View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by University of Hildesheim

User and Usage Profiling in a Multi-platform Service Environment

Armen Aghasaryan, Stéphane Betgé-Brezetz
Alcatel Research & Innovation, Marcoussis, France
armen.aghasaryan@alcatel.fr, stephane.betge-brezetz@alcatel.fr

Guillaume Raschia, Marc Gelgon
LINA Atlas-GRIM team, University of Nantes, France
guillaume.raschia@univ-nantes.fr, marc.gelgon@univ-nantes.fr

Abstract

In this paper, we present a beginning work on a behavioural profiling approach within a multi-service environment where the service usage and content consumptions data are collected on different service delivery platforms and/or on the user terminals. Therefore, it is necessary to be able to aggregate these potentially heterogeneous data so that a faithful user profile can be induced. The information structure used for profile data representation must be designed in a way to allow its utilization by a number of real-time multi-media personalized applications.

1 Introduction

The explosion of services offered by the telecommunications operators (voice services, video on demand, IPTV, etc.) and the diversity of access environments (terminals and networks) make more and more critical the capability of the operators to measure and to analyze in details the usage of the provided services by end-users. Indeed, the usage measurement allows the operators to be well informed on their best revenue generating services that ensure the success of their offer today and tomorrow.

In addition, a per user analysis aimed at the discovery of user's interest centres will also create the possibility for operators to propose more attractive applications adapted to the profile of each user. A good example is the targeted advertisement, a publicity selected for a particular user in accordance with his interest centers. While business analysts observe a certain saturation in the market of classical broadcast TV advertisement, the targeted ad is seen as a serious revenue relay for advertisers, see [Meyer, 2005].

A range of personalized applications have been successfully industrialized in the domain of Web technologies and e-commerce, e.g. Google AdWords, AdSense, or Amazon recommender system [Linden et al, 2003]. We note that these are "monolithic" solutions where personalization application and profile learning approach are very closely linked, or in other words, the profiling technology is dedicated to a particular personalization application.

In this context, the objective of our research project is twofold. First, we aim at extending the domain of profiling and personalization to the environments of next-generation network operators where multiple content and communications services co-exist. Second, we advocate a a generalized profiling approach where personalization and the profile learning techniques are clearly decoupled.

This paper is structured as follows. The section 2 presents the concept of a multi-platform / multi-application profiling module within a heterogeneous environment of modern large operators. In section 3, we describe the high level architectural approach. Finally, in section 4, both the user and content description as well as the scope of possible algorithmic approaches are presented.

2 Profiling in a multi-service environment

Telecom operators are now proposing content services (TV, music, …) in addition to their traditional communication services. Next stage will be the personalization of these services in order to better fit customer expectations and to get promising associated revenues. Various personalized services can be envisaged: targeted ad, personalized interactivity, personalized search, social networking applications, etc. Figure 1 illustrates such personalized applications in the context of IPTV (similar services can also be proposed in a mobile environment). At the initial step (a), the customers are watching a broadcasted TV program (all the viewers are watching the same program). Then, at the Ad time, each customer receives a targeted Ad depending on his profile: Alice - *who likes buying clothes!* - receives an Ad on dresses (b), John - *who likes cars!* - receives an Ad on cars (c). Furthermore, as shown in (b) and (c), interactivity menus can also depend on profiles.

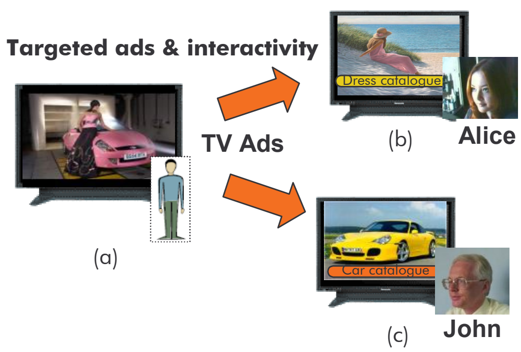

Figure 1 : Example of service personalization: targeted Ad and personalized interactivity on IPTV.

In order to propose such personalized services, operators need to have a reliable user profile, faithful to the actual user interests and expectations. This is the purpose of the multi-platform / multi-application user and usage profiling module that builds and updates the user profile taking into account the usage data on different services.

The principle of this module is shown in Figure 2. On one hand, it must integrate usage information of different Service Delivery Platforms (SDP): IPTV SDP for TV/video on fixed networks (e.g. xDSL), mobile video SDP for TV/video on mobile networks (e.g. either streamed on 3G or broadcasted on DVB-H networks) and IMS SDP (IP Multimedia Sub-system) for the next generation of communication networks. On the other hand, this generic profiling module should offer semantic user profile information allowing to address the various types of applications.

