users with similar tastes and preferences. Using a collaborative filtering approach, a recommender system identifies users who share the same rating patterns with the active user, proposing items which the like-minded users favored and the active user has not yet seen. On its own, collaborative filtering provides much more diverse recommendations than the content-based approaches, as it is based on the experience of the user's neighbors and does not rely solely on the universe of items rated by the individual. In addition, collaborative filtering techniques do not require the aforementioned resource-demanding content descriptions, as they identify correlations among the users' ratings instead of among contents. However, collaborative filtering has some drawbacks, such as lack of flexibility in estimating the similarity among users, and the so-called gray sheep problem, associated with those users whose preferences are "strange" or very different.

Hybrid Recommenders

To overcome these limitations, some studies propose to adopt a hybrid approach between content-based matching and collaborative filtering, which can be enriched with additional user preference information usually provided in a social network (Martínez et al., 2008). This approach mixes content-based methods and collaborative filtering, taking advantage of synergistic effects and mitigating the inherent deficiencies of either paradigm. This way, users are provided with recommendations that are more accurate than those offered by each strategy individually (Burke 2002).

The use of semantic information in recommender systems has already been proposed in many systems (O'Sullivan et al., 2004). In the simplest proposals, the semantic descriptions are used with the goal of providing the users with additional information about the TV contents they are watching (Dimitrova et al., 2003). By contrast, some of the more elaborate approaches also include these semantic attributes in the recommendation process (Mobasher et al., 2004). However, these proposals do not infer complex semantic relationships from the knowledge provided by the semantic descriptions.

Fernández (2007) presents an approach for automatic content recommendation that considerably reduces these deficiencies. Fernández proposes a hybrid recommendation strategy that combines content-based methods and collaborative filtering. The cornerstone of this technique is a new and flexible metric that quantifies the semantic similarity between specific TV contents.

Lekakos and Giaglis (2006) demonstrated how the concept of lifestyle can be incorporated in the recommendation process to improve prediction accuracy by efficiently managing the problem of limited data availability. They propose two approaches: relying on lifestyle alone, and integrating lifestyle within the nearest neighbor approach.

Recommendations to Groups

Recommender systems have traditionally recommended items to individual users, but recently there has been a proliferation of recommenders that address their suggestions to groups of users. The shift of focus from individuals to groups makes more of a difference than one might at first expect.

Jameson and Smyth (2007) provide a comprehensive description of the unique challenges faced by recommender systems for groups. They break down the process of group recommendation into four steps, explain how they differ from individual recommendation systems, and discuss issues raised by each step.

A Family Interactive TV system named FIT was introduced to filter TV programs according to the different viewers' preferences (Goren-Bar & Glinansky, 2004). It assumes that the choice of a viewer may change in the presence of other family members.

Yu et al. (2006) introduce three alternative strategies to generate program recommendations for multiple television viewers. After reviewing and analyzing the advantages and disadvantages respectively, they chose user profile merging as their solution. This strategy first merges all user profiles to construct a common "group profile," and then generates a common program recommendation list for the group according to the merged user profile.

Cosley et al. (2001) introduce PolyLens, a collaborative filtering recommender system designed to recommend items for groups of users, rather than for individuals. Group recommenders are more appropriate and useful for domains in which several people participate in a single activity, as is often the case with watching movies and eating at restaurants. The study presents an analysis of the primary design issues for group recommenders, including questions about the nature of groups, the rights of group members, social value functions for groups, and interfaces for displaying group recommendations. The authors found that users not only valued group recommendations, but were also willing to yield some privacy to gain the benefits of group recommendations.

SURVEY

We surveyed a group of Netflix users in order to better understand their behaviors and demographic information, as well as the relationships between those factors. The survey itself was hosted by SurveyMonkey. We used Facebook (a popular social networking site), as well as word of mouth, to recruit volunteers. This had the effect of skewing our sample toward our acquaintances, who may be more tech savvy than the typical Netflix user. The survey was available for several weeks, but we only received responses during the first eight days. In total, 60 people completed the survey, and 5 people began the survey but fail to complete it (answered less than 50% of the questions).

Survey Results

The majority of the respondents were male (61%), and the majority fell between the ages of 23 and 44 years old (90%). We were not able to find the actual demographics breakdown of Netflix users, nor were we certain that those statistics had ever been measured. Demographic data for the actual Netflix user population, if available, could prove useful in normalizing the survey results.

