Modeling a group of television viewers

Judith Masthoff

Faculty of Information Technology
University of Brighton, Brighton BN2 4GJ UK
Judith.Masthoff@brighton.ac.uk

Abstract. Watching television tends to be a social activity. So, adaptive television needs to adapt to user groups rather than to individual users. In this paper, we explore different ways in which individual user models can be combined to adapt to groups. We present an experiment on how real people make decisions for groups based on data about individuals. It is explored how this relates to the issue of modeling groups of learners.

Keywords Group modeling, interactive television, interactive instruction, adaptation, recommender systems

1. Introduction

In Britain, 100% of children have television in their homes, compared to only 53% PC, and 7% internet. They spend two and a half hours per day watching it (Livingstone & Bovill, 1999). One-third of children say the television is on when they come home from school, and two-thirds say it is still on when they go to bed. Working class children are far more likely to have their own television (and spend longer watching it), while they are far less likely to have access to a PC. The potential for education via television is clear. Interactive television could bring adaptive instruction into people's living rooms and children's bedrooms. For instruction to be effective, the system (television in this case) needs to adapt its broadcast to the interests, capabilities and knowledge of the individual student. In case of a television, this is even more important than in a classroom, as television competes with other entertainment forms. Adapting broadcasts to individual viewers is a topic in itself, and a lot of research has already been done in the area of electronic program guides. In this paper, we will explore an even more difficult issue: adaptation to a group of viewers.

Television viewing is largely a family or social activity (Barwise & Ehrenberg, 1988; Kasari & Nurmi, 1992). Unfortunately, television-viewing statistics do not include data on the average number of people watching television together and who watches television with whom (as also noted by Gillard, 1999). It is very likely to be culturally dependent, as the number of televisions per household varies widely. According to a large research in the UK (Livingstone & Bovill, 1999), television is the medium most often shared with family. Watching television together is top of the list of activities shared between parents and children, and more than two thirds of children watch their favorite programme with somebody else, nearly always family. Children most often watch with their siblings (Van Evra, 1998). Young people would very much like to watch television with friends, though (due to a lack of resource) many do not manage to do so (Livingstone & Bovill, 1999). Given the rising number of televisions in bedrooms it is likely that watching television with friends will be an increasingly popular activity. Already, television is the most popular conversation topic of young people with friends (Livingstone & Bovill, 1999). So, we believe that for adaptive instruction via television to become popular, it should be possible to learn while groups of people are watching together. Age, gender, intelligence, and personality influence what types of TV programmes children watch (Kotler, Wright & Huston, 2001; Gillard, 1999; Livingstone & Bovill,1999). Even within children and adults of the same age, differences will occur with respect to interests, knowledge and learning capabilities. The question then arises as to how you can adapt to a group of viewers, in such a way that each individual benefits from the broadcast.

2 Strategies for combining user models

A lot of research has been performed on user modeling, both in the general and the educational domain. Combining individual user models into a group model is, however, not easy, and has hardly been investigated. In this section, we will discuss some of the issues, and present a number of example strategies. We will also show what strategies are being used in existing systems. In this section, we will discuss the general case of group modeling. Towards the end of this paper, we will explore how this relates to modeling groups of learners.

2.1 Example strategies

Many strategies can be devised that combine individual user models. We will discuss some simple ones. We will assume the system has predicted preference ratings for each individual (say from 10, really like, to 1, really hate). We will illustrate each strategy with an example (the same example will be the basis of the experiment presented in Section 3). Individual ratings of three people are given for ten items. The example shows the group ratings resulting for the strategy and the corresponding 'group list', a sequence indicating in which order the items would be chosen. Sometimes, two items score the same, like E and F in the Average Strategy. That is indicated in the group list by placing them between brackets. This means that either E is followed by F, or F followed by E.

1. *Average Strategy*. Make a new list of ratings with the average of the individual ratings (in the example, we have given the sum rather than the average, as this will result in exactly the same list). Items get selected based on their rating on that list, the higher the sooner.

