Recommending TV Programs: How far can we get at zero user effort?

Patrick Baudisch
Integrated Publication and Information Systems Institute IPSI
German National Research Center for Information Technology GMD
64293 Darmstadt, Germany
+49-6151-869-854
baudisch@gmd.de

RECOMMENDATION REQUIRES EFFORT

Collaborative filtering methods have been applied to a number of domains like books, videos, audio CDs and Usenet news. These systems require some effort on the part of users before they can generate recommendations (see for example [5] on the "cold-start" problem). How much effort they require depends on domain and application. Some recommender systems for books or videos require rating a specific number of items before they generate the first prediction. If users can save money by not buying the wrong books and videos this effort might pay off. The Usenet news filter GroupLens [2] gathers ratings while users read articles so the process of building the profile and getting recommendations is interleaved. This strategy might lower the threshold for getting started. How much effort a user is willing to make depends not only on the domain but also on the personal value of the recommendation. Users who depend on the information in a newsgroup might make every required effort, but occasional users with a less serious information interest might not. To summarize: The higher the required effort the more potential users will quit.

Recommending TV programs has much in common with recommending Usenet news articles. Like news articles TV programs are a daily updating stream unlike the rather stable databases within book and video recommender systems. There is a relatively high number of programs/ newsgroup articles every day. Recommendation is supposed to help users to cope with the flood and to find the few interesting items for them.

How much effort are users going to spend on TV recommendations? Interviews showed: Not much. "I don't watch much TV anyway", "I just want to know what's on tonight" and so on. It seems that TV doesn't have much value to people, although statistics show that people do spend a lot of time watching. All users we asked already had experience with printed guides, which gave them a rather clear expectation of what a TV guide should do. Traditionally, both TV and TV guides have been broadcast media. TV makers broadcast, viewers watch. Editors write, readers read. Interactive and collaborative concepts are new here. It is obviously a risky undertaking to try to introduce new techniques in such a "where is the beef" situation.

DESIGN GOALS

The low disposition of interviewed persons to spend much effort on TV guides was the motivation for our group to try out a new way of gathering ratings. We made two basic design decisions: (a) don't ask users to do something they don't have an immediate benefit from and (b) don't overtax users with concepts unknown to them like rating, relevance or recommendation.

Figure 1 gives an overview of our recommendation engine. Before programs reach users they are first globally rated and ranked according to the *size-of-the-audience* ratings (see below). Then they are selected according to user-individual genre profiles. Personal rating offsets are added. Next, programs are presented to the users as ordered lists or tables. Individual programs in the lists and tables are color-coded to indicate the personalized ratings. From these lists and tables users choose the programs they actually want to watch and collect them in their *personal laundry* lists. The laundry list serves as a reminder which can be printed out and taken to the VCR or TV set. Preliminary user tests suggest that the laundry list is perceived as a useful tool.

To avoid any additional questioning we use the content of users' laundry lists as an input for recommendation generation. Programs in the laundry lists are interpreted as being rated high and recommended to other users. This approach has similarities with the implicit ratings deduced from reading time of newsgroup articles described in [6]. Laundry list content is processed by the two modules *size-of-the-audience rating* and *opinion leaders*.

"SIZE OF THE AUDIENCE"

The size-of-the-audience rating generates a general ranking of all programs. The rating of an individual program is calculated as the number of times the program occurs in users' laundry lists. The more often a program has been selected for viewing the higher the rating. The size-of-the-audience rating can lead users' attention to unexpected events that are not covered by personal profiles. In Germany, for example, the *Tour de France* came first into the focus of public interest when the German Telekom team was very successful in 1997. Although (or because) it is not personalized, the size-of-the-audience rating can be useful for detecting such events.

OPINION LEADERS

Since the ratings delivered by the size-of-the-audience ratings are general to all users, so-called *opinion leaders* are added as a source of more distinct, personal and higher quality recommendation. Opinion leaders are users that publish their otherwise private laundry list along with their names in a public folder. The system offers and recommends opinion leaders like it recommends genres. “You could have found all these programs more easily if you had subscribed to the genre Basketball and the opinion leader Lars Brückner. Do you want to do that now?”