61% of the respondents reported that they watch 1-10 Netflix programs per typical month, split evenly between those who watch 1-5 programs and those who watch 6-10 programs. 63% either "mostly" or "always" used Netflix's streaming functionality. One question asked users which device they used to stream Netflix content. This question was interesting in that it had the fewest respondents, 51 out of 61. This is most likely due to the way the question was phrased. We hypothesize that after reading the question, "Primarily, what type of device have you used to watch Netflix streaming content?", the users who don't stream content decided to skip the question rather then read the possible responses, the first of which was "I do not stream Netflix content." Of the users who do stream content, we found that 59% of them use a PC.

Netflix's recommendation engine, as well as our proposed prototype, rely on users rating content on a one to five star scale. We found that the frequency of content rating was fairly normally distributed with a median frequency of "around half the time." When presented with the statement, "Netflix's recommendations are useful," a majority (63%) "slightly agreed" that they were useful, 23% "strongly agreed," and only one user "strongly disagreed."

Two questions pertained to group watching of Netflix content. We found that 78% of users share their Netflix account, but that only one of those 46 individuals used multiple profiles. This is very important, as our proposed functionality relies heavily on each user having his or her own personal profile. We also found that most people watch Netflix content in groups either "some of the time" or "most of the time," with less people watching in groups "half of the time." We hypothesize that these two maxima correspond to whether the user lives alone or in a household with multiple viewers (roommates, couples, families, etc.)

The last question explained our proposed group recommendation functionality, and asked how interested the user would be. We found that people were generally more interested than uninterested. Unfortunately, we received some reports that participants found the wording of this question somewhat confusing. We also recognize that since the participants in this survey are primarily our friends, the results might be skewed toward the positive.

Crosstabs

We cross-tabulated the results between the questions, and found some interesting correlations. The three factors that showed the most interesting correlations were gender, number of Netflix programs viewed per month, and type of device used to stream content.

Figure 1 indicates that people who use PCs to watch streaming video are far more likely to rate and review content. We hypothesize that this is because the PC user experience is far more conducive to reviewing movies, but it is also possible that users who prefer streaming over their PCs are the type of people who enjoy reviewing content more than other users. We also found that male users were much more likely to engage in group watching behavior. Initially, we hypothesized that this was because males were more likely to engage in non-PC streaming (45% vs. 30% for females), which typically implies streaming to a television and thus providing a more suitable situation for groups of users to watch together. However, we found almost no correlation between group watching behavior and the type of streaming device used.

We found a very strong positive correlation between frequency of streaming content and number of viewed programs per month ($p<0.01$, with a null hypothesis that there was no correlation). We hypothesize that this is because streaming content is much faster and easier than receiving DVDs through the mail, although we have no way to absolutely determine if there is a causal relationship between the two behaviors, and if that is the case, which way the causal relationship is directed.

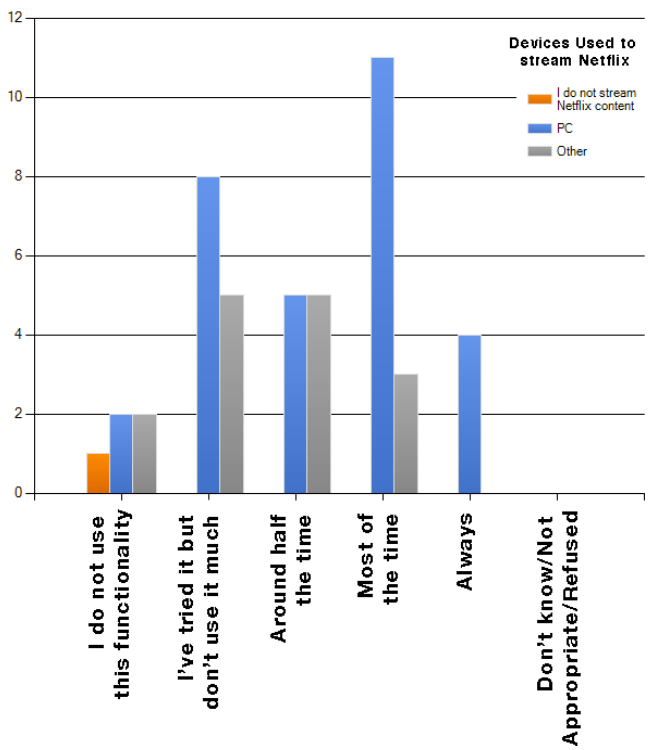

Figure 1. How frequently the user rates content vs. streaming device used.

Initially, we expected that the respondents who used Netflix most frequently would behave like typical "power users" and be more likely to take advantage of Netflix's features. We actually ended up finding a negative correlation between the number of programs viewed per month and the frequency of reviewing content ($p<0.082$, with a null hypothesis that the correlation would be greater than zero). Figure 2 illustrates this correlation.

