Collaborative Filtering

“Other people who also liked ... bought ...”

Earlier, many sources of recommendations were either based on intrinsic properties of the items (“Mystery novels set in 19th century Canada”) or popularity (“look at all the trucks parked outside that restaurant”). Reviewers gave their personal opinion (“two thumbs up!” “****!”). Friends could say “you would enjoy...” but might not know enough possible choices.

Then Will Hill, Mark Rosenstein, George Furnas, and Larry Stead (also Jim Hollan) did the “movie browser” experiment.
Will & colleagues asked volunteers to rate movies on a 1-10 scale. Each person got a list of 500 movies; the average respondent rated almost 200 of them. To focus down the ratings and get a lot of overlaps the list of 500 for each person always included 250 particularly popular movies.

Then they computed each person's ratings as a function of other people's ratings. So person A, for example, might be modeled as a linear combination of the ratings of C, F, G, and H.

That let them predict what A would think of other movies. About 10% of the ratings were held back from the computation to serve as a test set, and those used to see how accurate the results were. These were compared with ordinary ratings by a movie critic.
What was sent to the participants

Your must-see list with predicted ratings:
7.0 "Alien (1979)"
6.5 "Blade Runner"
6.2 "Close Encounters Of The Third Kind (1977)

Correlation with target viewer:

0.59 viewer-130 (unlisted@merl.com)
0.55 bullertjane r (bullert@cc.bellcore.com)
0.51 jan_arst (jan_arst@khdld.decenet.philips.nl)
0.46 Ken Cross (moose@denali.EE.CORNELL.EDU)
0.42 rskt (rskt@cc.bellcore.com)
0.41 kkgg (kkgg@Athena.MIT.EDU)
0.41 bnn (bnn@cc.bellcore.com)
Figure 3  Two Scatterplots of Actual Ratings by Predicted Ratings. Plot on left shows movie critics as predictor ($r=0.22$). Plot on right shows virtual community as predictor ($r=0.62$) (all values are jittered for the purpose of visual presentation, 3269 predictions each for 291 users).
Personalized suggestions

These recommendations are based not descriptions of the items, nor on one person's view of the items. They represent a community model, in which some people like some things, and other people like other things. So they can do better than any recommendation which is made without knowing anything about the person using the recommendations.

By contrast, for example, the pure PageRank algorithm behind Google merely relied on the static web; the value given to each page did not depend on who was doing the search (Google no longer uses only that algorithm).
Comments by the users

People were generally enthusiastic about the movie lists. The participants had been given a choice of being named or being anonymous, and most of them wanted to be named. So the lists of movies suggested included, as shown before, the names of the people making up the model of your preferences.

This produced instances of “you found my best friend from junior high school!” and “but you didn't tell me whether these people were single or not”. The system would make recommendations for pairs of people – movies both of them would like. One person wrote to say that he viewed it as a good sign that his girlfriend liked the same movies.
Technology transfer

A MIT student came to work with us at Bellcore for a summer. He went back to MIT, where Patty Maes realized this was a good idea and expanded on it, creating a company named Firefly Networks which did music recommendations (and other agent-related work).

In 1998 Microsoft bought Firefly (rumored price $40M), but closed it down in 1999. Collaborative filtering is alive and well – it's best known in Amazon's lists of what other people who bought the books in your history bought next.
KNOW WHAT YOUR CUSTOMERS WANT EVEN BEFORE THEY DO

WORD OF MOUSE
THE MARKETING POWER OF COLLABORATIVE FILTERING

JOHN RIEDL and JOSEPH KONSTAN
Many current uses

Two years ago Yahoo! introduced a new personalization algorithm for its “Today” box, which has four news stories selected for you; clicks on items in that box are up 270%.

Google, Amazon, Netflix, and every other e-business vendor is trying to tailor results to users.

For example, if you have been searching for “travel” and search on Egypt you'll get one set of replies; if you have been searching for politics, you'll get a different set. If you search for “traffic” Google looks to see where you are and selects accordingly.

Some failures: newspapers offer personalized lists of stories, but few readers seem to want them.
Four ways to suggest things

1. Content. Search for things by descriptor, or things similar to something you've previously liked. You can be the only user.

2. Communication. Let other people say what they like: best known as “+1”. No content description involved.

3. Context of use. Watch other people and collect what they do (hyperlinking, reading, ...) but average over all users.

4. Collaborative filtering. Try to select people whose preferences are similar, relying on explicitly stated preferences.