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
John	10	4	3	6	10	9	6	8	10	8
Adam	1	9	8	9	7	9	6	9	3	8
Mary	10	5	2	7	9	8	5	6	7	6
Group	21	18	13	22	26	26	17	23	20	22

Group List:
(E, F), H, (D, J), A, I, B, G, C

   A disadvantage is that an individual viewer might always loose out, because their opinion happens to be a minority view. This is the more likely to cause problems the larger the group. After all, in a small group the opinion of each individual will have a large impact on the average.

2. *Least Misery Strategy*. Make a new list of ratings with the minimum of the individual ratings. Items get selected based on their rating on that list, the higher the sooner. The idea behind this strategy is that a group is as happy as its least happy member. A disadvantage is that a minority opinion can dictate the group: if everybody really wants to see something, but one person does not like it, then it will never be seen.

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
John	10	4	3	6	10	9	6	8	10	8
Adam	1	9	8	9	7	9	6	9	3	8
Mary	10	5	2	7	9	8	5	6	7	6
Group	1	4	2	6	7	8	5	6	3	6

Group List:
F, E, (H, J, D), G, B, I, C, A

3. *Most Pleasure Strategy*. Make a new list of ratings with the maximum of the individual ratings. Items get selected based on their rating on that list, the higher the sooner.

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
John	10	4	3	6	10	9	6	8	10	8
Adam	1	9	8	9	7	9	6	9	3	8
Mary	10	5	2	7	9	8	5	6	7	6
Group	10	9	8	9	10	9	6	9	10	8

Group List:
(A, E, I), (B, D, F, H), (C, J), G

4. *Average Without Misery Strategy*. Make a new list of ratings with the average of the individual ratings, but without items that score below a certain threshold (say 4) for individuals.

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
John	10	4	3	6	10	9	6	8	10	8
Adam	1	9	8	9	7	9	6	9	3	8
Mary	10	5	2	7	9	8	5	6	7	6
Group	-	18	-	22	26	26	17	23	-	22

Group List:
(E, F), H, (D, J), B, G

5. *Fairness Strategy*. Top items from all individuals are selected. When items are rated equally, the others' opinions are taken into account. The idea behind this strategy is that it is not so bad to watch something you hate, as long as you get to watch the things you really love as well. This strategy is often applied when people try to fairly divide a set of items: one person chooses first, then another, till everybody has made one choice. Next, everybody chooses a second item, often starting with the person who had to choose last on the previous round. It continues till all items have been used. In our example, if we assume John chooses first, then John would like A, E, or I. He could choose E because it causes the least misery to others and has the highest average. Next it is Adam's turn. Adam would like B, D, F, or H. He could choose F because it has the best ratings for the others. Mary would choose A (her highest rating). Next, Mary would like E, which has already been shown, and then F, which also has already been shown. Therefore, it makes sense to let Adam choose, who likes B, D, or H. He chooses H, as that has the best ratings for the others. Following this strategy, we could end up with a group list like: E, F, A, H, I, D, B, etc. The list would, of course, be different if we let Mary of Adam choose first. However, we would expect A to be within the first three items, as it is the item Mary prefers most.

6. *Most Respected Person Strategy*. The ratings of the most respected person are used -in our example Adam-, only taking the ratings of the others into account to choose between similarly rated items. The idea behind this strategy is that groups may be dominated by one person. For instance, some research shows that the television remote control is most often operated by the oldest male present. Similarly, adults may have more influence than children (could depend on the time of day, adults having more influence later in the day). Visitors may have more influence than inhabitants of the house. Special circumstances, like birthdays, illness, etc. can influence who is "the most respected" person on a particular moment. A more sophisticated use of differences in social status would be to assign weights to the individuals' ratings.

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
John	10	4	3	6	10	9	6	8	10	8
Adam	1	9	8	9	7	9	6	9	3	8
Mary	10	5	2	7	9	8	5	6	7	6
Group	1	9	8	9	7	9	6	9	3	8

Group List:
F, H, D, B, J, C, E, G, I, A

2.2 Existing systems using group modeling

There are only two main systems that use group modeling (in the way intended in this paper): MusicFX (McCarty & Anagnost, 1998) and PolyLens (O' Conner, Cosley, Konstan & Riedl, 2001). MusicFX is used in a company's fitness center to select background music to suit a group of people working out at any given time. Users rate all music stations, from +2 (really love this music) to -2 (really hate this music). These ratings are converted to positive numbers (by adding 2) and then squared to widen the gap between popular and less popular stations. An Average Without Misery strategy is used to generate a group list. To avoid starvation and always picking the same station, a weighted random selection is made from the top m stations of the list (m being a system parameter).