We found that the respondents who use their PC to stream content were more likely to be "very interested" in our proposed functionality than those who use other streaming devices (37% vs. 20%). This adds further credibility to our hypothesis that the PC user interface is superior to the other options, in that it encourages user interaction.

PROTOTYPE CREATION

Engaging media systems collaboratively requires new tools (Irish & Trigg, 1989). For our focus group we set out to create a prototype application based on real data that enables multi-person recommendations in the context of the Netflix streaming service that would enable both group recommendations and a way to interact with those recommendations in a collaborative manner. Our use case involves two or more people, each with their own Netflix queue, body of ratings, and individual recommendations, interacting with Netflix streaming services to select and view movies collaboratively.

We used the excellent Netflix REST API to provide rental history and predicted ratings as the basis of the system. Our approach was to build a body of recommendations among a

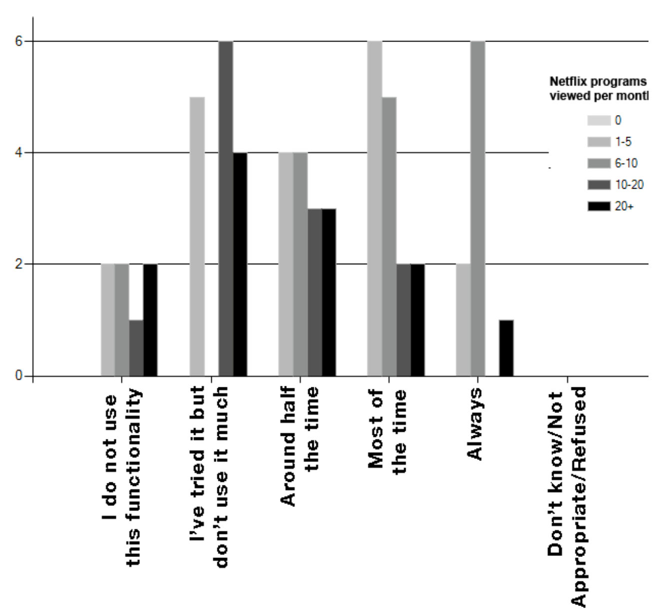

Figure 2. How frequently the user rates content vs. Netflix programs viewed per month.

group by retrieving each individual's recommendations from Netflix, combining them and sorting them to display what movies all viewers would most enjoy. The prototype was implemented as a web interface using mod_perl and Apache on a hosted Linux machine.

Authentication

In order to authenticate users against the API and add them to our system, we had participating users enter their Netflix logging information into the system, which retrieved connection credentials and stored them until explicitly deleted. The next screen in the interface presented a menu listing all authenticated users, as seen in Figure 3. The users could select two or more participants from this list and submit the form to generate the recommendation view.

Recommender View

Our Recommendation view has two primary sections: Merged Queues and Merged Recommendations.

The fact that a user has added a movie to his or her queue is significant in that the user has explicitly indicated that this is a film they would like to see. We decided to separate merged queues from merged recommendations because they differ semantically and create separate game plans for evaluating what to watch.

The Merged Queue list is generated using the API by retrieving what movies are in each participant's queue. We combine these queues and have the API generate a predicted rating for each viewer for every movie in this combined set. For any given movie, we created a "joint predicted rating" which was an average of the predicted ratings of all participating users for that movie. In addition to the average of the two ratings, we calculated the standard deviation amongst the ratings values as our "least misery index" – a larger standard deviation value suggests a larger difference in predicted ratings and more potential disagreement on the suitability of the selection for the group. Any movies that appeared in the intersection of queues are displayed first, followed by all remaining movies sorted by joint predicted rating descending, joint standard deviation ascending.

The Merged Recommender list was produced in much the same way but without attributing special significance to the intersection of recommendations across users. All recommendations were retrieved from the API, merged, joint predicted ratings were calculated and the movies were sorted by joint predicted rating descending, joint standard deviation ascending.

Finally, these two blocks were then rendered as two simple, horizontally scrolling lists of movie titles with the box art from the film and the joint predicted rating (see Figure 4).

We began evaluating this interface and, while the recommendations seemed to be relevant and useful, we felt there was a lack of enough information or visible "game plan" on how the recommendations were found and created. Just showing the joint predicted rating was opaque, and we believed that exposing the information and strategy used to promote any given recommendation could be helpful in selecting which movie to watch and also add to the collaborative nature of the experience.

The recommendation view was changed to expose the individual predicted ratings along with the joint predicted rating. We identified if the movie came from a particular user's queue or recommendation stream by rendering his or her name in green. These changes added a new dimension of information to the recommender view that was very helpful, but this additional information pointed out another potential issue inherent in the leveraging of Netflix' subjective ratings system when applied in a group context.

