Profiling Attitudes for Personalized Information Provision *

Yannis Ioannidis, Maria Vayanou
Thodoris Georgiou, Katerina Iatropoulou, Manos Karvounis, Vivi Katifori, Marialena Kyriakidi
Natalia Manola, Alexandros Mouzakidis, Lefteris Stamatogiannakis, Mei Li Triantafyllidi

MaD*g*IK Lab, Dept. of Informatics & Telecom, University of Athens, Hellas (Greece)
yannis@di.uoa.gr — http://www.madgik.di.uoa.gr/

Abstract

PAROS is a generic system under design whose goal is to offer personalization, recommendation, and other adaptation services to information providing systems. In its heart lies a rich user model able to capture several diverse aspects of user behavior, interests, preferences, and other attitudes. The user model is instantiated with profiles of users, which are obtained by analyzing and appropriately interpreting potentially arbitrary pieces of user-relevant information coming from diverse sources. These profiles are maintained by the system, updated incrementally as additional data on users becomes available, and used by a variety of information systems to adapt the functionality to the users' characteristics.

1 Introduction

The quality of human interactions is often intimately affected by the level of understanding that the persons involved have for each other. To a large extent, this holds for interactions between persons and information-providing systems as well. This paper outlines the philosophy and general approach of PAROS, a system under design and initial development at the Univ. of Athens, whose goal is to obtain such an understanding for its users by maintaining profiles of their attitudes[1] so that it may offer them information and other services in a personalized and adaptive fashion. The system aims to provide rich relevant functionality at a generic level, on top of which, particular applications may be built. The key contribution of our work around PAROS is fourfold:

(a) User-model independence: Currently, personalization and other user-based adaptations are mostly offered in ways that are tightly coupled with the main information system or application concerned. They are based on custom user models and profiles that are intimately dependent on the specific application domain, and profiling is based on user interactions with that same system alone. In reality, however, users typically interact with a variety of systems and applications, and their behavior within each environment is useful not only locally but for all other environments as well. PAROS moves away from the conventional restrictive approach and decouples the user model and its associated profiles from its environment-dependent functionality, i.e., user-data analysis and application adaptation. Hence, one's profile may be synthesized from observations coming

Bulletin of the IEEE Computer Society Technical Committee on Data Engineering

*Partially supported by the eContent+ and FP7 programs of the EC through the TELplus, Driver, OpenAIRE, and CHESS projects.
[1] Attitudes may be emotional preferences, behavioral tendencies, or cognitive evaluations (beliefs).

from several applications and be used to adapt the behavior of several other applications, achieving the desired independence to a large degree.

(b) User-model expressive power: As a consequence of their application dependence, current user models often deal with a single type of user preferences or attitudes (e.g., just behavioral tendencies, such as like to watch or like to read) for just a single class of entities (e.g., just movies or just books). PAROS supports a labeled-graph-based user model that can express, in principle, any kind of attitude towards any entity managed by the applications being personalized. Furthermore, the users, the world of relevant entities and relationships, and most importantly, the various kinds of attitude and their intensity are modeled in a uniform fashion, allowing for the effective expression of interactions and influences between them.

(c) Profile-derivation formulation: Derivation (prediction) of unknown user attitudes is a major part of most related efforts' investigation and is usually approached in seemingly ad hoc ways. Based on the graph form of its user model and the resulting profiles, PAROS formulates most types of profile derivation as path computations on these graphs. Having this common foundation, it is able to capture a wide variety of approaches by simply choosing different functions for two function-valued parameters that are the essential elements of the definition of path computations. Thus, the effectiveness of each approach depends on the cognitive and psychological plausibility of these two functions, while its efficiency depends on the algebraic properties of these functions and the performance of the suite of available path computation algorithms.

(d) Application environment novelty: The flexibility and richness afforded by the design of PAROS have inspired relevant work in several novel and challenging environments, whether with respect to profiling or to application adaptation or both. Traditional web logs recording user actions, large discussion threads, multimedia file streams, and user traces from a variety of sensors (e.g., geo locations from smart phones) are some of the sources we are dealing with for user profiling in different contexts. Recommendations in digital libraries, personalized storytelling during museum visits, and opinion summarization of controversial discussions are some of the application environments we are working on.