PolyLens is a group recommender extension of MovieLens, which recommends movies based on an individual's taste as inferred from ratings and social filtering. It allows users to create groups and ask for a recommendation for that group. PolyLens uses the Least Misery strategy, assuming groups of people going to watch a movie together tend to be small and a small group to be as happy as its least happy member.

The application domains of both PolyLens and MusicFX differ from television viewing in the sense that these systems do not need to select a group of items: people normally only see one movie per evening, and music stations can play forever. Our view on adaptive interactive television is that reasonably small video segments (in the order of minutes at most) would be concatenated. The smaller the segments the more adaptation and real interactivity can take place.

Though some exploratory evaluation of MusicFX and PolyLens has taken place, for neither system it has been investigated how effective the strategies really are, and what the effect would be of using a different strategy. We believe that such research is vital.

2.3 Some outstanding issues

In reality, group modeling is even more complex. To mention just some of the issues:

- *Differences in rating tendencies*. Not all people have the same rating behavior. Some people only use the ends of the scale, they either "really hate" or "really love" an item. Others only use the middle, never being very positive, and never being really negative. A "7" by Pete, who is always very negative, may be a far more positive review than a "9" by Tim, who likes everything. These differences in behavior should be taken into account when using the ratings as input for the group model. One way to do this is to normalize them.
- *Ratings are not necessarily linear*. Is the difference between a rating of 10 and 9 really as large as the difference between a rating of 8 and 9?
- *Transparency versus privacy*. From a usability/acceptability point of view, it is important that viewers understand why a certain decision has been made. One way for the system to make its reasoning transparent is by showing the individual ratings and how its group rating relates to these. However, while this serves transparency, not everybody may like their individual ratings to be known to the other members of the group. For instance, somebody may be afraid of horror movies, but not like that to be known to their friends. In an educational setting, it can be even more embarrassing: lack in knowledge (like being illiterate) is something people may not be willing to share.

3 How real people do it

One can easily create hundreds of strategies (the strategies above are only the tip of the iceberg). The important question is which strategy is most effective and will be most liked by viewers. As a starting point, we want to determine what strategy real people use. We have performed a first experiment to explore this.

3.1 Experimental design

Method

Subjects were divided into two groups, experiencing different experimental conditions. In both conditions, subjects were given the same individual ratings of three people, John, Adam, and Mary, for a set of video clips. In seven questions, they were asked which clips the three should view as a group, given that they only had time to see respectively 1, 2, 3, 4, 5, 6, or 7 clips, and why they made that selection. The task presented to both groups differed only in that in condition 2, "John, Mary and Adam" had been replaced by "John (29), Mary (32), and their grandfather Adam (81)" (see Appendix for exact task wordings). The individual ratings had been chosen to enable differentiating between the strategies we expected subjects to use (same ratings were used as in Section 2). A between subject design was used as a pilot test revealed large order effects: subject felt compelled to change their group ratings in favor of Adam, if they received condition 2 after condition 1.

Research questions

We wanted answers to the following research questions:

- Do subjects follow a clear strategy? Is it possible to describe subjects' individual behavior in terms of a logical strategy? Are the strategies discussed above being used?
- Is there a dominant strategy? Is one strategy used by a majority of subjects, and, if so, which strategy is it?

- Do subjects take pleasure, misery, and fairness into account? Which do they find most important?
- Is social standing taken into account? Does subjects' behavior change if one person in the group can be regarded as more important?

We expected the results of both experimental conditions to provide some answers to the first three questions, and the difference between the conditions to provide some insight in the last question. Our hypothesis was that in Condition 2, Adam would be regarded as more important (because of his age), and the selections would be more geared toward his taste.