Normalizing Rankings

All users carry their own internal model of what a particular star rating means, and they have their own distribution

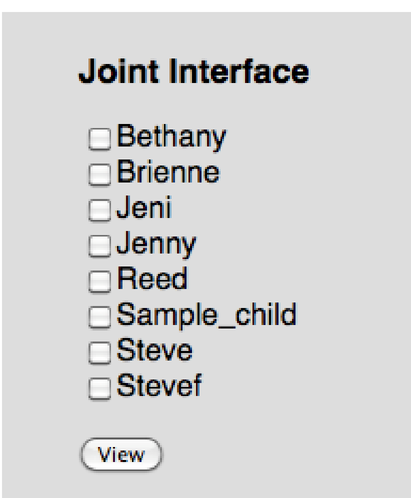

Figure 3. Multiple User Interface

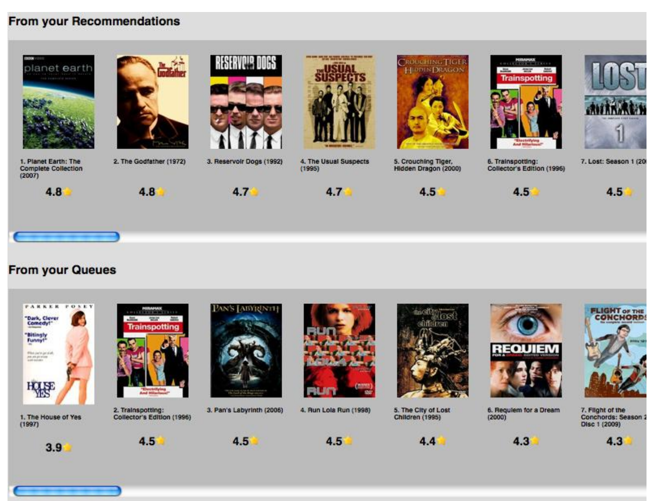

Figure 4. Basic Recommender View

patterns for assigning these star ratings. We found that some of users commonly give “5” or “1” stars to films, while others used those ratings judiciously. One user’s interpretation of a “4 star” rating may be the equivalent of another user’s “3 star” rating. This distorted the effectiveness of the recommendations by favoring the recommendations of whoever gives most frequently the highest ratings to movies.

To counteract this effect, we normalized each predicted rating based on the user’s actual rating habits. When a user authenticates, we download his or her entire rental history and generate a predicted rating for every movie in the set. This set is then used as the basis of a normalization function that accepts a predicted star rating as an argument and returns an indexed score based on the percentage position of that score amongst the dataset. This indexed score was then used in place of the predicted ratings when calculating the joint predicted rating. We found this had some noticeable effect on the results, reordering them according to normalized average scores rather than simple averages.

The “New User” problem was very evident in our attempt at normalization. Users without some corpus of rental history did not have enough rating data to make reliable normalization calculations. To defeat this problem, we primed the rating dataset with 50 ratings that match the “normal” Netflix ratings distribution. As we add actual user ratings to the dataset they will influence the normalization outcomes away from the weighted “normal” ratings distribution.

Finally, we found that displaying the normalized index instead of the joint star rating was confusing and explaining it in the context of the interface was difficult. We used a python binding to the system GD libraries to render a scatterplot of the dataset used to feed the normalizing algorithm. The graphic is a sparkline representation of the scatterplot and is shown next to each user’s rating information (Figure 5). Giving an at-a-glance interpretation of each user’s rating patterns was deemed to be very useful.

FOCUS GROUP

We conducted a focus group in order to explore Netflix viewing behavior in multi-person situations in an anecdotal, more in-depth manner than our survey could provide. We also hoped to collect feedback about our prototype from typical Netflix users who might find group recommendation functionality useful.

Design and Procedure

The three participants in the focus group were escorted to the IX Lab upon arriving at the iSchool. The discussion moderators introduced themselves and gave a short description explaining the project and the focus group procedure. The moderators discussed the confidentiality of the participants, and their written consent to be recorded and to participate in the discussion were obtained. Next, the group members were asked to introduce themselves and the

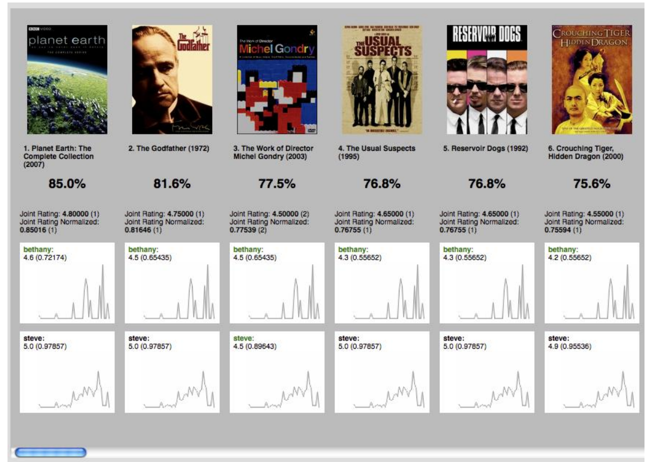

Figure 5. Normalized View with Visible “Game Plan” Details

discussion began. For the first part of the discussion, questions were designed to:

