Decentralized Mediation of User Models for a Better Personalization

Shlomo Berkovsky*

University of Haifa, 31905, Haifa, Israel
slavax@cs.haifa.ac.il

Abstract. The growth of available personalization services and the heterogeneity in content and representation of therein exploited User Models (UMs), raise a need for a mechanism allowing to aggregate partial UMs generated by other services. Such a mechanism will allow reuse of partial UMs in multiple personalization services that may need it. This paper discusses the details of a decentralized mediator for cross-domain and cross-technique translation and aggregation of partial UMs. The mediator facilitates enriching UMs managed by personalization services and improving the quality of the provided personalization.

1 Introduction

Providing accurate personalized information services to consumers requires modeling their preferences, interests and needs. This data is referred to in the literature as the User Model (*UM*) [7]. Typically, service providers build and maintain proprietary UMs, tailored to the specific contents offered by the service, and the personalization technique being exploited. Since the quality of the provided personalized service depends largely on the characteristics and richness of the UMs, different services would benefit from enriching their UMs through importing and aggregating partial UMs, i.e., the UMs built locally by other, possibly related, services. This can be achieved through *mediation* of partial UMs.

UM mediation raises a number of issues. The first issue is the commercial nature of the nowadays information world. Due to competition, personalization services usually neither cooperate, nor share their partial UMs. The second issue is customer's privacy. Partial UMs built by service providers may contain customer's private data, which should not be disclosed to untrusted parties [4]. The third and fourth issues are the structural heterogeneity and incompleteness of the UMs contents, since every service refers to a specific application domain only. The lack of standard representation, and specific requirements imposed by different personalization techniques, result in personalization services building their models in different, ad-hoc forms. As a result, large amounts of heterogeneously represented and possibly overlapping (or conflicting) data are scattered among various service providers.

Generation of a central UM, as a composition of partial UMs stored by various personalization services, is discussed in [6]. For this, each service maintains a mechanism

* The author would like to thank his advisors, Tsvi Kuflik and Francesco Ricci, for their valuable comments on this paper.

V. Wade, H. Ashman, and B. Smyth (Eds.): AH 2006, LNCS 4018, pp. 404 – 408, 2006.

capable of accessing the relevant parts of the central UM. To provide personalization, each service extracts the required data from the central UM and later updates the central UM. However, the centrality of the UM poses a severe problem that should be treated. In [8], the authors discuss agent-based sharing and management of partial UMs, which are centrally aggregated into a global UM. However, neither the sharing policy, nor the translation between different representations is defined, such that UMs sharing between the services should be implemented explicitly.

Unlike the above studies, this work aims at handling the aggregation through a *decentralized UM mediator*, capable of aggregating partial UMs. The mediator provides a scalable platform for privacy-enhanced data exchange and facilitates an ad-hoc (i.e., for a specific purpose, and not derived from a general, continuously maintained UM) generation of the UMs for the target service through translation and aggregation of partial UMs built by other services. Thus, the mediator bootstraps empty UMs, or enriches the existing UMs, leveraging the quality of the provided personalization.

2 Mediation of User Models

Principal architecture of the mediator was discussed in [1], whereas this paper elaborates on the mediation process and the ways of applying the mediator in a decentralized distributed environment. The main functionality of the mediator is to facilitate aggregation of partial UMs built by different services. Thus, it provides a common interface for user modeling data exchange. Figure 1 illustrates the mediation process.

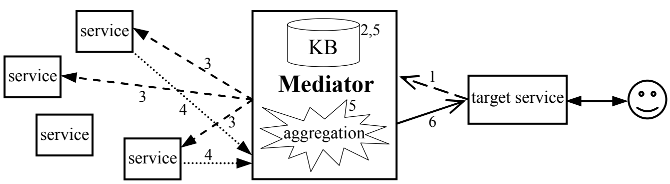

Fig. 1. Architecture and stages of the UM mediation

The mediation process is partitioned to the following stages:

1. A target service, required to provide personalization to a user, queries the mediator for the UM related to the application domain of the provided service.
2. The mediator identifies the required personalization domain and the UMs representation in the target service.
3. The mediator determines a set of other services that can potentially provide partial domain-related UMs of the given user and queries them.
4. Services, actually storing the needed data, answer the query, and send to the mediator their partial UMs of the given user.
5. The mediator translates and aggregates the acquired partial UMs (using the KB) into a single domain-related UM, represented according to the target service.
6. The generated domain-related UM is sent to the target service, which is capable of providing more accurate personalized service.