Each of the above aspects of PAROS presents significant new research (and engineering) challenges, which are part of our current and future work. In addition, PAROS faces extensive syntactic and semantic diversity of data that are part of user profiles, serve as input to the profiling processes, or are otherwise managed by the applications. Hence, the old and ever-present issue of heterogeneous data integration and harmonization arises in all PAROS components as well, but it is not our focus in the context of this work. The remaining sections of the paper are in more or less one-to-one correspondence with the four contribution areas above, identifying the relevant challenges and highlighting the corresponding directions of our work.

2 System Architecture and User-Model Independence

Figure 1 presents the key PAROS components and their interactions. ***User Model and Profile Management*** serves as the foundation on top of which all other parts are built. It provides the basic functionality required by the PAROS user model, which captures several levels of user attitudes towards the relevant world entities and concepts. Following the database tradition, such functionality may be partitioned into *profile definition* and *profile instantiation*. The former is reminiscent of database schema definition and consists of defining types of attitude to be captured in a particular environment, functions that are available for use in profile derivation, and classes of users, relevant world entities, and their relationships. The latter is reminiscent of database manipulation and consists of storing and maintaining the actual profiles, expressing particular users having particular attitudes towards particular world entities. ***Pattern Extraction*** includes techniques that analyze user-behavior data, compute statistical measures and correlations, and extract usage patterns. Such data may be collected explicitly, by having users provide related input directly, or implicitly, by having the system monitor and document users' behavior in different contexts. ***Profiling*** encompasses mechanisms for creating user profiles based on the patterns identified. The latter are appropriately interpreted within the framework of the user model to obtain a first set of profile elements, which may then be recursively processed further to expand profiles with additional elements. Futhermore, as additional user data becomes available (whether online or in batch mode) and usage patterns are updated, the corresponding user profiles are modified accordingly to capture short-term mood swings or long-term attitude changes. To be effective, profile interpretation, expansion, and evolution should all be done according to cognitive and psychological theories. ***Adaptation*** represents the actual selection of the appropriate profiles and their use to adapt the behavior of and the content provided by information systems and other applications to the users' preferences and interests. In principle, a separate component is required for each external system customized, although PAROS may take advantage of certain commonalities.

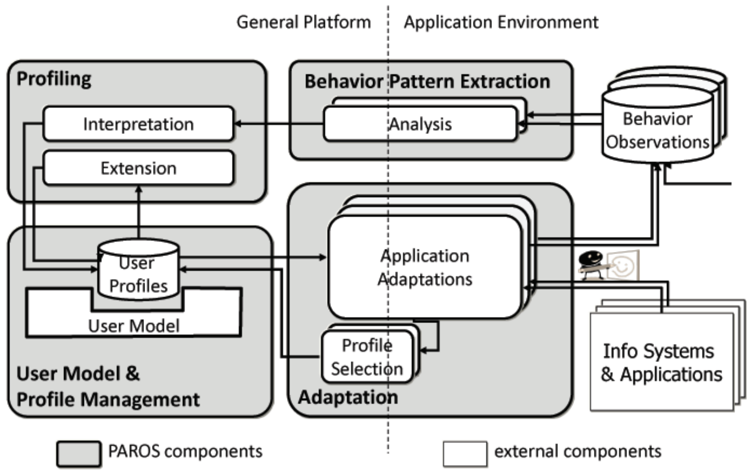

Figure 1: PAROS Architecture

Separation of the four main conceptual components of PAROS is the critical element that leads to the desirable user model independence. User profile elements may come from multiple sources (from applications customized by PAROS or possibly from completely external ones), based on different interaction modes (direct user input or indirect monitoring of user behavior), and after different statistical analyses and other processing based on different techniques. Nevertheless, they are all captured in a unified profile that expresses all aspects of user attitudes. This one profile can then be used to adapt any relevant application, independent of the latter's role in shaping up that profile.

