Providing Context-Sensitive Information to Groups

Berardina De Carolis and Sebastiano Pizzutilo

Intelligent Interfaces, Department of Computer Science
University of Bari, Italy
<decarolis|pizzutilo>@di.uniba.it

Abstract. This paper describes how to provide *background information* adapted to the model of a group of people present in an Active Environment.

1 Introduction

Today most public places are provided with large-screen, digital displays or other output devices typical of the particular environment: examples are cardio-fitness machines in a fitness center, displays of a food dispensers, bus/train/plane notice-boards, etc.. In opposite to on-line information seeking, such displays promote the experience of "encountering" the information while carrying on another activity [1]. We denote with "*background information*" contents and news that are secondary to the main reason or task that led users to that particular environment.

In this paper we propose an approach to group modelling that aims at providing background information adapted to presumed information needs of people present in a communal space and to context features such as the particular activity zone in which information is displayed, the time of the day and so on. The system adopts an approach in computing the profile of the user group that considers the fact that people to which information is addressed may be totally **unknown**, or may be **known** in full or in part, for example if the profiles of all or some of the users may be transferred to the environment. In order to test the system we selected as active environment a Fitness Center[1]. This type of environment is interesting for the main purpose of this research since: i) people subscribe a contract with the center and, contextually, it is possible to ask them to fill a questionnaire about their interests; ii) users are often provided with magnetic badges that allows identifying their entrance in the environment; iii) users are heterogeneous and have different interests, furthermore for certain period of time their presence is quite stable with some turn-over periods; iv) the overall environment can be divided in different activity zones in which it is plausible that people have different information needs (i.e. reception, fitness room, locker rooms, etc.); v)it is possible to make a statistical forecast of how many and which categories of users are present in an activity zone in a given time slice and therefore to combine this information with the profiles accessible by the system.

In the rest of the paper we describe briefly the system architecture and the group modelling strategy and , in the last Section, we discuss results.

1 A.S.D. BodyEnergy, Mola di Bari, Italy.

2 Group Modeling in GAIN

The approach presented in this paper is an evolution of GAIN (Group Adapted Interaction for News) [2]. GAIN is based on a Service Oriented Architecture [3]. The main component of the system is the GAIN Web Application (WA) that is used for displaying news to people attending the target communal space. Adaptation to the group and context is realized by the Group Modeling WS, responsible for computing the presumed preferences and interests of the group, and the RSS News Reader WS, responsible for searching news on the Internet. The user profiles in GAIN are formalized using the situational statement language UserML from UbisWorld [4], since it provides a language that allows representing the user model in a very detailed way by introducing information useful for the contextualization. The RSS News Reader WS allows to search for news using the RSS Feed technology. Each RSS feed follows a standard syntax and consists of a set of news, each with a title, a category, a description (which is an abstract of the news item) and a link to a web page where the news item is located. The list of filtered news is sorted according to the interest scores that the group modeling component calculated for every news category and shown on the display.

2.1. Group Modeling Strategy

MusicFx [5] is a group recommender system employed in a fitness center that is for some aspects very similar to GAIN. MusicFx chooses music according to the preferences of the groups of users present in a fitness center. However, while in MusicFx all users present in the shared space are known by the system, in GAIN we want to combine preferences about ***known*** people with the presumed preferences of the ***unknown*** ones. For this reason we decided to combine information about the statistical distribution of preferences of people that usually attend that place, with those that are eventually known by the environment. In order to collect these statistical data, we conducted a study concerning the people distribution and their interests about news in different activity zones of the Fitness Center. Groups in these places are made up of people that spend there a *limited* period of time (short or long). Group formation is *accidental* however it is implicitly regulated by the type of activity that people performs (i.e. a collective courses, a individual training, and so on).
The study involved 170 subjects (typical representation of the target users of our application during a day. Subjects were requested to fill a questionnaire that was divided in three sections aiming at: i) collecting some *demographic data* about the subjects (gender, age, category); ii) understanding the *frequency of attendance (*at *what time*, for *how long* and *how many days* during the week subjects were attending the place) and in which activity zone subjects are supposed to stay in each time slice according to the habitual activity performed in the place; iii) understanding which were the possible *topics* of interest by asking subjects to score them in a list using a 1-5 Likert scale (from 1 " *I really hate it*" to 5 "*I really like it"*) for each activity zone. From this data we derived some information about the habits of every user in term of average number of minutes spent in each activity zone during a day, and about their distribution in every time slice. Figure 2 shows, in particular, the distribution of subjects' interests when being in the Fitness room in two different time slices: around 10.00

a.m. and 18.00 p.m.. These time slices were selected as being quite different: in the morning the fitness room is attended prevalently from women while in the early evening from young male students.