All of these are in use, recommending web pages, music, movies, meals, hotels, and cat food.
Details on collaborative filtering

Should the data age? Are ratings from years ago as useful as ratings today? Obviously, this differs by domain: books and movies are static, restaurants and hotels change.

Are some users more reliable than others? Should ratings be weighted by the apparent consistency or utility of the person doing the rating?

Should collaborative filtering be combined with other methods such as content based or overall popularity?

Should specific kinds of recommendations be tracked (for example, rating restaurants by price, service, or cuisine)?

Do you also use data like location, gender, age, ...?
What are the problems?

*Privacy.* Most obviously, the information held by the system operator; but also, observations of recommendations can be used to infer something about the other people making the reviews. Remember that multiple systems interact: if you find somebody reporting what they've read on Facebook you can then see from Amazon what else they might have read. System operators use anonymization and data fuzzing to make this difficult (if only for commercial reasons).

*Inaccuracy.* Vendors try to manipulate the data to get their products recommended as much as possible. Mechanical Turk can be used to buy favorable reviews on Yelp for 25 cents each. Yelp does try to filter out the paid-for reviews.
In the bubble

In politics, as an example, people most read things with which they agree. Filtering is a way of helping them avoid anything else.

Researchers found that most Google queries get personalized results and these are based on data from different groups of searchers, not just individual history.

“Personal Web searching in the age of semantic capitalism: Diagnosing the mechanisms of personalization” by Martin Feuz, Matthew Fuller, and Felix Stalder, First Monday, v. 16, Feb. 2011
Usage bias

Some products have only a few varieties, each of which is used by a great many people (cars, some movies). Some products come with incredible choice, some of which only a relatively few people may have used (books, restaurants).

If the recommender system does not cope with this, it will recommend the most popular items over and over again. This reinforces the concentration on a few items.

For example, traditional library cataloging writes down about the same amount of information for each book. So content-based searching in a catalog leaves all books roughly equal. By contrast, my 1997 book *Practical Digital Libraries* has two reviews on Amazon; *Harry Potter and the Deathly Hallows* has 3,704 reviews.

Another view of this problem: who writes the first review, if nobody looks at anything until another reviewer suggests it?
Netflix gave the computing community a challenge in 2006-2009, to try to improve algorithms for suggesting movies to people. They distributed a huge data set – 99 million ratings for training, 3 million for testing. The goal was to improve over the way Netflix started out recommending movies with the grand prize being given for a 10% improvement.

After two years of improvements of 8% and 9%, the prize was won in 2009 by a team combining researchers from AT&T Laboratories, Commando Research & Consulting, and Pragmatic Research.
Privacy again

Although Netflix anonymized the dataset, researchers at the University of Texas compared the lines in the dataset with ratings posted at IMDB, and identified some people, whose other movie preferences (and other information from the postings at IMDB) were now public. A lawsuit resulted, with a claim that Netflix had violated the law enforcing confidentiality of video rental records (an outcome of the confirmation hearings for Robert Bork, when a newspaper disclosed that he had rented *A Day at the Races* and *The Man Who Knew Too Much*).

The lawsuit, *Doe v. Netflix*, was filed in 2009 and as far as I can see is still unresolved.
Both Google and LinkedIn have recently made changes to their rules (Google very loudly, LinkedIn somewhat quietly). But every major e-commerce company is using collaborative filtering techniques. The Google change was to say that data from multiple Google applications (Gmail, Google Docs, ...) will be merged. This gives you better predictions of what you might search for and want, at a cost in privacy. LinkedIn now wishes to use your name and picture in advertising – this is not directly about collaborative filtering.

In general, as the various systems get bigger: a) they'll be able to give you better recommendations, b) your data will be more private to outsiders (since some of the techniques used depend on rarely seen items), c) the company collecting your data will have more of it in one place.
Future applications

Recommending movies is perhaps less important than what one might do with this technology.

Can you run an online dating service with it? See Brozovsky, “ColFi - Recommender System for a Dating Service”

What about hiring and promotion? See Baez and Casati, “Resource Space Management Systems”

Can you predict diseases somebody might get by collaborative filtering on their medical history? See Davis, Chawla, Blumm, Christakis and Barabas, “Predicting individual disease risk...”

Can you find terrorists? See Al Hasan, Chaoji, Salem, and Zaki, “Link prediction using supervised learning”.