On the Construction of TV Viewer Stereotypes Starting from Lifestyles Surveys[1]

Cristina Gena and Liliana Ardissono
Dipartimento di Informatica, Università di Torino
Corso Svizzera 185; I-10149 Torino, Italy
{cgena, liliana@di.unito.it}

Abstract. This paper describes how a user modeling knowledge base for personalized TV servers can be generated starting from an analysis of lifestyles surveys. The aim of this research is the construction of well-designed stereotypes for generating adaptive electronic program guides (EPGs) which filter the information about TV events depending on the user's interests.

INTRODUCTION

With satellite and cable TV, the convergence of TV and Internet and the advent of digital networks, the offer of TV channels will dramatically increase, overloading users such as in the WWW. In this scenario, personalized filtering techniques are essential to reduce the huge number of broadcasted programs; see [7] for references to the related work. This paper describes the design of a knowledge base for user modeling starting from an analysis of lifestyles surveys. The work is part of a more extensive research for the development of a system generating adaptive EPGs. The system includes a User Modeling Component which exploits alternative strategies to handle the user models [2,7] and uses stereotypical information to personalize the interaction since the first time the user logs into the system. The stereotypes enable the system to initialize the user model with estimates on the user's interests starting from a small amount of information about him [8]. Then, the user model is revised during the interaction, after monitoring his behavior.

STEREOTYPICAL KNOWLEDGE

The exploitation of sociological stereotypes seems to be usual in the mass-media world. Thus, we decided to generate our user modeling knowledge base starting from an analysis of existing surveys about TV viewers. Particularly, we examined the lifestyles surveys which cluster the population into groups according to consumers'

---

[1] This work has been developed in cooperation with TI LABS within the project "Modellizzazione automatica dell' utente nell'interazione su Web". We thank L. Console and P. Torasso for having contributed to this work with suggestions and discussions. Moreover, we thank F. Portis for his helpful advice and his work in the development of the system.

preferences, socio-cultural trends and homogeneous behaviors. Given the completeness of the considered viewpoint, we decided to build our stereotypes knowledge base starting from a lifestyles study, *Sinottica*, conducted by Eurisko data analyzers [4]. However, the information regarding the lifestyles is not defined in a formalized way; thus, we chose our own representation format, based on [1]. We assumed a plausible correlation among homogeneous user groups and their preferences, then we structured the stereotypes in two main parts: *i) a profile*, containing the classification data of individuals belonging to the represented stereotype; *ii) a prediction part*, containing the typical preferences of such individuals. While we use classification data to evaluate how close the individual TV viewer matches a stereotypical description, the preferences are used to make initial predictions on his preferences for TV programs. The Eurisko lifestyles description has been used for the profile of the stereotypes, which has been further split into two main parts: *personal data* (age, gender, education level, type of job, geographic zone) and *interests*.

Regarding the prediction part of the stereotypes, we initially analyzed a survey on the exposure to the TV, made by Eurisko in collaboration with Auditel[2], which measures the audience of each lifestyles class [5]. We analyzed these information items considering the average audience reception rating[3] and the share[4]. To obtain more detailed information about viewer preferences, we decided to merge the Eurisko/Auditel audience data and the information about interests. We assumed an existing correlation between a user's interests and the programs concerning his interests. Moreover, we refined such collected data by comparing it with the audience data of Eurisko Big Map [6], a sociographic analysis of the Italian society which places the Eurisko lifestyles in a multi-dimensional map and provides the corresponding audience data for each national TV channel. Finally, we included two temporal dimensions in the prediction part: the watching frequency and the viewing time. The first one represents the TV watching frequency. The second one represents the situations where user preferences may occur during four time intervals in which the day is subdivided (morning, afternoon, evening, night).

STRUCTURE OF THE STEREOTYPES

Similar to the representation of stereotypical information adopted in SETA [1], we defined a family of stereotypes describing lifestyles. The features of a stereotype are represented as slots, according to the formalism introduced in Torasso and Console [9]. Each slot includes an *Importance* facet, representing the impact of the feature on the overall description of the stereotype. Moreover, it includes a list of <*linguistic value*, *likelihood*> pairs, which can be interpreted as a probability distribution for the linguistic values of the feature.

---

[2] This "super partes" company daily picks up information about TV audience. The Auditel survey classifies the Italian population in several socio-demographic panels according to age, gender, education level, type of job, geographic zone. For each panel, the daily audience data are available, grouped by viewing time and TV channels.

[3] Number of average viewers in every minute of a program.

[4] Percentage of viewers of a program compared to the overall audience in the same time slot.