3 User Model

The basis of any personalization or recommendation system is its user model, which must be able to represent in sufficient detail the objects, concepts, and relationships on which the system would offer customized services. The PAROS user model attempts to represent in a uniform way the entire universe of users and their attitudes , including i) individual users, groups of users, communities, and their relations, ii) entities that are the objects of users' attitudes and their relationships to other entities, categories, or concepts, and iii) user attitudes towards these objects. For the sake of readability under space constraints, in this section, we present a somewhat simplified version of the PAROS user model, which is a graph GUM(UOF,DAA), with UOF as its set of nodes and DAA as its set of edges. Nodes in set UOF primarily represent classes of (if any) and individual ***Users*** of the system, ***Objects*** of the world represented in the system being personalized, and ***Functions*** between the above (also captured in edge set DAA). Nodes may also represent particular user groups or communities, object sets, function families, or their combination, if such aggregations play particular roles in the environment concerned.

Edges in set DAA represent two main kinds of relationships among the graph nodes. The first kind is ***Data*** relationships, such as those typically found in databases or other data collections, capturing object structure, association, membership, and inheritance. The second and most interesting kind is ***Attitudinal*** relationships, whose corresponding edges emanate from user nodes and are incident to any kind of node and capture user preferences and beliefs or user actions. Examples of the former may target users (e.g., *trusts* a friend, *is inspired by* a celebrity, *follows* an authority), world objects (e.g., *likes* a concrete object, *supports* a cause, *believes* an ideal), or functions (e.g., *values* a function, in the sense that an attitude towards one of its ends affects the attitude towards the other). Examples of the latter include actions towards users (e.g., *tag* a friend, *respond to* a forum member), actual world objects (e.g., *buy* a book, *watch* a movie), or virtual objects (e.g., *query* a term, *visit* a page, *send* a comment). Figure 2 gives a simple example of a profile illustrating various kinds of relationships (for simplicity, no class nodes or action edges are shown). Note that functions have a dual representation in GUM, being both edges and nodes, a flexibility that facilitates the expression of certain attitudes.

Each attitudinal edge in DAA (and maybe other edges as well) is potentially associated with a label vector, whose structure depends on the edge type. An example of a potential singleton label vector is *trusts*[*ti*](u,v), where for users u and v, trust intensity $ti \in [-1,1]$ captures the level of trust that u has for v, where 1 indicates complete trust, -1 indicates complete distrust, and 0 means indifference. Similarly for other attitudinal edges, such as *likes*[*li*](u,o), capturing the level of likeness that user u has for object o, and *values*[*vi*](u,f), capturing the level of importance that user u places in function f for calculating *li* of *likes*(u,o) based on the *li* of *likes*(u,o), for any o, o' such that f(o)=o'. See Figure 2 for such examples. Possible enhancements of all of these label vectors include at least three other elements: [*co*, *so*, *tmp*]. Confidence $co \in [0,1]$ captures the confidence level for the value of the corresponding intensity (*ti*, *li*, or *vi*), source *so* indicates where intensity and confidence values have been extracted/derived from, and timestamp *tmp* indicates the time period during which the other elements of the vector are valid. To a large extent, the structure, semantics, and detailed mechanisms of the label vectors determine the complexity of the overall user model. Simpler or more complex label vectors result in correspondingly less or more refined representations of users' attitudes.

The above model is rather general and can capture most models appearing in the literature (even if using different notations and representation languages). This generality allows one to remain in the same framework during experimentation with and comparison of different models, as well as in possible future enrichments.