How to become an opinion leader? Based on user profiles and laundry lists a program can estimate which users’ selections could be of interest for the community. Candidates should a) correlate with many other users, b) cover interest areas not yet covered by enough opinion leaders and c) use the laundry list on a regular basis. If these requirements are fulfilled the system will suggest that a user becomes an opinion leader.

Being an opinion leader *requires* no extra work. But to deserve their name we *want* them to select more accurately or even to rate or annotate programs. Why should users do the extra work? “Ways to provide compensatory benefits to those group members need to be found” [3, p. 810]. We have to reward them. One way is to organize the opinion leaders in a competitive way and to reward the winners. The success of an opinion leader can be measured in subscriptions, i.e. the number of users that selected them as their favorites. Concept *Tour de France:* a) display opinion leaders ordered by their success as a *leader board*. The currently first and most successful one wears the *yellow jersey* b) reward the best ones by giving awards and prices at the end of each *stage* c) drop the least successful one when new users apply.

DISCUSSION

The central element of the suggested no-extra-effort rating system is the laundry list. Certainly: The laundry list is a magnitude less powerful than explicit seven-items scale ratings: It provides only Boolean input, reasons for selecting and not selecting items are ambiguous and it can lead to self-fulfilling prophecies. The applied technique shows similarities to last year’s experiments at MyYahoo to correlate users according to uploaded bookmark lists.

On the other hand the laundry list approach frees us from explaining users that they will help generating recommendations using it. The direct utility of the laundry list avoids lots of the trouble of explaining, convincing and persuading users to rate. This solves much of the general problem that many users experience the promised benefit associated with rating as being far away. After all, recommendation is a rather abstract concept and the question is how many first time users are willing to learn enough about it for it to become an incentive.

But at the end, those users that want more than the mainstream recommendation of the size-of-the-audience rating *do* have to spend some effort. The system can suggest opinion leaders, but since these suggestions are based on the less informative laundry lists, users will have to check and fine tune to get equally good results as they would with systems using ratings on seven-items scales.

How do opinion leaders relate to those systems where the virtual community is selected by the computer? In those systems one’s virtual community can change by the hour. Opinion leaders are an attempt using active collaborative filtering [5] to provide “names” behind recommendations, like a large set of editors – along with the option to become one. Hill et al. state: “Recommendation and evaluations should come from people, not a black box machine or so-called agent” [4, p. 196].

The opinion leader concept is built on the assumption of role specialization: “Yet there is evidence that people naturally prefer to play distinct producer/consumer roles in the information ecology” [1, p. 60]. It divides users into two distinct groups: A large majority of consumers that contribute only a little to a mass-average and a group of self-selected leaders that are willing to spend more effort on the system for getting social or monetary reward.

How far can we get at zero user effort? We need more user tests to find out how much information we can draw out of implicit ratings and which other techniques can be applied to edit them and offer them to users. This will be subject to evaluations that will follow our online debut in May.

REFERENCES

1. Loren Terveen, Will Hill, Brain Amento, David, McDonald and Josh Creter, Phoaks a System for Sharing Recommendations, *CACM 40,* 3, 1997
2. Bradley N. Miller, John T. Riedl, Joseph A. Konstan, Experiences with GroupLens: Making Usenet Useful Again, *1997 Usenix Annual Technical Conference*, Anaheim, CA, 1997.
3. Grudin, J., Social Evaluation of the User Interface: Who Does the Work and Who Gets the BENEFIT?, Proceedings of the IFIP INTERACT’87: Human-Computer Interaction, 1987, 805-811
4. Will Hill, Larry Stead, Mark Rosenstein and George Furnas: Recommending and Evaluating Choices in a Virtual Community of Use, In *Human Factors in Computing Systems CHI*’95 *conference proceedings*,, pages 194-201, 1995
5. David Maltz and Kate Ehrlich: Pointing the way: Active collaborative filtering, CHI’95 Human Factors in Computing Systems, p. 202-209, 1995
6. Masahiro Morita, Yoichi Shinoda: Information Filtering Based on User Behaviour Analysis and Best Match Text Retrieval. SIGIR 1994: 272-281