Housewife

Profile

Age *[personal data]*:	Importance: 1.0 Values: (less_then_15, 0.0) (15/24, 0.0) (25/34, 0.0) (35/44, 0.5) (45/54, 0.5) (55/64, 0.0) (more_then_64, 0.0)
Gender *[personal data]*	Importance: 1.0 Values: (male, 0.0) (female, 1.0)
Books*[interest]*:	Importance: 0.6 Values: (low, 0.8) (medium, 0.2) (high, 0.0)
Family*[interest]*:	Importance: 1.0 Values: (low, 0.0) (medium, 0.2) (high, 0.8)

Prediction part

Movies:	Values: (action movies, 0.1) (soap-romance movies, 0.4) (cult movies, 0.0) (TV movies, 0.3) (comedy, 0.2)

Fig. 1. A portion of the "Housewife" stereotype, a female class of users.

Profile of a stereotype. The personal data and the interests are described as features with two or more corresponding linguistic values. In the slot, a numeric value in the range [0..1] is associated to each linguistic value, in order to describe the compatibility of the linguistic value with the description of the user class represented by the stereotype. For instance, consider the feature *Gender* in the stereotype "Housewife" shown in Figure 1. The 100% of the people belonging to the class are female. Also the *Importance* facet, describing how relevant is the feature to the description of the represented stereotype, takes values in the range [0..1], where "0" means that the feature is irrelevant, while "1" means that it is extremely important.
Prediction part. These slots make predictions on the preferences for TV programs of a typical user belonging to a stereotype, on his watching frequency and on the viewing times in which such preferences may occur. Again, we consider a set of linguistic values for each feature and we associate with each linguistic value a numeric value in [0..1]. This numeric value represents the conditional probability of the linguistic value for the datum, given that the user belongs to the stereotype. For instance, consider the preferences for movies shown in "Housewife". The female users represented by this stereotype likely prefer soap-romance movies and TV movies, they quite like comedy and action movies, while they certainly don't like cult movies.

DISCUSSION AND FUTURE WORK

This paper has described the analysis and definition of a stereotype knowledge base used within a system for the generation of personalized EPGs to predict user preferences on TV programs [2,7]. To define the knowledge base, we used stereotypical descriptions starting from an analysis of a lifestyles survey, assuming a plausible correlation among homogeneous user groups and their preferences.

A preliminary test of our stereotypical approach showed that the system's predictions on viewer preferences were excellent for users fitting the personal data and the interests of a specific stereotype. In contrast, the system failed when the user's interests were different from those evaluated according to his personal data. Basically, stereotypical knowledge does not correctly handle users matching different lifestyles in different aspects of their behavior, because of the major selectivity of the personal data in the classification of users, in spite of interests. Of course, interests could be made more selective than personal data, but sometimes the explicit user's preferences are not reliable, because they are influenced by factors such as misunderstandings and the desire of being considered in a positive way. Thus, the personal data fit the aim of classifying the users in stereotypical descriptions better than the user's explicit interests. In order to cope with the complexity of human behavior and to solve these problems, we structured the User Modeling Component (UMC) of our system into different user modeling modules, which exploit alternative sources of information about users. In addition to the Stereotypical Module, an Explicit Preferences Module manages the user's declared preferences and the UMC is in charge of solving the possible conflicts in the predictions of the two modules [2].We are persuaded that the stereotypes represent useful information during the initial interaction with the user. Moreover, the existence of stereotypical behavior in watching TV is also supported by empirical studies. At the moment, we are testing new solutions starting from information about different types of people, to simulate the real behavior of the system. During this second test phase, we will also concentrate on: *i) refining* the prediction part of the stereotypes by analyzing new audience data; *ii) reclassifying* the user in more suitable stereotypes, after monitoring the user's program selections. The reclassification will only concern the interests; *iii) exploiting* personalization strategies to adapt the content and layout of the TV guide to the user model.

REFERENCES

[1]Ardissono L., Goy A. (2000): Tailoring the Interaction With Users in Web stores. User Modeling and User-Adapted Interaction, 10(4), pp. 251-303, Kluwer Academic Publishers.

[2]Ardissono L., Portis F., Torasso P., Bellifemine F., Chiarotto A., Difino A. (2001): Architecture of a system for the generation of personalized Electronic Program Guides. To appear in Proceedings of the UM2001 Workshop on Personalization in Future TV, Sonthofen, Germany, 2001.

[3]Auditel (2000): http://www.auditel.it

[4]Calvi G. (1986): Indagine sociale italiana. Rapporto Eurisko 1986. FrancoAngeli

[5]Casetti F., Di Chio F. (1998): Analisi della televisione. Strumenti Bompiani.

[6]Eurisko. Sinottica (2000): http://www.eurisko.it

[7]Gena C. (2001): Designing TV Viewer Stereotypes for an Electronic Program Guide. To appear on the Proceedings of the Eight International Conference on User Modeling.

[8]Rich E. (1989): Stereotypes and User Modeling. In A. Kobsa and W. Wahlster, editors, User Models in Dialog Systems, pages 31-51. Springer Verlag, Berlin.

[9]Torasso P., Console L.(1989): Diagnostic Problem Solving. North Oxford Academic.