- determine individual participants’ decision-making process in selecting movies to add to the queue or watch instantly,
- explore participants’ movie-rating behavior and their perceptions of the usefulness of Netflix’s ratings and recommendation system,
- reveal decision-making process in a group setting and its social context, and
- probe potential privacy concerns associated with Netflix’s social networking “friends” feature.

In the second half of the focus group, we presented our prototype to the participants, asked for their feedback on the perceived usability and practicality of our current design, and discussed their suggestions as to how the functionality of the prototype could be improved.

The participants were also asked to fill out a copy of the survey form so that we could obtain background information about their Netflix habits. The results are as follows:

Participant 1

- Watches 6-10 Netflix movies per month (both DVDs by mail and streaming video)
- Has tried watching streaming video, but does not use the service much
- Uses a PC when she does watch streaming content
- Shares one Netflix profile with family members
- Watches Netflix movies with others most of the time

Participant 2

- Watches 10-20 Netflix movies per month
- Always watches movies using Netflix’s streaming service

- Uses an Xbox 360 to watch streaming content
- Household members use separate profiles
- Watches Netflix movies with others all of the time

Participant 3

- Watches 1-5 Netflix movies per month
- Uses Netflix's streaming service around half of the time
- Uses a PC to watch streaming content
- Shares one Netflix profile with household members
- Watches Netflix movies with others around half of the time

The discussion was transcribed and coded to reveal trends and patterns in the responses of the different participants. While the relatively small number of participants we could recruit is a limitation, we were still able to improve our group recommendation system with the information gained through the discussion of these points, by focusing on users' current behaviors so that our design works with the way they watch movies on Netflix. We also used their direct suggestions and feedback in modifying the prototype.

Participants' Viewing Behavior

Most of the discussion about the participants' viewing behavior centered on their use of Netflix's ratings and recommendation system. We also touched on Netflix's social networking "friends" feature and some privacy issues that could become a concern with group profiles and recommendations.

Ratings and Recommendations

Much of the discussion in the first half of the focus group revolved around participants' opinions about Netflix's current ratings and recommendation system. This topic was significant to the project because users' current perceptions about the system will influence their willingness to try our proposed group recommendation functionality.

When asked about her rating behavior, Participant 1 responded that she rates Netflix movies for her household almost all of the time – unless a disc happens to be scratched when it arrives, she usually does assign a star rating when the Netflix online interface prompts her. However, she also expressed uncertainty about the usefulness of providing ratings with the statement, "we're hoping that the rating system actually works someday … It sort of works, but maybe the more information they have the better they'll get at it." Participant 1's household also shares one account profile, and her family will "discuss at the end of the movie" what star rating to assign. When she and her spouse disagree in their opinions about the movie, they typically compromise by averaging the two ratings that each person would provide. She also notes, "if one of us really hated it, it definitely becomes a two, even if the other person really liked it." This type of rating behavior was interesting to us, because it reinforces our idea that watching movies with other people is a highly social activity that leads to much discussion between the viewers.

Participant 2 expressed her hesitance to rate the Netflix movies she watched until she created her own profile. Previously, she felt that rating movies of the different genres she enjoyed could potentially dilute the ratings and recommendations of the main account, making them less useful overall. After switching to two profiles, she rates movies much more often.

Participant 3 didn't know that multiple profile functionality existed until this point was raised in the discussion, indicating that Netflix poorly advertises this useful feature. He stated, "Most often I rate movies that I've seen that I didn't see on Netflix." His personal method for assigning star ratings was also unique. He claimed, "I tend to rate movies that I feel strongly about either way. I don't like to give threes, but I love to give fives and I love to give ones." He also noted that he will assign five-star ratings to movies he likes when he feels Netflix's suggested rating for the movie is too low, in the hopes that his five will slightly improve the suggested rating overall.

When asked whether recommendations play a role in selecting which movies to add to the queue or to watch instantly, Participant 3 stated that for him, Netflix's suggestions are "probably 80% okay" because he has been a user for several years and has rated a large set of movies. Also, he likes very specific genres and types of movies, so his ratings and taste preferences help create accurate programming recommendations. In choosing Netflix content to watch, he responded that suggestions are "a factor in my programming decisions, but it's not the only factor." Overall, he thinks Netflix recommendations are more accurate than recommendations made by other systems.