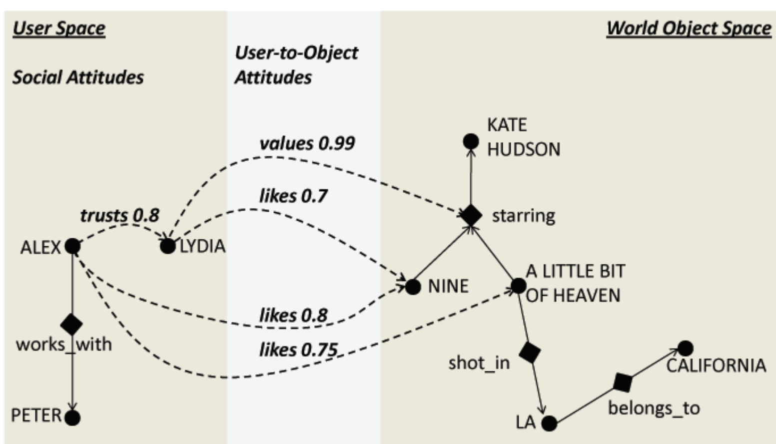

Figure 2: Profile example in the GUM graph user model

4 Profile Derivation Formulation

Profile derivation can be distinguished into first-level derivation (*profile interpretation*, from action edges to attitudinal edges) and second-level derivation (*profile expansion*, from existing attitudinal edges to other unknown attitudinal edges). As a simple example of interpretation, consider two nodes, user Alex and actor Kate Hudson, and an action edge between them, ***queries***[37](Alex,Kate Hudson), indicating that Alex has queried about

Kate Hudson 37 times. The process of interpretation (of attitudes from actions) could derive an attitudinal edge ***likes***[0.85](Alex,Kate Hudson) between the same nodes, predicting the intensity of Alex's attraction to the said actress based on his behavior related to her. Why 37 queries imply an 0.85 level of attraction (both numbers chosen arbitrarily here, for the sake of the example) is a matter of cognitive and psychological theory, which should be codified in the derivation process in some fashion.

Likewise, as an example of expansion, consider the following sequence of attitudinal and data edges and their end nodes: ***trusts***[0.9](Alex,Lydia), ***likes***[0.7](Lydia, Nine), ***values***[0.99](Lydia,starring), starring(Nine,Kate Hudson). Informally and by ignoring some sticky points, the last two edges, ***values***[0.99](Lydia,starring), starring(Nine,Kate Hudson), could be merged into one: ***values***[0.99](Nine,Kate Hudson). Together with the first two edges above (***trusts*** and ***likes***), this forms a path through nodes Alex, Lydia, Nine, and Kate Hudson. The process of expansion (of attitudes) could essentially derive a (probably missing) attitudinal edge, ***likes***[0.62](Alex,Kate Hudson), between the two end nodes, Alex and Kate Hudson, predicting the intensity of the former's attraction to the latter based on indirect information. Again, the methods that generate 0.62 as the attraction intensity should be scientifically justified and be realized in appropriate system components.

Whether as interpetation or as expansion, profile derivation in PAROS is abstracted as an optimal path computation problem, similar to the shortest path, most reliable path, or bill of materials problems. All these are defined on labeled directed graphs and essentially compute a property for an (existing or non-existing) edge (a,b) based on corresponding properties of all the paths between nodes a and b in the graph. Most path computation studies use a path algebra formalism [1], which associates a label with each edge and each path in the graph. A path label is computed as a function, called *CON* (for concatenate), of the sequence of labels of the edges in the path. In addition, a path set P is also associated with a label, which is computed as a function, called *AGG* (for aggregate), of the labels of the paths in P.

Profile derivation does not directly map to conventional path computations, having several additional complications or differences that raise some new theoretical and computational challenges. For example, edges typically have rich label vectors instead of just individual scalar values (e.g., labels could be of the form [intensity, confidence, timestamp, ...]), edges are of different types and affect each other in diverse ways (e.g., different combinations of attitudinal-edge types may be used to derive further edges of these or other types), or the algebraic properties of CON and AGG required for applying the known path computation algorithms do not always hold (e.g., associativity of CON is missing in some cases). Still, modeling derivation as path computation and using that as the foundation of this work is very valuable and serves as a promising starting point in our current and future efforts.

5 Adaptation Applications

In this section, we report on several interesting application environments and information systems we have worked on to personalize their behavior and make their services adaptive. In each case, we point out the most characteristic aspects of our work, whether in the analysis or the adaptation components of PAROS. The great variety of these environments offers a good indication of the applicability of our overall approach.

