

Building and Exploiting User Models

Elaine Rich
Computer Science Department
Carnegie-Mellon University
Pittsburgh, Pa. 15213

This paper describes the issues involved in building and exploiting individual user models in order to guide the performance of an interactive system. A system called Grundy, that recommends novels to people, is described and analyzed as a forum in which to explore those issues. One of the major techniques exploited in Grundy is the use of stereotypes as a mechanism for quickly producing an initial approximation to a model of the user. Experiments with the system show that despite the fact that stereotypes are inherently imperfect, their use does contribute significantly to the system's ability to build useful models of its users.

1. Introduction

As computers come to be used by a larger number of people to help perform a great variety of tasks, it is becoming more and more important for them to be easy for people to use. There are many factors that can contribute to the ease of use of a computer system, ranging from the good design of terminals to the speed of the system's response, the appropriateness of its response, and the naturalness of its input and output languages. Yet another factor that can significantly affect the ease with which people can use a system is the ability of the system to tailor its behavior to the specific needs of an individual user. It is this aspect of the habitability problem that will be discussed in this paper.

In order for a system to be able to respond to the individual needs of its users, it must have access to some kind of model of each user. Because the term "user model" can be used to mean several different things, it is important to specify what we mean by it. The three major dimensions along which user models can be classified are:

- Are they models of a canonical user or are they models of individual users?
- Are they constructed explicitly by the user himself or are they abstracted by the system on the basis of the user's behavior?
- Do they contain short-term, highly specific information or longer-term, more general information?

*This research was sponsored in part by the Defense Advanced Research Projects Agency (DOD), ARPA Order No. 3597, monitored by the Air Force Avionics Laboratory Under Contract F33615-78-C-1551. The author was supported by a grant from the Xerox Corporation.

The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the US Government.

The work described here has concentrated on user models in one corner of this space. Models of individual users were built because of the facility for personalization they provide that is lacking in a canonical user model. Implicitly constructed models were used because of the inherent inaccuracy and the annoyance of requiring users to construct their own models of themselves. The choice of position along the third dimension was much less clear cut. Both short-term knowledge (such as the topic currently being discussed or the goal now being pursued) and long-term knowledge (such as the level of the user's knowledge about the problem domain or goals that are likely to be pursued) can be important to the performance of an interactive system. But it appeared likely that different techniques would be appropriate for dealing with the two types of knowledge. Models of long-term user characteristics were chosen because very little work has been done with them, while short-term user models were already being explored, particularly in the context of natural language understanding. So the rest of this work will deal explicitly with the issues involved in individual user, implicit, long-term user modeling.

2. Some Requirements for an Individual-User, Implicit, Long-Term User Modeling System

The model must be built dynamically, as the system is interacting with the user and performing its task. Because the model is to be built implicitly and because users do not want to be forced to answer a long list of questions before they can begin to use the system, it is necessary that the system be able to exploit a partially constructed user model and that it be able to tell, as it is performing its major task, when it is appropriate to update the user model and how to do so.

The system should be able to infer as much knowledge about the user from as little information as possible. This is important if the user is not to be required to answer a large number of questions about himself. The major technique that can be exploited to do this is the use of stereotypes. A stereotype can be activated on the basis of a single action and can suggest a large

array of likely characteristics of the user. Stereotypes serve a function very similar to that of scripts [2]; they group together frequently co-occurring events to provide useful predictions than can guide the system.

The user modeling system must be able to cope with the uncertainty that will arise from the fact that most of what it believes about the user is really just a good guess. Because most of the knowledge about the user will be based only on likely inferences drawn from his behavior, the system must associate with those inferences some measure of how confident it is of them. It must also remember the basis for each inference so that when conflicts arise (as they are sure to do), it will be able to resolve them in some reasonable way. Of course, to do this, it must also possess some conflict resolution rules. In this respect, user modeling systems are similar to other probabilistic reasoning systems, such as MYCIN [3].

It must be possible to modify the data base of stereotypes so that they will eventually characterize the system's users as accurately as possible. Since it will almost never be possible to construct stereotypes that are completely accurate and since the stereotypes are so important to the function of the system, it is important that the system monitor their accuracy and change both them and their triggers as warranted by the actions of the users.

3. A Brief Description of Grundy

In order to have a forum in which to explore these issues and to be able to make a convincing case that some solutions had been found, it was deemed necessary to tackle the issues in the context of a specific task domain and actually to build a system into which the solutions to the issues could be incorporated. The task that was chosen for this experiment is to suggest to users novels that they might like to read. This is an appropriate domain since a wide variety of people read novels, but they do not all like the same ones; and there is enough flexibility in the task that knowledge about the individual user can be exploited.

The novel recommending program, called Grundy, first asks one or two questions, on the basis of which it begins to build its model of the user, usually by evoking one or more stereotypes. Then it begins the process of recommending novels. Each time it selects a book, it asks the user whether or not he has read it. If he has, he is then asked whether he liked it. If he did, Grundy can update its model of him on the basis of the known characteristics of the book. If he did not like the book, Grundy attempts to find out why not and then to update its model of the user appropriately. If, on the other hand, the user has not read the book, then he is given a description of it and then asked whether he thinks he would like it. If he thinks he would, Grundy need do nothing -- it appears to be on the right track. But if he thinks he would not like it, Grundy must try to discover why not and then to update its model of him. Thus, as Grundy interacts with the user, its model of him will evolve and begin to contain specific knowledge about him, inferred on the basis of his responses, in addition to its early beliefs, which were principally based on the predictions of stereotypes.