In response to Participant 3's thoughts about the accuracy of recommendations, Participant 1 said, "I was going to say it wasn't that useful yet, and then when you said that they don't send you the really popular bad movies, I thought, well, they never recommend those to me, so it actually probably is more like 80%." Furthermore, when asked if recommendations help her to choose movies, Participant 1 said, "if it's a recommendation of something that I've heard about, you know, I've heard it from a critic or somebody whose taste in movies I agree with, then I'll go ahead and add it [to the queue]. But if it's just the little synopsis, it's a rare occasion that I'll just pick it because it's recommended."

Friends and Privacy

Participant 3 identified his Netflix friends as a potential factor in his programming decisions. "[With] the friend suggestions, there's some when they suggest a movie, there's no way I'll watch it, and there's some when they suggest a movie, I'll watch it every time, depending on who that person is. Those [recommendations] are very strong."

Participant 1 also recognized her friend recommendations as being inconsistent, saying, "I have one friend whose taste in movies I can't stand, so it doesn't really help at all. But it's a good idea."

The other participants also thought Netflix's social networking features were a good idea, but they believed that these features were not very well advertised to users. For instance, Participant 2 did not have any friends through her Netflix account, and Participant 1 thought that only longtime Netflix users would be familiar with the social networking capabilities.

Participant 3's comment "I think Netflix is sacred, I don't know why, like I'd probably give out my bank password quicker," introduced the issue of privacy to the discussion. The participants were not aware that their Netflix friends could view their queues and recently watched movies, which could potentially lead to embarrassment or other negative social consequences. The option to "hide" movies in your profile would help alleviate this concern. Especially if a group recommendation system like the one we propose were implemented, the "friends" feature would become much more widely used, and more research on privacy and these social interactions would be necessary.

Prototype Discussion

After an introduction to our prototype and its working process in which the participants' Netflix profiles were combined to provide group recommendations to them, participants immediately pointed out the need for a filtering system to limit the results to specific genres. Participant 3 explained, "the movie decision process is, what kind of movie do you want to watch, or I really like this actress or actor or director, so that usually is how those conversations start." A way to view only the results linked to a certain genre or style, actor, or director would allow users to quickly locate movies they want to watch on a specific occasion, so they don't have to scroll through many irrelevant results.

Participant 2 also quickly asked, "is there a way that you can, after you watch a movie, put it at the very end" so that movies one users has watched recently don't continue to show up at the top of the recommendations. As the discussion continued, this issue of recently viewed movies being highly recommended emerged as a significant concern to the group members. Participant 1 observed:

> Participant 1: At least half the movies [recommended by the prototype] I've already seen and rated, so it's recommending that I see these movies again. Which isn't such a bad deal if you're getting a bunch of people together and you're saying, "Oh yeah, I saw that, that was great…"
>
> Moderator: Then maybe that aspect of the algorithm says, if there has been a rating, but rating was very very high, then maybe leave it in. Once you start throwing multiple people together… it's interesting.
>
> Participant 1: It is. For instance, if a bunch of people got together and they were totally in the mood for horror and I've seen Old Boy before, I probably would say, "Old Boy, yeah, I saw that three years ago, let's see it." But there's been a lot of stuff recommended that I've seen very recently, so for me it seems a little strange that it's being recommended right now.

The group eventually agreed that if the group recommendation results were limited to show only movies that have not been watched recently, a balance would need to be implemented in some way. For instance, if a movie is extremely highly rated, users may want to watch it more frequently in a group setting than other movies, regardless of how recently they have seen it.

Another major area of discussion concerned the Xbox 360 and its current Netflix streaming video interface. Participant 2 pointed out that the Xbox "knows when your friends are on … if you're already friends with someone through Xbox, it would be easier to log on the different profiles and introduce this prototype too."

The Xbox 360 also has a dual-login feature in which users can log into multiple gamer profiles using the same machine. This feature could potentially allow for different Netflix accounts or user profiles to be associated with the various Xbox gamer profiles. The ability for friends or family members to quickly log in during a group movie-watching setting seems to fit in with the current Xbox interface.

Towards the end of the discussion, when participants were asked to provide their overall impressions of the group recommendation system and its usefulness, the following dialog occurred:

> Participant 1: I got my parents on Netflix, and when they come to visit, they have a very limited group of movies that they want to see, and so it's a great idea to be able to link their account to my account and go, "Oh, well these are the movies you're gonna agree on."
>
> Participant 3: I find it to be more interesting in learning about my friends than in picking a movie, I think that's fascinating.
>
> Moderator: So the key is, to some extent, not so much about seek and find a movie, as create an interesting social experience while you're doing it as well.