Subjects

Thirty nine subjects participated in the experiment. All were final-year undergraduate students of the IT faculty attending a lecture of the Adaptive Interactive Systems module. The students were studying various courses (B.A. Computer and Information Systems, B.Sc. Computer Studies, B.Sc. Computer Science, and B.Sc. Software Engineering). The experiment took place in a lecture room. Subjects were assigned to experimental condition depending on where they sat: the left of the room was assigned to condition 1 (18 subjects, 16 male, 2 female, average age 28, standard deviation 9.7), the right to condition 2 (21 subjects, 15 male, 6 female, average age 24, standard deviation 3). The spread over courses was similar for both conditions.

3.2 Results and discussion

Subjects do not seem to answer the questions independently: they responded which new clip should be added to the sequence they had already chosen for the previous question. This made it possible to present the results in the way we have done in Tables 1 and 2 (for respectively, Condition 1 and 2), only showing the new clip selected for each question. However, from an experimental point of view, this is not ideal: it might have influenced their strategy, making it perhaps less likely that they use the "fair strategy" (which only makes sense when selecting a larger group of clips). We need to explore whether the results would be different if we asked the subjects immediately to select, say, six clips. A between subject design could be used to distinguish between different set sizes.

Table 1. Results for Condition 1 (see below for meaning of coloring and lines).

	1	**2**	**3**	**4**	**5**	**6**	**7**
sub14	E	F	H	J	D	A	I
sub2	F	E				I	B
sub11						G	B
sub10							
sub4							I
sub7							A
sub9				D or J	D or J	A	I
sub17				D	J	B	G
sub3			J	H	D	G	I
sub18					G	D	B
sub16				G	H		A
sub6			A	H	J		I
sub13							
sub5			D			G	B
sub1						A	
sub12		H	J	D	E	G	
sub15		G		E	D	H	
sub8			E	A	B	D	H

Table 1 shows the results for Condition 1, and Table 2 for Condition 2. As can be seen in the tables, subjects did not always make a unique selection for a clip, sometimes they answered "D or J". We have tried to keep the table as simple (and less crowded) as possible: if a cell does not have a clip name in it, then the first name above it

applies. For instance, sub11 responded F to the first question. Subjects have been ordered to make the table as easy to view as possible.

The tables includes information about the how well the subjects' responses fit some of the strategies discussed in Section 2:

- Bold border lines indicate responses that are in correspondence with the Average Strategy.
- Gray cell colors indicate responses that are in correspondence with the Least Misery Strategy.
- Bold dotted border lines indicate responses that are in correspondence with the Average Without Misery Strategy.

So, for instance, sub7's selections were in correspondence with the Average Strategy for the first five clips, and in correspondence with the Least Misery Strategy for the first six clips. Sub2 followed the Average Without Misery Strategy throughout.

Table 2. Results for Condition 2.

	1	**2**	**3**	**4**	**5**	**6**	**7**
SUB1	F	E	H	J	G	D	B
SUB2					D	B	G
SUB14							
SUB 8						A	I
SUB 16							
SUB 11							
SUB 20						G	A
SUB 12							B
SUB 5						I	
SUB 9					A	D	I
SUB 15				A	D	J	
SUB 19				D	J	A	
SUB 10						G	B
SUB 3							A
SUB 4			D	H			I
SUB 18			G	J	H	D	B
SUB 21		G	D			B	E
SUB 17		J	H	E	D	G	B
SUB 13		H	D		J		
SUB 6	E or F	E or F	H	J or D	D or J	A	I
SUB 7	E	F	J	A	H	G	D

The results suggest the following answers to our research questions:

- There is evidence that human subjects use the strategies mentioned above, particularly the Average Strategy, the Average Without Misery Strategy and the Least Misery Strategy.
  - *Average Strategy*. Two subjects in Condition 1 (sub14 and 9) and five subjects in Condition 2 (SUB8, 16, 11, 19, and 6) exactly followed the Average Strategy.
  - *Least Misery Strategy.* Three subjects in Condition 1 (sub11, 10, and 5) and two subjects in Condition 2 (SUB12, and 10) exactly followed the Least Misery Strategy. Three subjects in Condition 1 (sub4, 7, 3) and three in Condition 2 (SUB 20, 3, 4) followed the Least Misery strategy almost completely. For clip 7, three selected A ("while Adam hates it, the others really like it so he will just have to put up with it", "Mary seems to loose out in most clips, so A is for her", "Might as well please two out of three"), three selected I ("Closest", "At least the majority will be satisfied").
  - *Average Without Misery Strategy*. Two subjects in Condition 1 (sub 17, 2) and three subjects in Condition 2 (SUB2, 14, 5) exactly followed the Average Without Misery Strategy. Three subjects used

a threshold of 4 or 5 and two subjects (sub2, SUB5) used a threshold of 2 or 3. As expressed by one subject "I try to please all of them making sure that no is lower than five".

