Collaborative Filtering: GroupLens

- High volume and personal taste makes Usenet news an ideal candidate for collaborative filtering techniques

- GroupLens project started in 1992 at University of Minnesota completed a pilot study to establish the feasibility of using collaborative filtering for Usenet news
GroupLens

- Usenet news is high volume, high turnover discussion list service on the Internet
- The newsgroups may carry hundreds of messages each day
- While in theory the newsgroup organization allows readers to select the content that most interests them, in practice most newsgroups carry a wide enough spread of messages to make most individuals consider Usenet news to be a high noise information source

GroupLens …

- Further more, each values a different set of messages. Both taste and prior knowledge are major factors in evaluating news articles
- For example, take comp.lang.c++ newsgroup. Introductory questions and answers may be uninteresting to an expert programmer. Debates over subtle and advanced language features may be useless to the novice
GroupLens : Design decisions

• Several critical decisions were made as a part of the pilot study:
  – Integration with existing news reading applications
  – Support facility to rate an article with minimum effort on the part of the user
  – To provide prediction of the rating that the system expects a user will give to an unseen article

GroupLens: Challenges

• The overall challenges for the project were:
  – Integration of collaborative filtering into an information system with existing users, existing applications and interfaces
  – Addressing the dynamic, distributed nature of Usenet news. Articles have short life spans and there is no central repository of news articles
  – Working with extremely sparse set of ratings. Typical users read only a tiny subset of total Usenet news articles
  – Delivering acceptable performance to users and providing mechanisms to scale the system as number of users and articles grow
GroupLens: Architecture

Figure 4. GroupLens architecture overview. Usenet clients connect to the GroupLens server through the GroupLens client library, and to a separate NNTP server as usual. The GroupLens server accepts ratings and provides predictions for articles delivered by the NNTP server.

GroupLens: Server Architecture

Figure 3. GroupLens server architecture. The beige box encloses the GroupLens server. The ratings broker serves as a single point of contact for clients to the server.
GroupLens: News Article Statistics

Figure 2. Ratings profiles for four Usenet news groups. The percentage of articles assigned each rating varies significantly from newsgroup to newsgroup. Most articles in rec.humor were given the worst rating (1 out of a possible 6), while the ratings in comp.os.linux.development.system were distributed more uniformly.

Figure 3. User pair correlations for three newsgroups. One way to compare the similarity of users is to compute the Pearson coefficient between their ratings. Here, the number of user pairs with each Pearson coefficient is plotted for three different newsgroups. The presence of many high correlations in the rec.humor newsgroup indicates general agreement about quality in that domain. In the moderated newsgroup rec.food.recipes, correlations are mainly narrowly distributed about the mean, suggesting that individual tastes matter more in this domain.
Conclusions

• In the trial, the authors found out that highly rated articles are read more often than less highly rated articles

• Usenet represents a different set of challenges to collaborative filtering than domains such as music or movies where new items are relatively infrequent and lifetimes are long
Current Work

- After their famous Usenet trial, the project has expanded its scope to research overall information filtering solutions, integrating in content-based methods as well as improving current collaborative filtering technology.
- MovieLens is a web-based recommender system for movies, based on GroupLens technology. MovieLens is not only a great service to moviegoers, it is also an experimental data source and a framework for researching user interface issues relating to recommender systems.
- Net Perceptions develops, sells, and supports recommender system software based on the GroupLens technology developed in this research group.

References