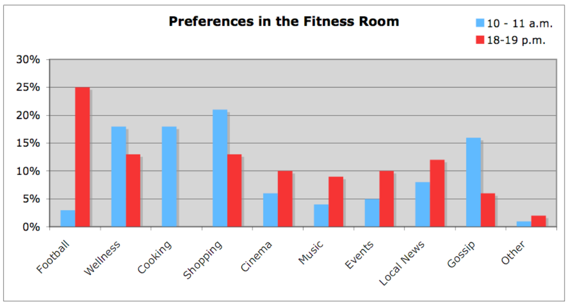

Figure 2. Comparison of interests in two time slices

Then, the definition of the group profile is made according to the formula we propose in (1) where different weights may be assigned to the ***known*** and ***unknown*** groups, according to the relative importance one wants to give to one group with respect to the other. In (1), we denote as:

- $\boldsymbol{P_{SURE}}$, the weight (from 0 to 1) given to the preferences of known group.
- $\boldsymbol{P_{PROBABLE}}$[2] , the weight (from 0 to 1) given to the preferences of the unknown group.
- K, the number of topics;
- UM^{i}_{j} the score for ***topic j*** in the activity zone A from ***user i;***
- b the base of the votes; can be 1, 5, 10, 100, etc..
- N , the number of known users;
- M the number of profiles that constitute the statistical dataset (initially M was equal to the number of profiles collected in the preliminary study);
- f_i , the frequency of the attendance for every user of the selected activity zone, calculated as the number of days attended by ***user i*** divided by number of working days;
- t_i , the average number of minutes in which the ***user i*** is in the activity zone in the considered time slice;
- $F_m = \sum_{m=N+1}^{N+M} (f_m * t_m)$ the frequency in the statistical dataset;

Then, C_j, indicating the confidence value for a ***topic j*** to be shown to the group in the activity zone A, is computed as follows:

$$\forall j \in \{1,..,K\} : C_j = \frac{1}{b}\left[\frac{P_{SURE}}{N}\sum_{i=1}^{N}\left(UM_i^j\right) + P_{PROBABLE}\sum_{i=N+1}^{N+M}\left(UM_i^j * \frac{f_i * t_i}{F_m}\right)\right]^{3} \quad (1)$$

with N>0 and M>0 and b <>0.

This formula is a variation of the Additive strategy [5] in which the weight for the unknown part of the group cannot be uniformly distributed, since people are present in a place with different frequencies. We introduced $\frac{f_i * t_i}{F_m}$ for filtering news according to the fact that some users are more likely to be in a given activity zone at a certain time than others. This frequency is calculated by approximating the presence of users according to questionnaire answers. In particular, we used the data collected in the preliminary survey in order to calculate $\boldsymbol{f_i}$ and $\boldsymbol{t_i}$ as the average number of minutes that a ***user i*** spends in the activity zone, during the week, in the time slice in which the

---

2 We were interested in having the confidence of each topic expressed as a percentage. For this reason we set $\boldsymbol{P_{PROBABLE}} = f(\boldsymbol{P_{SURE}})$, being f a function that relates these two values.

3 This is valid in case PSURE+PPROBABLE=1 otherwise it is necessary to divide the value of Cj for the value of PSURE+PPROBABLE in order to obtain a result between 0 and 1.

group modelling function is activated. Obviously, when N=0 and M>0 P_{SURE} should be set to 0 and $P_{PROBABLE}$ to 1. In the opposite case, when N>0 and M=0, $P_{PROBABLE}$ should be set to 0, and P_{SURE}=1. Once the list of preferences is computed by the group modeling web service, it is used to filter the news by the GAIN web application.