- *Fairness Strategy*. Two subjects (sub6, 13) used some kind of a Fairness Strategy. Both selected A relatively early. They made comments like "Although Adam gave 1 mark for A, he gets to see F", "Although some gave some clips low marks, they all get to see some they rated highly". Other subjects applied fairness towards the end: "Mary 's average ratings have been low, so give her something she will enjoy" (sub1, explaining selecting A), "Mary seems to loose out in most clips, so A is for her" (SUB3), "As Adam did not like A, pick D next as he scored it a 9" (SUB15).

- There does not seem to be a clearly dominant strategy, but Average, Average Without Misery, and Least Misery are all plausible candidates for implementation. Fairness plays a role, but our human subjects did not have a clear strategy for applying it.
- Subjects sometimes exhibit completely unexpected behavior. Four subjects in Condition 1 (sub18, 16, 15, and 8) and four in Condition 2 (SUB1, 18, 21,7) selected G (ratings 6,6,5) before D (ratings 6,9,7). This seems rather illogical. Sub15 explained using disparities in ratings as a basis for selection. This would mean that a group is happy if everybody were equally happy or miserable. SUB1 also mentioned less disparity as a reason for preferring G to D.
- Many subjects take misery into account, as evidenced by the high proportion of subjects using the Least Misery and Average Without Misery strategies. Even subjects that do completely deviate from the Least Misery or Average With Least Misery strategies, like sub12, sub15, SUB21, SUB18, SUB17, and SUB13 avoid misery: all left out A and I from their selection. Therefore, preference should be given to a strategy that takes misery into account.
- There may be a tendency in favor of the hypothesis that social status influences selection, but the difference is small (no statistics have been done yet, probably not significant). We calculated Adam's average pleasure per question for each condition by taking the average of the sum of his ratings. The results are shown in the table below. Only one subject explicitly mentioned age as a reason for a selection: "A is not chosen because only the young ones like the topic" (SUB5). Our intention of making Adam the most respected person did not completely succeed: one subject (SUB7) actually mentioned "Adam's scores have been ignored to some extend because of age", another (SUB21) said "overall pick the average highest, if there is any difference attempt to match the two people with the same age". Overall, it seems that this part of the experiment was not successful: in future we will have to make it more obvious that one person is socially more important (perhaps by making it their birthday).

	1	2	3	4	5	6	7
Condition 1	8.9	16	23.7	31.6	40.1	46.2	51.7
Condition 2	8.9	16.1	24.9	32.4	40.6	46.3	51.9

4 Applicability to adaptive instruction

The experiment presented focussed on the more general issue of group modeling. The question arises to what extend this is applicable to adaptive instruction. In adaptive instruction, there are many reasons for selecting an item (topic, explanation, exercise), such as the student's existing knowledge (does it fulfill the prerequisites), learning goal and learning style, the educational flow (does it built on what has been explained before), etc. One way to match this onto the modeling as discussed above would be to construct a single rating for each item based on how it scores on these aspects. This might mean that a learner might be confronted sometimes with material that is above or below him, particularly when the group of learners is very heterogeneous. The prevention of misery strategies could avoid some of the worst problems.

5 Conclusions

Our research has only just started, but the issues involved in group modeling are clearly interesting, and merit a wide range of investigation. Our experimental research will be extended as follows:

- Instead of asking one user to select a set of items for a group, we are going to ask groups of users to agree on a selection for another group.

- Instead of asking one user to select a set of items for a group, we are going to ask groups of users (each playing the part of a viewer) to negotiate a selection for their group, using given individual preference ratings.
- Different variations of the scenarios could be explored, such as time of day (children may be more likely to determine what is on television early in the day, and adults late in the evening), special occasions (like it being somebody's birthday), etc.
- The scenarios could be made more explicitly related to the educational domain
- After implementation of one or more algorithms, we are going to let the system generate a selection, and let a subject judge how good the selection is, what are its strengths and weaknesses.
- We could also perform something like a "Turing test": ask users to compare the selection of the system with the selection of humans. Which do they prefer and why?

