Position Statement — Explanations in Recommender Systems

Jonathan L. Herlocker
herlocke@cs.umn.edu
Dept. of Computer Science and Engineering
University of Minnesota
http://www.cs.umn.edu/~herlocke

1 Introduction

Automated collaborative filtering (ACF) systems predict a person's affinity for unexperienced items based on the past experiences of that person and the past and current experiences of a community of people. ACF systems have been successful in research, with projects such as GroupLens[7], Ringo[10], and Video Recommender[4] gaining large followings on the Internet. Commercially, some of the highest profile web sites like Amazon.com, CDNow.com, and MovieFinder.com have made successful use of ACF technology.

While automated collaborative filtering systems have proven to be generally accurate, their failure rates still remain unacceptable for certain domains or individuals. While a user may be willing to risk purchasing a music CD based on the recommendation of an ACF system, he will probably not risk choosing a honeymoon vacation spot based on such a recommendation. However, there is no reason why the higher- risk domains should not benefit from ACF technology.

There are several key problems obstructing the development of trust in an automated collaborative filtering system as a decision aid. The primary problem is that current ACF systems are stochastic processes and will make mistakes from time to time, no matter how well implemented. A secondary problem is that ACF systems are black boxes, computerized oracles that give advice, but cannot be questioned. A user has no feeling when to trust a recommendation and when to doubt a recommendation. These problems can prevent acceptance of ACF systems as decision support aids.

Explanation capabilities provide a solution to building trust and may also improve the decision-making performance of automated collaborative filtering systems. An explanation behind the reasoning of a recommendation provides transparency into the workings of the ACF system. Users will be more likely to trust a recommendation when they know the reasons behind that recommendation. Explanations will help users understand the process of ACF, and know where its strengths and weaknesses are.

Research is necessary to determine how effective explanation facilities will be with ACF systems, and what is the proper way to implement them. The remainder of this statement describes our plans for user experiments related to the development of explanation facilities in ACF systems. We begin by describing the errors that are introduced into a automated collaborative filtering systems.

2 Errors in Collaborative Filtering

Errors in recommendations from collaborative filtering systems are unavoidable. Recommendations are based on a combination of ratings supplied by other users. Those who have conducted surveys will tell you that when you ask a person a question like "how much do you like this item?", her answer may vary greatly depending both on the environment in which the question is asked and her current mental state. Thus the same question asked at different times may yield different answers. Now take a recommendation that is a weighted aggregate of hundreds or thousands of human ratings, and you have a computation with many sources of error. If your model of computing recommendations is sound, then your computational process can use large numbers of people to extract the true patterns from the random variance, but you can never escape the occurrence of occasional statistical errors.

Even if it were possible to capture human ratings truly, the models which we use to mimic word-of-mouth recommendations are imperfect. There are many examples of model-based error. We often assume that a user's interests are consistent over time, when they are not. We may assume that people will have similar rating distributions. Computing a prediction based on a weighted combination of similar user's ratings seems intuitively correct, but we know that it is not always accurate. Our measure that compute similarity between users may introduce error. All of these factors and more introduce error into the recommendation process above and beyond the random error from the human input.

3 Explanations

The ability to request an explanation provides us with a mechanism for handling the unavoidable error that comes with a recommendation. Consider how we as humans handle suggestions as they are give to us by other humans. We recognize that other humans are imperfect recommenders. In the process of deciding to accept the suggestions, we might consider the previous performance of the recommender or we may compare how the recommender's general interests compare to ours in the domain of the suggestion. But if there is any doubt, we will ask "Why?", and let the recommender explain their reasoning behind a suggestion. Then we can analyze the logic of the suggestion and determine for ourselves if the evidence is strong enough.

It seems sensible to provide explanation facilities for recommender systems such as automated collaborative filtering systems. Previous work with another type of decision aide — expert systems — has shown that explanations can provide considerable benefit. The same benefits seem possible for automated collaborative filtering systems. Most expert systems that provided explanation facilities, such as MYCIN, used rule-based reasoning to arrive at conclusions. Explanation in that domain involved providing a trace of the rules used. Since collaborative filtering does not generally use rule-based reasoning, the problems of explanation there will require different approaches and different solutions.

Building an explanation facility into a recommender system can benefit the user in many ways. It removes the black box from around the recommender system, and provides transparency. Some of the benefits provided are:

- **Justification**. User understanding of the reasoning behind a recommendation, so that he may decide how much confidence to place in that recommendation.
- **User Involvement**. User involvement in the recommendation process, allowing the user to add his knowledge and inference skills to the complete decision process.
- **Education**. Education of the user as to the processes used in generating a recommendation, so that he may better understand the strengths and limitations of the system.
- **Acceptance**. Greater acceptance of the recommender system as a decision aide, since its limits and strengths are fully visible and its suggestions are justified.

Together, the potential for increasing the impact of automated collaborative filtering systems is great.

4 Related Work

Expert systems commonly employed explanations as part of the user interface to their expert knowledge and reasoning engine. The best documented use of explanation occurred in the early medical expert system MYCIN[1]. MYCIN provided explanations by translating traces of rules followed from LISP to English. A user could ask both why a conclusion was arrived at and how much knew about a certain concept. Other work describing explanation facilities in expert systems includes Hovitz, Breese, & Henrion[5], and Miller & Larson[9].

Work related to explanations can be found in cognitive science, psychology, and philosophy. Johnson & Johnson have begun research into the components of a unified theory of explanation in human-computer interfaces[6]. To support their theories, they performed empirical experiments to help determine the logical components of an explanation. There has also been considerable study into the psychology of questioning and question answering with humans and how it can be applied to human- computer interfaces[8, 3]. Philosophers have studied the rules and logic of human discourse — such as in the book "The Uses of Argument" by Toulmin[11].

