



## Feature-based and Clique-based User Models for Movie Selection: A Comparative Study

JOSHUA ALSPECTOR<sup>1</sup>, ALEKSANDER KOŁCZ<sup>2</sup> and N. KARUNANITHI<sup>3</sup>

<sup>1</sup>*ECE Department, University of Colorado, Colorado Springs, Colorado 90918, U.S.A.*

*e-mail: josh@eas.uccs.edu*

<sup>2</sup>*ECE Department, University of Colorado, Colorado Springs, Colorado 90918, U.S.A.*

*e-mail: ark@eas.uccs.edu*

<sup>3</sup>*1F-319B, Bellcore, 445 South Street, Morristown, NJ 07960.*

*e-mail: karun@bellcore.com*

(Received 16 August 1996; accepted in revised form 12 January 1998)

**Abstract.** The huge amount of information available in the currently evolving world wide information infrastructure at any one time can easily overwhelm end-users. One way to address the information explosion is to use an ‘information filtering agent’ which can select information according to the interest and/or need of an end-user. However, at present few information filtering agents exist for the evolving world wide multimedia information infrastructure. In this study, we evaluate the use of feature-based approaches to user modeling with the purpose of creating a filtering agent for the video-on-demand application. We evaluate several feature and clique-based models for 10 voluntary subjects who provided ratings for the movies. Our preliminary results suggest that feature-based selection can be a useful tool to recommend movies according to the taste of the user and can be as effective as a movie rating expert. We compare our feature-based approach with a clique-based approach, which has advantages where information from other users is available.

**Key words:** User modeling, information filtering, collaborative filtering, feature extraction, neural networks, linear models, regression trees, bagging, CART.

### 1. Introduction

In recent years, computer-network-based information services have gained wide acceptance both within commercial and noncommercial sectors. This is evidenced by the explosion of information utilities and services on the World-Wide-Web portion of the internet. The information content in such services is mostly textual. However, the currently evolving internet is expected to support not only a variety of text-based information services but also various multimedia (hypertext, audio and video-based) information services. Some of the potential application domains in which the information infrastructure is likely to have impact are: banking at home, access to electronic libraries, distance learning and laboratories, delivery of news and entertainment on demand, electronic shopping malls, law enforcement and security alertness, legal services, national health care and weather services,

and telecommuting. Thus, the internet has the potential to change the way we work, communicate, travel, and generally access information.

The huge amount of information available in the information infrastructure at any one time can easily overwhelm end-users. Even within existing computer-network-based information services, providing information that is of interest to a particular end-user is not an easy task. For example, filtering relevant e-mail information in the internet is not easy because a single message may be sent over a set of mailers (e.g., filtering out messages that had passed through a particular server may stop many desired messages and still allow 'e-mail spam' to arrive via a different route), a message may consist of a 'thread' (a sequence of 'replies' to the original mail), or the header may not reflect the actual content. This situation is likely to worsen in future multimedia information networks unless the end-user has the ability to filter information based on what is relevant to him/her.

Several useful text-based tools exist for navigational purposes (Obraczka, Danzig, and Li, 1993) on the internet. An example of a common-to-all user interface for the internet is Mosaic, Netscape(TM) or Microsoft Internet Explorer(TM). These are hypertext-based easy-to-use interfaces built on top of various internet navigational and browsing tools such as Gopher, WAIS and World Wide Web, and incorporating search/retrieve services such as Archie and FTP. Even with such a common interface, these navigational aids require network support, and active participation of the end-user. Moreover, despite rapid progress, at present few equivalent filtering systems exist for the evolving multimedia information infrastructure.

An important part of an information filter is a user model to predict what a target user would like to filter. In the most straightforward approach, the users may be required to state their preferences in a more or less structured way (e.g., by way of creating a personalized profile). However, such an approach has obvious limitations and what is really desired is a system capable of modeling a particular user's preferences (or taste) on the basis of the actual choices and decisions made by the user during the course of his/her interaction with the information provider. In this study we concentrate on one particular type of the user modeling problem involving movie selection. This is a relevant problem considering that the data on the majority of movies made so far is being made available over the internet (e.g., the Internet Movie Database<sup>1</sup>), and in the future video-on-demand services may well be able to supply customers with any movie desired. Considering the very large number of potential choices, the problem of selecting a movie conforming with the user's taste will certainly be of importance. Many other applications of recommendation systems (e.g., music selection) have been reported in the literature (Maes, 1994).

The problem of designing a movie rating system tailored to the preferences of a particular user is approached from two angles. In the first model, we use a collaborative filtering approach (Goldberg et al., 1992), where a 'clique' of users whose taste is similar to that of the target user is found. Subsequent predictions of