We also intend to research the literature on voting, after all, reaching group decisions is a vital political issue, and many different systems are employed for this purpose. We are exploring the use of a multi-agent architecture, in which an agent will represent each viewer, and the agents negotiate on the basis of knowledge about their viewer's preferences. Agents could use all kinds of strategies, including tactical voting.

References

Barwise, P., and Ehrenberg, A. (1988). Television and its audience. London: Sage.

O' Conner, M., Cosley, D., Konstan, J.A., and Riedl, J. (2001). PolyLens: A recommender system for groups of users. In Proceedings of ECSCW 2001, Bonn, Germany, pp. 199-218. As accessed on http://www.cs.umn.edu/Research/GroupLens/ poly-camera-final.pdf .

Gillard, P. (1999). The child audience: Who are they and how are they using TV and new media? Paper presented at the Children's Television Policy Forum and Reception, Sydney, 22 July. As accessed on http://www.aba.gov.au/abanews/conf/1999/ pdfrtf/20years_gillard.rtf .

Kasari, H., Nurmi, S. (1992). TV audience segments based on viewing behaviour. In Advertising Research Foundation (ARF) and European Society for Opinion and Marketing Research (ESOMAR), Worldwide broadcast audience research symposium. Toronto.

Kotler, J., Wright, J., and Huston, A. (2001). Television use in families with children. In J. Bryant and J.A. Bryant (Eds.), Television and the American family, pp 33-48. Mahwah, N.J.: Lawrence Erlbaum Associates.

Livingstone, S., and Bovill, M. (1999). Young people, new media. Summary report of the research project: Children, young people and the changing media environment. As accessed on http://www.lse.ac.uk/Depts/Media/people/slivingstone/young_people_report.pdf .

McCarthy, J., and Anagnost, T. (1998). MusicFX: An arbiter of group preferences for computer supported collaborative workouts., in proceedings of the ACM 1998 conference on CSCW, Seattle, WA, pp. 363-372.

Van Evra, J. (1998). Television and child development. Mahwah, N.J.: Lawrence Erlbaum Associates.

Appendix

Scenario 1

John, Mary, and Adam are goring to watch video clips together. We know how interested they are in the topics of the clips. Each clip is rated from 1 - really hate this topic - to 10 - really like this topic.

Item	John	Adam	Mary
A	10	1	10
B	4	9	5
C	3	8	2
D	6	9	7
E	10	7	9
F	9	9	8
G	6	6	5
H	8	9	6
I	10	3	7
J	8	8	6

1. They only have time to watch one clip. Which clip should they watch? Why?
2. They only have time to watch two clips. Which clips should they watch? Why?
3. They only have time to watch three clips. Which clips should they watch? Why?
4. They only have time to watch four clips. Which clips should they watch? Why?
5. They only have time to watch five clips. Which clips should they watch? Why?
6. They only have time to watch six clips. Which clips should they watch? Why?
7. They only have time to watch seven clips. Which clips should they watch? Why?

Scenario 2

John (29), Mary (32), and their grandfather Adam (81) are going to watch video clips together. We know how interested they are in the topics of the clips. Each clip is rated from 1 - really hate this topic - to 10 - really like this topic.

Clip	John	Adam	Mary
A	10	1	10
B	4	9	5
C	3	8	2
D	6	9	7
E	10	7	9
F	9	9	8
G	6	6	5
H	8	9	6
I	10	3	7
J	8	8	6

1. They only have time to watch one clip. Which clip should they watch? Why?
2. They only have time to watch two clips. Which clips should they watch? Why?
3. They only have time to watch three clips. Which clips should they watch? Why?
4. They only have time to watch four clips. Which clips should they watch? Why?
5. They only have time to watch five clips. Which clips should they watch? Why?
6. They only have time to watch six clips. Which clips should they watch? Why?
7. They only have time to watch seven clips. Which clips should they watch? Why?