5 Research Questions

There are three key research questions that we hope to answer about the use of explanations with automated collaborative filtering systems.

- **Can explanation facilities increase the acceptance of automated collaborative filtering systems?** We believe that by providing transparency into the workings of the ACF process, we will build users' confidence in the system, and increase their willingness to use the ACF system as a decision aid.
- **Can explanation facilities increase the decision-making performance of ACF system users?** By performance, we mean the ratio of good decisions to bad decisions, where decisions are made based on recommendations and explanations given to them by the ACF system.
- **What models and techniques are effective in supporting explanation in an ACF system?** An ACF system's computational model can be complex. What is the right amount of detail to expose? How much information is too much? There are many such questions that can be answered through experimentation with users.

6 Explanation Models and Techniques

Explaining a prediction given by an automated collaborative filtering system requires explanation of a complex mathematical process. Planning the explanation and presenting the explanation components in a informative and understandable manner is not a small challenge. Most current recommender systems provide a simple interface, usually a tabular ranked list of best-bet recommendations. However, large amounts of data are used to compute a recommendation and presenting this data in a usable manner will require a more complex interaction.

The manner in which an explanation interacts with the user will affect the acceptance and performance of a system. Therefore, we believe that choosing the right model for interaction is important. For example, following are three possible models for explanation:

- **Data-Explorative Model**. In this model, the user can explore the data on which the recommendation was based. No attempt is made to explain the mathematical process used to create the prediction. Initially, the user may be shown what the recommender system thinks are the *key* data, but the user can select and zoom in on other parts of the data. This process allows them to validate using their own personal approaches.

- **Process-Explorative Model**. Here the recommender system attempts to explain at a high level the mathematical process used to arrive at a recommendation (for example by using as a flowchart). The user can zoom in on steps in the computation, and may be allowing to make changes to the computation.
- **Argumentative model**. In this model, the recommender system works as an agent that attempts to use logical argument techniques to support a conclusion. At each stage, the system makes a claim, and the user can challenge the inference and data behind a claim. This model minimizes the amount of data that the user must process at one time.

Further design and experimentation will help determine the right model for interaction while performing explanation.

7 Experiments

We are currently building explanation facilities for the MovieLens movie recommendation services[2]. MovieLens uses automated collaborative filtering technology to provide personalized recommendations for movies. Users rate movies they have seen on a scale from 1 to 5, and MovieLens predicts other movies that they are most likely to enjoy. In its current incarnation, MovieLens is currently a black box, providing no reasoning behind a prediction, or any indication of the confidence behind a prediction. The explanation facilities will allow the user to explore the system's reasoning behind a prediction, and perhaps also provide other sources of information that may support the prediction or provide evidence against it.

Participants in the experiment will be volunteer users of the MovieLens site. Current users of MovieLens will be offered the opportunity to try the new interface. Volunteer users will be randomly assigned to interfaces using different explanation techniques. They will be asked to return to MovieLens and fill out a very short questionnaire for every movie that they decide to see based on recommendations from MovieLens. Data from these questionnaires will be used to determine the error rates. Post-experiment questionnaires will be used to assess general appeal and success of the interface. Retention rate of users and length of use will be measured though login records.

We should have results from the experiments available in time for presentation at the recommender workshop.

8 Conclusion

Explanations have shown themselves to be very successful in previous work with expert systems. From this knowledge, it seems intuitive that they will prove to be successful in interfaces to automated collaborative filtering systems. The challenges will be to extract meaningful explanations from computational models that are more ad hoc than rule-based expert systems, and to provide a usable interface to the explanations. The result will hopefully be decision aids that are more accepted, more effective, more understandable, and which give greater control to the user.

References

[1] Bruce Buchanan and Edward Shortliffe, editors. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project.* Addison–Wesley, Reading, Massachusetts, 1984.

[2] Brent J. Dahlen, Joseph A. Konstan, Jon Herlocker, Nathaniel Good, Al Borchers, and John Riedl. Jump-starting movielens: User benefits of starting a collaborative filtering system with "dead data". Technical Report TR 98-017, University of Minnesota, 1998.

[3] Arthur C. Graesser and John B. Black, editors. *The Psychology of Questions.* Lawrence Erlbaum Associates, 1985.

[4] Will Hill, Larry Stead, Mark Rosenstein, and George Furnas. Recommending and evaluating choices in a virtual community of use. In *Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems*, pages 194–201, 1995.

[5] E. J. Horvitz, J. S. Breese, and M. Henrion. Decision theory in expert systems and artificial intelligence. *International Journal of Approximate Reasoning*, 2(3):247–302, 1988.

[6] Hilary Johnson and Peter Johnson. Explanation facilities and interactive systems. In *Proceedings of Intelligent User Interfaces '93*, pages 159–166, 1993.

[7] Joseph A. Konstan, Bradley N. Miller, David Maltz, Jonathan L. Herlocker, Lee R. Gordon, and John Riedl. GroupLens: Applying collaborative filtering to Usenet news. *Communications of the ACM*, 40(3):77–87, March 1997.

[8] Thomas W. Lauer, Eileen Peacock, and Arthur C. Graesser, editors. *Questions and Information Systems.* Lawrench Erlbaum Associates, 1992.

[9] Christopher A. Miller and Raymond Larson. An explanatory and "argumentative" interface for a model-based diagnostic system. In *Proceedings of UIST '92*, pages 43–52, 1992.

[10] Upendra Shardanand and Patti Maes. Social information filtering: Algorithms for automating "word of mouth". In *Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems*, pages 210–217, 1995.

[11] Stephen Edelston Toulmin. *The Uses of Argument.* Cambridge University Press, 1958.