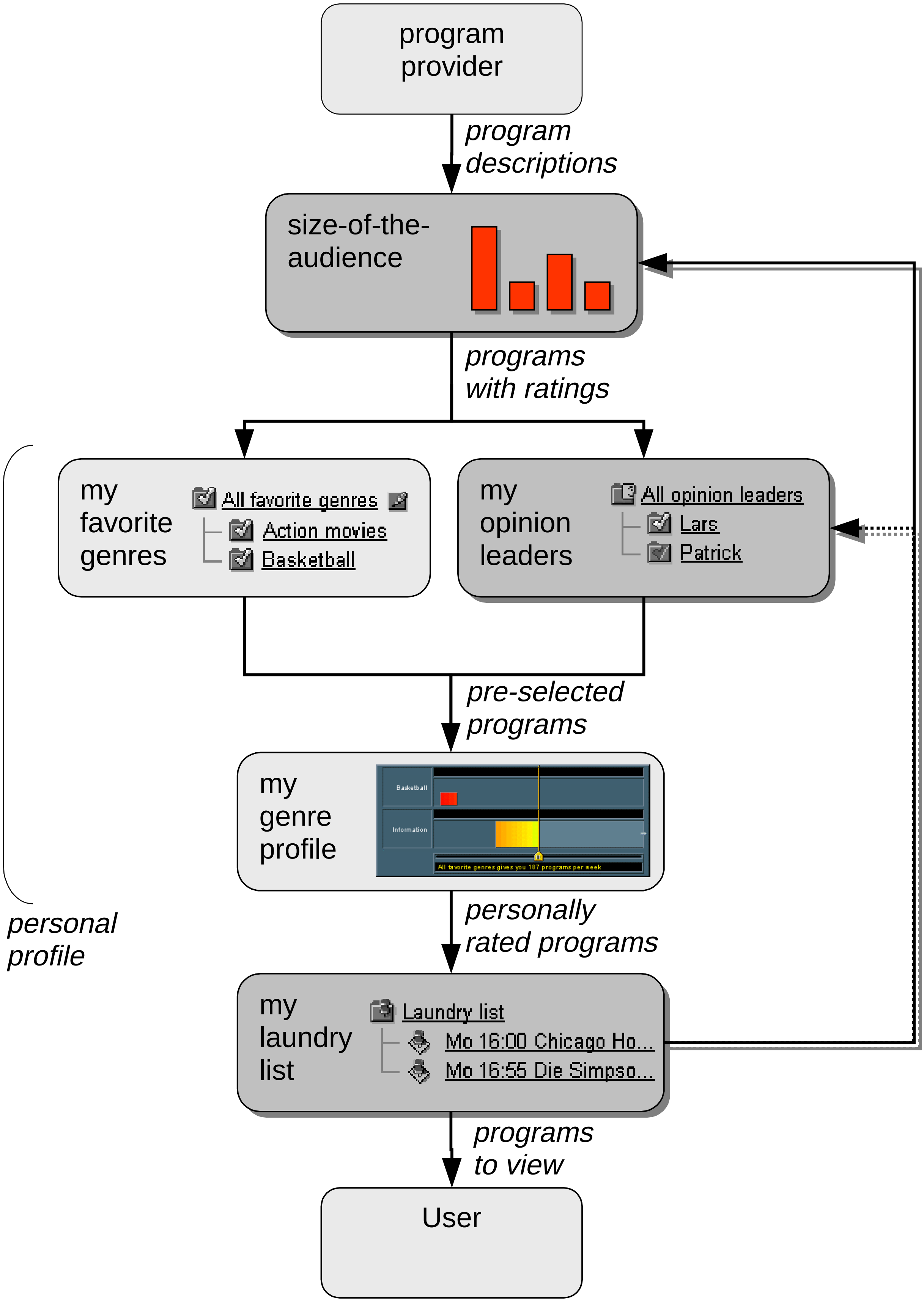

Figure 1: The path along which TV program descriptions reach the user. The three darker rendered components are discussed in this article.