Since the idea for a group recommendation system centers on the social experience of watching and discussing movies with others, it is important for the system to be interesting and relevant to the users. In this regard, we feel our prototype was successful, as it stimulated great conversation during our focus group.

CONCLUSIONS AND FUTURE WORK

With the intention of adding group recommendation functionality to Netflix's current streaming video system, in this study we carried out a preliminary survey, developed a prototype application, and conducted a focus group to evaluate the prototype and make suggestions for improvements. The user community that will benefit most from our proposed additions includes those who currently use Netflix, rate movies in Netflix, and watch Netflix rentals with friends and family.

Our survey provided us with a general view of the watching habits of Netflix users, and the possibility of our proposed group recommendation system fitting in with their current habits. Armed with the results gathered in this survey, we were able to determine how our prototype might be integrated with their current behavior. After researching the possibility of extending the Netflix recommendation system to include the option of providing group recommendations, we developed our prototype application. Additionally, we conducted a focus group with three Netflix users who have multiple viewers in their households to evaluate the prototype's usefulness and to further explore Netflix viewing behavior in multi-person situations. We found that participants valued group recommendations and were willing to use the proposed system.

The discussions from the focus group were encouraging, as they provided us with a valuable insight into the usage and advancement of our proposed group recommendation system. After analyzing the focus group data, we identified the following areas in which we can improve the prototype:

- The ability to easily refine the results by genre – users could interact with a menu of checkboxes to add and remove genres from the viewing list. We would add genre labels to each entry to help place the movie.
- A slider to control the MPAA ratings range – defaulted to span ratings "G" through "R." The results would be filtered within the ranges and each result marked clearly with the MPAA icon for that rating.
- The director and the top-billed stars of the film as links listed under each title. Clicking on these links will refine the recommender list to show all films from those individuals, ordered by joint predicted ratings.
- A search box to allow free text search. Results would be displayed in columns for people, titles, and studios with a recommender stream for those terms.

The proposed prototype application was specifically designed for users of Netflix. Further study is needed to understand the users of other recommendation sites to serve their special needs. Another limitation of this study was the relatively small sample size of our focus group. Taking into consideration such factors as our small budget and the limited time available to conduct the study, we believe the findings of this study are interesting and can be explored in different contexts and situations. Further study will provide useful insights for more revisions of our prototype.

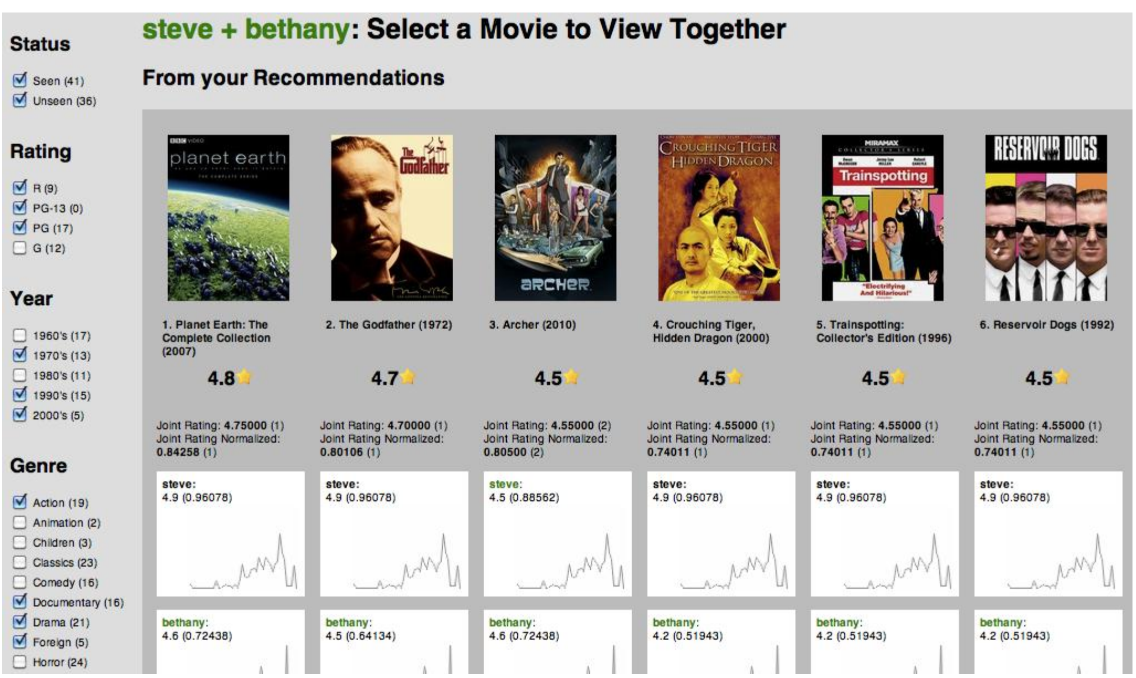

Figure 6. Possible interface modifications for future versions of the Netflix Group Recommender