2.1 Updating the group model

In the context in which GAIN is applied, the group model can be updated in different situations: in a **non-interactive** context, in an **"collective" interactive context** using a public touch screen display, in a **personal interactive** context through a personal device. In all cases, we believe it would be impracticable to update the model by asking people to explicitly rate the relevance of background news, especially if they are in that place for a different purpose [6]. Therefore, in all the three cases model updating occurs when **new users** come into the activity zone or when the **next time** slice is reached, according to the statistical distribution. The system re-applies the formula (1) to calculate the new confidence of all news categories. To avoid sudden changes of the topics list, a scanning of known users is done every **n(A)** minutes. This time interval corresponds to the average time that subjects involved in the preliminary survey declared to spend in each activity zone (**A**). In the second situation, the users may interact with the system by simply clicking on the proposed news**.** This is considered as a kind of implicit feedback, since we may assume that users do not click on news at random, but rather on news whose titles are interesting to them [7]. Therefore, the clicked links may be referred as positive samples which match the user preferences from a statistical point of view. However, in our application domain we do not know who is the member of the group that made the selection. For this reason, we created a temporary profile for every time slice and for every activity zone ($\mathbf{(GIP(A))_t}$: **G**roup **I**nteraction **P**rofile for the activity zone **A** in time slice **t**). This profile has the same structure of $UM(A)_i$ and contains a list of the news categories that we have in our domain, but with an initial confidence value equal to 0. Every time a user selects a news belonging to a category **x,** this is denoted as a positive feedback and the relative counter is incremented. At the end of the time slice the confidence of each category is updated by dividing the relative counter by the total number of selected news. For example if N is the number of all the news selected by the users, and we consider Kj as the counter of the selected news for each category, the confidence Cj in the $\mathbf{GIP(A)_t}$ for the category j, will be calculated as Kj/N. The temporary profile $\mathbf{GIP(A)_t}$ is used to update the group preferences for that activity zone **A** in the time slice, since it is added to the statistical dataset and used in the next computation according to the formula (1). In this case the number of profiles M, used to build the statistical dataset, is incremented. This approach enables us to tune the statistical calculation of the group profile, in order to reach a situation that is closer to the real usage of the system. However, a possible problem may regards the number of collected profiles (M) since they enrich the statistical dataset and are never discarded. To solve this problem the idea is to stop with this way of gathering information about the user when we will have a quite large number of usage profiles (around 1000) and to use machine learning techniques to extract information for building stereotypes relative to activity zones and time slices. With this new approach the temporary profiles will be considered new examples for updating stereotypes. The third situation regards the interaction through a personal device. In this case we can identify the user and then we may use feedback to update his/her personal profile as in a single user application.

4 Discussion about results

In order to validate our approach we tested the impact of the adaptation strategy first through a subjective evaluation study and then using a simulation of the system behaviour. This first study involved 80 people that usually attend the Fitness Center. One group of 40 people received a non-adapted random selection of news while for the other group news where selected according to the statistical profile. Both groups received news on a non-interactive display. We performed a t-test on the two datasets and results showed that the group that received news adapted to the statistical profile found in average more interesting (p-value=0,016), appropriate (p-value=0,004) and adequate (p-value=0,01) the proposed news than the group that received non-adapted ones. The second experiment, conducted using a simulation realized with a multiagent system, was aiming at understanding how to tune P_{sure} and $P_{probable}$ values according to the different situation that may occur. Results show that if profiles of known users are statistically different from the statistical one then the list of the news to propose to the group will vary considerably. This becomes even more evident when P_{sure} is high. In this case a plausible criteria could be to set P_{sure} according to the ratio between the number of known and unknown users. If known users are similar to the statistical profile, then the mark does not change very much and, obviously, a high P_{sure} may enforce the position of some categories in the classified results. The P_{sure} value can be used to carry weight the context features Results obtained so far seem to confirm that the mixing statistical information about the group with those about known users allows to handle efficiently news adaptation in active environments.

References

1. Elderez, S. (1997) "Information encountering: a conceptual framework for accidental information discovery". In *Proceedings of an Internation Conference on Information Seeking in Context (ISIC),* Tampere, Finland, 1997, pp. 412-421.
2. Sebastiano Pizzutilo, Berardina De Carolis, Giovanni Cozzolongo and Francesco Ambruoso: A Group Adaptive System in Public Environments, WSEAS TRANSACTION on SYSTEMS. Issue 11, Volume 4., pagg 1883-1890, November 2005.
3. Endrei, M. et al.. 2004. Patterns: Service-oriented Architecture and Web Services. IBM Redbook, ISBN 073845317X.
4. Heckmann D. Ubiquitous user modeling. IOS Press, 2005.
5. McCarthy, J., and Anagnost, T. 1998. MusicFX: An arbiter of group preferences for computer supported collaborative workouts., in Proceedings of the ACM conference on CSCW, Seattle, WA, pp. 363-372.
6. Masthoff, J. 2004. Group Modeling: Selecting a Sequence of Television Items to Suit Group of Viewers, User Modeling and User-Adapted Interaction, v.14 n.1, p.37-85.
7. Adomavicius, G. and Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734-749.
8. Joachims, T., Granka, L., Pan, B., Hembrooke, H., and Gay, G. 2005. Accurately interpreting clickthrough data as implicit feedback. In *Proceedings of SIGIR Conference on Research and Development in information Retrieval.* SIGIR '05. ACM,154-161.