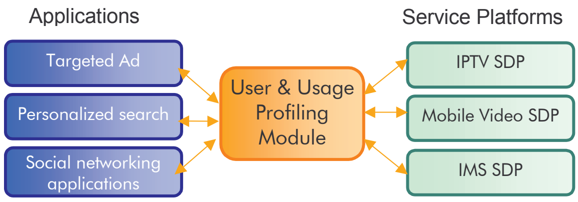

Figure 2. A multi-platform / multi-application profiling module.

In the sequel, we present the architecture of this module and the algorithmic approaches that can be used to aggregate the usage information into a consistent user profile.

3 Architectural approach for generic user profiling

3.1 Functional blocks

The overall architecture of the profiling module is depicted in Figure 3. As it was mentioned earlier, we advocate an approach where the personalization and the profile learning are clearly dissociated.

At first, we assume that multiple sources of raw data are used : server data (e.g. CDR – Call Data Records) coming from the different SDP (IPTV, mobile video and IMS), or data extracted from network devices, terminals or Set-Top-Boxes (e.g. cookies). Example of such raw data can be: the trace of a user session who has viewed *during 10mn a sport video on his TV*, or the trace of channel switching extracted from the DLSAM equipment.

The main components of the proposed architecture are the following (c.f. Figure 3):

- **User Profile Data-Base** that stores user profile information. In the scope of this paper, we are concerned with semantic user data such as the user interest domains (sports, entertainment, etc.), and his service habits (high/low user of video, service habits at home or on the move, etc.).
- **Explicit Profiling** allowing for the end-user to enter directly his user profile information (through a web site or a questionnaire). Explicit profiling is necessary, however – even if the end-user enters fairly these information it is not sure that it corresponds to the actual profile of the user. Therefore, having also an implicit profiling is mandatory.
- **Implicit Profiling** allowing to gather all the raw data, to learn and to update the user profile with respect to the real usage of services. Combination of such explicit and implicit profiling will allow the operator to have the most faithful and up to date profile.
- **Query Interface** used by the applications to offer customized services. Although the specific logic of personalization will be realized within each particular application, some frequent request patterns can be provided by the profiling module via its intelligent query interface. So, depending on the applications, various requests should be envisaged, e.g.: get all the users having certain profile criteria (e.g. interest in tennis, high video user), get the users having a profile "similar" to another one. This query interface (and the offered API) is critical as it will ensure the independence of the profiling module with respect to different personalized applications.

Note that all the SDPs as well as the profiling module are supposed to be owned by a single large operator. In the case where multiple operators are involved a profile sharing mechanism should be introduced, but this is out of the scope of this paper.

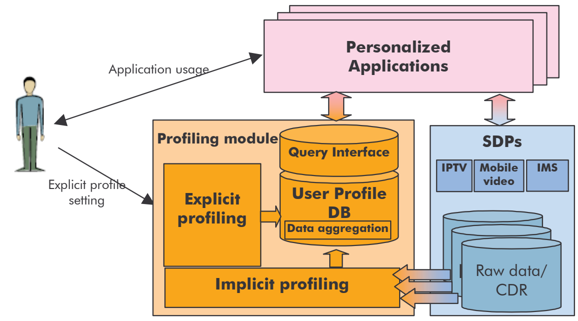

Figure 3. Main functional blocks for user profiling and personalization of services.

3.2 Non-Functional requirements

While defining a system of service usage measurement, an analysis of operational constraints of security, confidentiality, privacy, and scalability shall be taken into account especially for real-time multi-media applications. For example, the following aspects can be mentioned:

- limitations in terms of confidentiality guarantees, or the rights of the operator on the usage data of the customers. The related privacy legal constraints (national, European or international) have therefore to be guaranteed,
- capacity to support the profile evolution,
- performance issues related to temporal and spatial complexity and to the profile construction algorithms,
- possibility of evolution towards a distributed architecture for data acquisition, profile learning and service personalization.

4 Algorithmic profiling approaches

4.1 User profile and content description

The ETSI work on user profile management (ETSI STF265 EG 202 325 Document, oct. 2005) enumerates the very many aspects of user profile management entertaining to personal devices. As an example, it raises issues specific to the distributed infrastructure we are aiming at, such as the conflict of interest on how to disclose as little as possible from both the user profile and the content potentially delivered. Although user profile learning from their actions is out of scope of this ETSI standard, this issue can relate technically to how we model the user profile describing the user's interests.

Standardization has since long addressed the description of audiovisual content (programs and commercials)

[Evain *et al*, 2003.], including features such as 'genre' and 'channel' which are most amenable to learning user interests. It now also includes an MPEG-7 description form for users' interests and the history of user viewing actions, designed with similar goal as we focus on.

Descriptions conforming to this standard have been used in [Thawani *et al*., 2004] to suggest ads suited to a user profile. The matching criterion remains however limited to keywords, whereas the proposal involves a more powerful generalization/induction mechanism.