Query Recommendations in Digital Libraries: In the context of the portal of European National Libraries (TEL [2]), we have implemented a suite of mining techniques to analyze the usage logs maintained by the portal, including techniques for data collection, cleaning, transformation, and pattern extraction. All these form appropriate workflows of data processing operators that are built on top of and orchestrated by **madIS** [3], an open-source extensible data transformation and analysis system that we have developed for this purpose, but is proving a reliable platform in other ongoing efforts as well. In addition to their intrinsic value on how the portal is used, the statistical patterns being extracted by the analysis workflows have also formed the basis for identifying personal, national, and global profiles for its users, based on which, the portal's services have been extended with recommendations on term queries, library collections, and even similar users [4].

Opinion Mapping in Web Discussion Forums: Our work in the area has focused on analyzing online discussions, summarizing the key viewpoints, and profiling the discussants in terms of the side they support. We have used the posts of a discussion thread to construct a Reply Graph, where nodes denote posts and an edge from post A to post B indicates that A quotes B and replies to it. Careful analysis of this graph leads to *discussion topic extraction* and *like-minded discussant identification*. The topic extraction analysis algorithm is based on identifying the appropriate keywords within the scope of each post, which consists of the posts that are closely related to it, whether topologically (proximity in the Reply Graph) or chronologically (proximity of the posts timestamp). The end result is the set of topics of every post, even of small posts with no characteristic keywords in their bodies, and the overall set of topics of each discussion and sub-discussion. The group identification analysis algorithm is based on the observation that most discussion exchanges are expressing disagreement [5, 6, 7]. The end result is the set of discussant groups, where each group has a particular opinion or attitude for the discussion topics and members of each group largely disagree with members of other groups.

Other applications: In addition to ongoing work in the above two environments, we are also involved in some other efforts that take advantage of the PAROS approach but also enrich its range of applications. One such effort investigates the basic characteristics of trust between people and also its derivation from people's interactions in social networking sites. Another one is building on madIS to monitor the life-cycle of scientific publications, form the profiles of ad hoc research communities, and provide personalized search and alerting services to their members. Finally, a recent effort focuses on personalized and adaptive storytelling during museum visits: profiles of visitors are to be derived based on the route they follow while in the museum, on details of their interaction with relevant applications on their mobile phones, and on stereotypical information about them. The goal is for visitors to be guided around the exhibits through a story whose narration is dynamically adapted to their profiles.

6 Conclusions and Future Work

Based on the concept of user model independence, a general graph-based user model, and the use of path computations as the starting point for profile derivation, PAROS aims to become a comprehensive platform for user modeling, user profiling, and profile-based adaptation of information provision and other services. In addition to investigating the limits of the overall approach, the system will serve as a platform for studying the cognitive and psychological validity of profile derivation techniques, the scalability and efficiency of behavior analysis and pattern discovery algorithms, the effectiveness of adaptation strategies, and other relevant research issues. Design and initial implementation of PAROS are currently under way and preliminary experimentation with available parts is quite promising.

References

[1] B. Carre, "Graphs and Networks", Clarendon Press, Oxford, UK, 1979.

[2] http://www.theeuropeanlibrary.org

[3] http://code.google.com/p/madis/

[4] M. Kyriakidi, L. Stamatogiannakis, M. L. Triantafyllidi, M. Vayanou, Y. Ioannidis, "A Prototype Personalization System for the European Library Portal", *European Conf. on Digital Libraries*, 2010.

[5] M. Karvounis, T. Georgiou, Y. Ioannidis, "Utilizing the Quoting System of Online Web Forums to Estimate User Agreement", *ACM SIGMOD Conf., Undergraduate Research Poster Competition*, 2010.

[6] T. Georgiou, M. Karvounis, Y. Ioannidis, "Extracting Topics of Debate between Users on Web Discussion Boards", *ACM SIGMOD Conf., Undergraduate Research Poster Competition*, 2010.

[7] R. Agrawal, S. Rajagopalan, R. Srikant, Y. Xu, "Mining Newsgroups Using Networks Arising from Social Behavior", *Proc. 12th International Conference on World Wide Web*, 2003.