4. The Performance of Grundy

In order to collect evidence that Grundy is successfully exploiting user models, twenty-three people were observed using the system. Several aspects of Grundy's performance were investigated.

4.1. Grundy's Success at Recommending Novels

Although building the ideal novel recommending program was not the primary goal of this research, it is important to establish that Grundy exhibits some form of intelligent action in order to show that the user modeling techniques are effective. In order to measure the system's success, a small experiment was conducted at the end of each conversation Grundy had with a user. Some books were selected randomly from the data base and then suggested to the user. It was then possible to compare Grundy's rate of success at recommending books that looked good to the users both when it was exploiting the user model and when it was not. Figure 1 shows the results of the comparison. It shows the number of good suggestions (i.e. the user said he might like) and the number of bad ones (i.e. the user said he would not like) in both the controlled and random modes. The data clearly demonstrate that the user models do contribute significantly to Grundy's performance as a novel recommender.

	CONTROLLED	RANDOM
GOOD	102	54
BAD	39	60

Figure 1: GRUNDY's Success Rate

4.2. Grundy's Success at Building Models of Its Users

Several attempts were made to test more directly Grundy's success at building models of its users, by comparing the models built by Grundy to self-descriptions provided by the users. This is not a perfect measure of Grundy's success, however, because people's descriptions of themselves do not tend to be very accurate (or even consistent). Despite this, the data indicate that Grundy is doing quite a good job at building models of its users. This conclusion is not, of course, surprising in light of the data on Grundy's success at recommending books.

4.3. Learning in Grundy

As has already been discussed, it is important that the stereotypes and triggers that form the bulk of the database of a user modeling system be able to change to reflect the actual body of users of the overall system. On the other hand, it is important that no single user be allowed to have a momentous effect on a

system's global view of the class of all users. So in running Grundy with about twenty users, only the most commonly invoked stereotypes could be expected to change significantly. The most frequently activated stereotype was the MAN stereotype, and it did show interesting and significant change over the period Grundy was run. For example, the stereotype initially predicted that men would like books that are very fast-moving and full of suspense. The values of both of those characteristics decreased considerably on the basis of Grundy's experience with its users. Both of these changes can easily be explained by the difference between the group the stereotype was originally intended to characterize (all adult male Americans) and the group it was actually applied to in practice (some adult male intellectuals). The people who used the system, although they were men and did seem to like fast-moving, exciting books, also tended to like philosophical novels and literary classics, which tend to be much calmer. Thus by the end of the experiment, Grundy had a more accurate picture of the men it was encountering than it had had at the outset. This suggests that the learning mechanisms are capable of causing at least some sorts of improvement in the stereotype data base of a user modeling system.

4.4. The Effectiveness of Stereotypes

In looking at the results of GRUNDY's attempts to use stereotypes to build useful models of people, one is reminded of one important characteristic of stereotypes and thus of systems that use them. Although a stereotype may often provide highly appropriate and useful information about a person, it is, at best, an expression of a tendency and not an absolute truth. One example will serve to illustrate this point clearly.

Several stereotypes in GRUNDY predict that people who fit them will be interested in the conflict "art-vs-science". Those predictions caused GRUNDY to believe, correctly, that several users were interested in that conflict and thus to recommend books involving it. But one person, who fit three stereotypes that predict an interest in that conflict said, on the second questionnaire, that he was definitely not interested in it. Figure 2 shows all the situations in which GRUNDY believed the user was interested in art-vs-science.

5. Conclusion

If one were to distill this thesis into its most essential points, there would, I think, be three.

The first is the awareness of the potential usefulness of user models of this sort. This awareness is just now becoming more important as the use of computers spreads further and further into the daily lives of more and more people. If this thesis does nothing except make people think about building these computer systems so that they can better serve their users, this effort will not have been wasted.

But realizing a need and satisfying it are not the same thing. The second significant point of this thesis is that, for at least one such system, in at least one domain, an individual user model was made to work and to guide a performance system successfully.

User	Stereotypes Predicting Interest	Interested?
4	Intellectual, Contemplative, Scientist	yes
6	Intellectual	yes
7	Intellectual, Contemplative	yes
8	Intellectual	yes
9	Contemplative	yes
10	Intellectual, Artist	yes
11	Intellectual	yes
13	Intellectual	yes
20	Intellectual	yes
22	Scientist, Artist, Contemplative	NO

Figure 2: An Illustration of the Fallibility of Stereotypes

And yet the third contribution, after arguing for necessity, and demonstrating feasibility, is suggesting technique. Not only are user models useful and doable, specific techniques for building them have been developed. Although no claim is being made that these techniques are the perfect ones, they have been shown to be workable in one domain and extensible to others.

These three points, the usefulness, the feasibility, and the techniques of user modeling, constitute the major contribution of this thesis.

A more detailed discussion of all of the issues that were mentioned in this paper, as well as several that were not, including a description of the actual mechanisms employed, and a discussion of how to extend the ideas to other domains, can be found in [1].

References

- [1] Rich, E. A.
Building and Exploiting User Models.
PhD thesis, CMU, 1979.
- [2] Schank, R. C. and R. P. Ableson.
Goals, Plans, Scripts and Understanding: An Enquiry into Human Knowledge Structures.
Erlbaum Press, 1977.
- [3] Shortliffe, E. H.
MYCIN: Computer-based Consultations in Medical Consultations.
American Elsevier, 1976.