Hence, in our proposal, the user profile model is defined over the semantic compression technique proposed in [Saint-Paul *et al*., 2005]. Such an approach is somehow similar to that of [Pigeau, *et al*, 2003] applied to TV recommender systems.

4.2 Profiling algorithms

Specific requirements of the multi-platform service environment are addressed to propose a novel and well-suited approach to user profiling.

The main idea is to compute a classification tree of content services from their metadata and then, to mark subtrees corresponding to consumed services by a given user. Both tasks perform online and profiles are incrementally maintained. A conceptual clustering algorithm build concepts (or summaries) that represent classes of services thanks to their metadata. Profiles are then defined as a ranked collection of summaries associated to classes of services. Ranks are calculated as aggregated preferences taken from usage information. And the more the summary is high in the tree, the more it provides general information about the user profile.

Furthermore, using a fuzzy-set based compression technique, our approach relies on background knowledge built from an user-defined vocabulary with a high semantic level. Indeed, the fuzzy set theory provides a symbolic/numerical interface that leads to a numerical computation and a human-friendly interpretation of data structures. Each node from the classification tree is described by a set of linguistic labels on metadata attributes. Then it offers a straight way to query services from personalized applications. Such a querying mechanism has been defined in [Voglozin *et al*., 2006]. In the same time, this summary model allows the end-user to provide by hand an explicit profile with its own vocabulary. The system incorporates explicit summaries into the (implicit) list and resolves inconsistency following a given policy such as "explicit first" or "implicit first" or "middle-term".

Finally, such a system is expected to fulfil both the functional and non-functional requirements as stated in section 3: a content-based profiling approach thanks to services metadata and usage information, a joint explicit and implicit profile, a personalized access to services provided by a profile-oriented filtering mechanism, an online update of profiles, a linear temporal complexity of the algorithm w.r.t. metadata as well as a controlled memory footprint, and last, the possibility to deploy the solution onto distributed architectures (see [Saint-Paul *et al*., 2005] for details about complexity and distribution).

Technical and legal precautions should be considered to conform to privacy constraints, even if we claim that they do not involve scientific problems. However, privacy issues arising from building user profiles of TV viewers have recently been surveyed in [Spangler *et al*., 2006], especially when data mining is employed to analyse these profiles for marketing purposes.

5 Conclusion

In this paper, we have presented the architecture of a multi-platform/multi-application profiling module integrating heterogeneous service usage information and enabling various personalized applications (targeted ad, personalized search, social-based networking applications, etc.). Telecom operators should take benefit of this profiling module to offer these next generation of revenue generating applications.

However, several challenges have yet to be addressed before the deployment of this system in an operational environment. There are technical challenges, e.g. challenges related to the complexity and the exactitude of the profiling algorithms, their capacity to reflect faithfully the real interest centers of an user, or the possibilities of distributing these algorithms closer to the end-user for scalability purposes. But, other challenges are related to the human acceptance of such profiling and its intrusiveness in the private life of the customers. Even if privacy constraints are guaranteed, this dimension has to be taken into account in the very early design of both the profiling module and of the personalized applications. Nevertheless, one can observe that – even if all the users may not accept a deeper profiling - today they are more and more open to provide personal information on the web. Moreover, operators should also propose incentives for the users to be profiled (e.g. service discount). So, the latter should be also interested by more personalized services in return of their authorization to be profiled.

References

[Evain *et al*, 2003.] TV-Anytime Phase 1, J-P. Evain and H.Murret-Labarthe, EBU Technical Review, July 2003.

[Linden *et al*., 2003] G. Linden, B. Smith, and J. York, Amazon.com Recommendations, Item-to-Item Collaborative Filtering, Industry Report Amazon.com, 2003.

[Pigeau, *et al*., 2003] A. Pigeau, G. Raschia, M. Gelgon, N. Mouaddib, R. Saint-Paul, A Fuzzy Linguistic Summarization Technique for TV Recommender Systems, The IEEE International Conf. on Fuzzy Systems, 2003.

[Meyer, 2005] L. Meyer, Television 2015 - The future of TV financing in Europe, IDATE, DigiWorld Analysis, 2005.

[Saint-Paul *et al*., 2005] R. Saint-Paul, G. Raschia, N. Mouaddib, General Purpose Database Summarization, in VLDB'05, August 30-September 2 2005.

[Spangler *et al*., 2006] W. Spangler, K. Hartzel, M. Gal-Or, Exploring The Privacy Implications Of Addressable Advertising And Viewer Profiling, Communications of the ACM, May 2006.

[Thawani *et al*., 2004] A. Thawani, S. Gopalan, V. Sridhar, Context Aware Personalized Ad Insertion in an Interactive TV Environment, TV'04, Eindhoven, 2004.

[Voglozin *et al*., 2006] W.A. Voglozin, G. Raschia, L. Ughetto and N. Mouaddib. Querying a Summary of Database, Journal of Intelligent Information Systems (JIIS), Volume 26(1):59-73, January 2006.