Future work on this project should focus on refining the interface. Specifically incorporating the advantages of interactive search of merged recommendations (Albertson, 2010) to create the most effective and useful possible group recommender combined with collaborative information retrieval (Raya et al., 2004). Figure 6 illustrates how these changes might be incorporated into the existing prototype.

Additionally, we would conduct further user testing, especially tests which measure the effectiveness of various joint recommendation algorithms to add nuance and complexity, such as Bayesian learning (Lim et al., 2007) and perhaps better use of dimensions such as age of items in the dataset.

ACKNOWLEDGMENTS

The School of Information at the University of Texas at Austin provided the Information eXperience Lab (IX Lab) in which we conducted the focus group.

REFERENCES

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, *17*(6), 734--749.

Albertson, D. (2010). Analyzing User Interaction with the ViewFinder Video Retrieval System. *Journal of ASSIS&T. 61(2),* 238-252.

Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction*, *12*(4), 331-370.

Cosley, D., Konstan, J. A., & Riedl, J. (2001). PolyLens: A recommender system for groups of users. *In Proceedings of the European Conference On Computerr-Supported Cooperative Work*, 199--218.

De Meo, P., Quattrone, G., & Ursino, D. (2010). A query expansion and user profile enrichment approach to improve the performance of recommender systems operating on a folksonomy. *User Modeling and User-Adapted Interaction*, *20*(1), 41-86.

Fernández, Y. (2007). Avatar: Enhancing the Personalized Television by Semantic Inference. *International Journal of Pattern Recognition and Artificial Intelligence*, *21*(2), 397.

Goren-Bar, D., & Glinansky, O. (2004). FIT-recommend ing TV programs to family members. *Computers & Graphics*, *28*(2), 149-156. doi:doi: DOI: 10.1016/j.cag.2003.12.003

Irish, P. M., Trigg, R. H. (1989). Automatic User Preference Learning for Personalized Electronic Program Guide Applications. *Journal of ASSIS&T. 40(3), 192-199.*

Jameson, A., & Smyth, B. (2007). Recommendation to Groups. In *The Adaptive Web* (pp. 627, 596).

Leban, R. (2008). Netflix is a service, not a web site. Retrieved March 5, 2010 from http://www.thisuser.com/2008/06/Netflix-is-service-not-web-site.html.

Lekakos, G., & Giaglis, G. M. (2006). Improving the prediction accuracy of recommendation algorithms: Approaches anchored on human factors. *Interact. Comput.*, *18*(3), 410-431.

Lim, J., Kang, S., Kim, M. (2007). Automatic user preference learning for personalized electronic program guide applications. *Journal of the ASSIS&T. 58(9), 1346-1356.*

Martínez, L., Pérez, L. G., Barranco, M., & Espinilla, M. (2008). Improving the Effectiveness of Knowledge Based Recommender Systems Using Incomplete Linguistic Preference Relations. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *16*(supp), 33.

O'Sullivan, D., Smyth, B., Wilson, D. C., Mcdonald, K., & Smeaton, A. (2004). Improving the Quality of the Personalized Electronic Program Guide. *User Modeling and User-Adapted Interaction*, *14*(1), 5-36.

Osborn, J. (2008). Goodbye, Netflix profiles? Retrieved March 5, 2010 from http://www.metafilter.com/72626/goodbye-Netflix-profiles.

Pogue, D. (2008). Where are the Netflix profiles? Retrieved March 7, 2010 from http://pogue.blogs.nytimes.com/2008/06/23/monday-2/.

Rashid, A. M., Albert, I., Cosley, D., Lam, S. K., Mcnee, S. M., Konstan, J. A., & Riedl, J. (2002). Getting to Know You: Learning New User Preferences in Recommender Systems, 127--134.

Raya, F., & Pejtersen, A. M., & Cleal, B. & Bruce, H. (2004), A Multidimensional Approach to the Study of Human-information Interaction: A case study of collaborative information retrieval. *Journal of the ASSIS&T*. 55(11), 939-953.

Yu, Z., Zhou, X., Hao, Y., & Gu, J. (2006). TV Program Recommendation for Multiple Viewers Based on user Profile Merging. *User Modeling and User-Adapted Interaction*, *16*(1), 63-82.