Two major issues that should be resolved to facilitate proper functionality of the mediator are: (1) "Which services can provide valuable partial UMs?", and (2) "How to translate and aggregate the acquired heterogeneous UMs?", i.e., stages *3* and *5* in figure 1. Thus, in the rest of the paper we focus on these stages.

First, let us analyze the distribution of partial UMs among the services. Nowadays, personalization services are exploited in a wide variety of application domains (e.g., movies, music, tourism, etc...). The contents of the UMs may vary between application in the same domain, and certainly between applications in different domains. Thus, the UM is considered as an aggregation of partial domain-related UMs: $UM=aggr(UM_1,UM_2,\ldots,UM_k)$. Moreover, within a given domain, the services may exploit various personalization techniques (e.g., collaborative, content-based, case-based reasoning, etc) that impose different representations of the UMs. As a result, domain-related UM is considered as an aggregation of partial technique-related UMs: $UM_d=aggr(UM_d^1,UM_d^2,\ldots,UM_d^n)$, where UM_d^t denotes the partial UM referred to application domain *d*, built by a service exploiting personalization technique *t*.

Stage *3* of the mediation aims at determining the set of services that should be queried by the mediator. We assume that three groups of services will provide valuable partial UMs for building UM_d^t for a service from domain *d* exploiting technique *t*: (1) other services from *d* that also exploit *t*, (2) services from *d* that exploit another technique *t'*, and (3) services from another, relatively similar, domain *d'* that also exploit *t*. Although other services, i.e., with different combinations of techniques and domains, can potentially provide valuable partial UMs, we refrain from querying them, since their mediation requires multiple translations, which may 'contaminate' the data.

To alleviate the task of determining and querying the relevant services, we propose to organize the available services in a hierarchical semantically demarcated structure. The upper level of the hierarchy represents different application domains of the services. The domains are represented by the nodes of an undirected graph, where the weights of the edges reflect the similarity between the respective domains. The similarity values allow determining whether partial domain-related UM_j can be valuable for aggregating another domain-related UM_i. The bottom layer represents specific services within the domains, such that the services are grouped according to the personalization techniques they exploit. This organization of services inherently restricts the queries for partial UMs only to the services referring to the same application domain, or to different but relatively similar domains, and exploiting the same personalization technique (or both, i.e., the same domain and technique). Analyzing

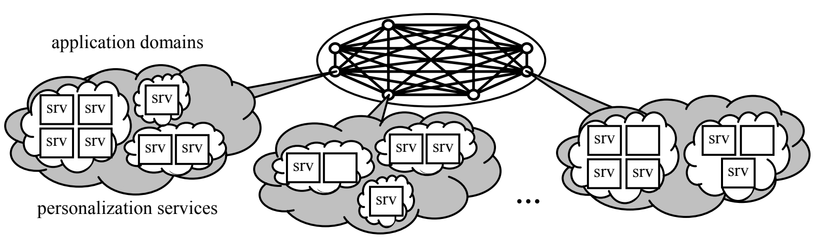

Fig. 2. Organization of services: graph of domains and grouping by the techniques

UM representations in various techniques will allow determining the applicability of partial UM_d^m for aggregating in another UM_d^n, and further restricting querying of services. Figure 2 illustrates the organization of the personalization services.

Another important issue is translating and aggregating the acquired partial UMs (stage *5*). Clearly, different application domains require different information to be stored in the UMs. Even within the same domain, different services may store different information in their partial UM, according to the exploited technique (e.g., ratings vector in collaborative UM vs. a list of interest topics in content-based UM). Moreover, partial UMs from the same domain and technique may use different terms to de-scribe equivalent semantic concepts. Thus, successful translation of partial UMs requires rich inter- and intra-domain *Knowledge Base* (KB) that allows identifying commonalities between partial UMs and inferring the required data. According to the above-mentioned groups of services providing valuable partial UMs, we define *3* types of possible translations: (1) simple concatenation of partial UMs, (2) *cross-technique* translation, and (3) *cross-domain* translation. Clearly, each type of translation requires specific inference mechanism and exploits different data from the KB.

In addition to the main task of UMs translation, stage *5* of the mediation process is also responsible for aggregating partial UMs, i.e., resolving conflicts and inconsistencies. This is not unreasonable that different services will provide partial UMs with different levels of accuracy, relevance and 'freshness'. Although we highlight the importance of resolving these issues, they are beyond the scope of the current work.

Consider an example scenario where content-based movie recommender requests the mediator for a UM of a given user (stage *1*). The mediator identifies the domain and the representation of the UM, and queries other services that can potentially provide valuable partial UMs (stages *2* and *3*). Let us assume that other content-based movie recommenders, movie recommenders exploiting other techniques, and TV, books and music recommenders are queried. Services, storing the user's UMs, answer the query and send their partial UMs (stage *4*). The mediator exploits various KBs to translate and aggregate the acquired partial UMs into a single movie-related content-based UM (stage *5*). For example, cross-technique translation from collaborative to content-based movies UM exploits a KB of movies data (e.g., genres, directors and actors), which allows the mediator to generalize a set of collaborative ratings into the content-based UM, containing a list of genres, directors and actors liked/disliked by the user. Conversely, cross-domain translation from books to movies content-based UMs exploits a KB of books and movies genres that facilitates the translation through identifying the correlations between the contents of the UMs (e.g., liked/disliked genres, common to movies and books). Then, the aggregated UM is sent to the movies recommender, which provides the user more accurate personalization (stage *6*).

Finally, we would like to highlight three hypothesized advantages of the proposed mediation mechanism: (1) the designed one – better personalization provided by the target service, (2) scalability and robustness – achieved due to the lack of a centralized user modeling mechanism and UM representation, (3) data encapsulation and privacy – achieved due to an independent management of domain-related data by the services, and direct communication between them, where attacking a single service will expose partial UMs only. Nonetheless, we should raise the main disadvantages: necessity of strong inference mechanisms and rich inter- and intra-domain KBs.

3 Preliminary Results and Future Research

Preliminary evaluations of cross-domain mediation in collaborative movie recommender were reported in [2]. There, the datasets from different (but similar) domains were simulated by splitting movies UMs according to the genres of the movies. The accuracy of the recommendations over the aggregated UMs was similar to the accuracy of centralized collaborative recommendations, providing an initial validation to the feasibility of cross-domain UMs mediation. In the future, we plan to investigate the applicability of cross-domain mediation in less similar application domains.

Cross-technique mediation between collaborative and content-based movie recommenders was discussed and evaluated in [3]. To achieve this, we exploited IMDB database (*www.imdb.com*) for extracting the features of the movies (e.g., genres, actors, directors, etc) and building a weighted content-based UM that served as a basis for generating content-based recommendations. Experiments showed that for small UMs, accuracy of the recommendations using the translated content-based UMs is better than of the recommendations using the original collaborative UMs.

Currently, we are also working on UMs mediation between case-based reasoning and content-based tourism personalization systems. For this, we exploit IR techniques for analyzing the contents of the UMs built by Trip@dvice tourism planning system (*tripadvice.itc.it*) and initializing the UMs of museum visitors. In parallel, we exploit similar IR techniques for analyzing textual contents of Web-sites classified in Web-directories for devising distances between different application domains.

In the future, we plan to investigate the possibility of exploiting UM ontologies, e.g., GUMO [5], for the purposes of using parts of generic UM representations in the mediation process. Finally, we will extensively evaluate the proposed approach through combining cross-technique and cross-domain mediations, and will deduce the conditions, where UM mediation improves the quality of provided personalization.

References

[1] S.Berkovsky, "*Ubiquitous User Modeling in Recommender Systems*", in proc. of the UM Conference, 2005.

[2] S.Berkovsky, P.Busetta, Y.Eytani, T.Kuflik, F.Ricci, "*Collaborative Filtering over Distributed Environment*", in proc. of the DASUM Workshop, 2005.

[3] S.Berkovsky, T.Kuflik, F.Ricci, "*Cross-Technique Mediation of User Models*", in proc. of the AH Conference, 2006.

[4] L.F.Cranor, J.Reagle, M.S.Ackerman, "*Beyond Concern: Understanding Net Users' Attitudes about Online Privacy*", Technical Report, AT&T Labs-Research, 1999.

[5] D.Heckmann, T.Schwartz, B.Brandherm, M.Schmitz, M.von Wilamowitz-Moellendorff, "*GUMO – the General User Model Ontology*", in proc. of the UM Conference, 2005.

[6] J.Kay, B.Kummerfeld, P.Lauder, "*Managing Private User Models and Shared Personas*", in proc. of the UbiUM Workshop, 2003.

[7] A.Kobsa, "*Generic User Modeling Systems*", in User Modeling and User-Adapted Interaction, vol.11(1-2), pp.49-63, 2001.

[8] A.Lorenz, "*A Specification for Agent-Based Distributed User Modeling in Ubiquitous Computing*", in proc. of the DASUM Workshop, 2005.