User Profile Modeling and Learning

Evangelia Nidelkou
Informatics and Telematics Institute, Centre for Research and Technology Hellas, Greece

Vasileios Papastathis
Informatics and Telematics Institute, Centre for Research and Technology Hellas, Greece

Maria Papadogiorgaki
Informatics and Telematics Institute, Centre for Research and Technology Hellas, Greece

Ioannis Kompatsiaris
Informatics and Telematics Institute, Centre for Research and Technology Hellas, Greece

Ben Bratu
Motorola Ltd, France

Myriam Ribiere
Motorola Ltd, France

Simon Waddington
Motorola Ltd, UK

INTRODUCTION

A major theme of Information Science and Technology research is the study of personalization. The key issue of personalization is the problem of understanding human behaviour and its simulation by machines, in a sense that machines can treat users as individuals with respect to their distinct personalities, preferences, goals and so forth. The general fields of research in personalization are user modeling and adaptive systems, which can be traced back to the late 70s, with the use of models of agents by Perrault, Allen, and Cohen (1978) and the introduction of stereotypes by Rich (1979). With the wide progress in hardware and telecommunications technologies that has led to a vast increase in the services, volume and multimodality (text and multimedia) of content, in the last decade, the need for personalization systems is critical, in order to enable both consumers to manage the volume and complexity of available information and vendors to be competitive in the market.

BACKGROUND

The goal of personalization is to endow software systems with the capability to change (adapt) aspects of their functionality, appearance or both at runtime to the particularities of users to better suit their needs. The recent rapid advances in storage and communication technologies stress the need for personalization. This need is more evident in consumer-oriented fields, like news content personalization systems, recommendation systems, user interfaces, and applications like home audiovisual material collection and organization, search engines in multimedia browsing and retrieval systems, providing services for personalized presentation of interactive video content. Among these applications, some are Web-based, but there are also versions for PDAs and mobile devices (Tuoriniemi & Parkkinen, 2007) and mobile devices.

In this article, current approaches of user modeling and user profile representation are discussed, and then the focus is on methods for automatic learning of user models and profiles. The presented learning approaches cover a wide range of machine learning (vector-based or probabilistic) methods and also extend to support the most recent advances in personalization systems such as collaborative filtering, ontology-based user modeling and user social context.

OVERVIEW OF LEARNING AND ADAPTATION METHODS IN PERSONALIZATION SYSTEMS

User Modeling–User Profile Representation

User modeling describes the process of creating a set of system assumptions about all aspects of the user, which are relevant to the adaptation of the current user interactions. This can include user goals, interests, level of expertise, abilities and preferences. The most reliable method of user modeling is by explicit entry of information by the user. In most practical systems, this is too time-consuming and complex for the user. Hence implicit user modeling, based on analysis of past and current user interactions, is critical. The user profile is a machine-processable description of the user model.

The information included in user profiles can be divided into a number of categories such as user demographic information, semantic interests, context and location information, and privacy and user interface preferences (Heckmann & Krueger, 2003). Semantic preferences reflect user preferences for particular content topics. User interests and semantic user preferences are the most important source of information widely used in the personalization systems. More specifically, user interests are distinguished between *short-term* that are determined by a particular user interaction or current context, and *long-term* interests which are determined by the user behaviour and preferences over a longer period of time. User interests can also be classified into *gradual* (as a result of user experience), *abrupt* (as a result of an external stimulus) or *repetitive*. Loeb (1992) mentions two types of repetitive changes, *repetitive but predictable* (according to time of day) and *repetitive but unpredictable* (according to user mood).

There is a variety of structures and paradigms that have been used in the academic literature and in commercial personalization systems for the representation of the knowledge and information concerning the user, including the ones listed below. *Attribute-value pairs* are a fundamental data representation in many computing systems and applications. The advantage of such a structure is that it is an open-ended data structure, thus allowing for future extension without any need for modification. In such situations, all or part of the data model may be expressed as a collection of tuples (attribute name, value), where each element is an attribute-value pair. Several attempts have been put forward to standardize this type of user information structure, such as the IEEE Personal and Private Information (PAPI) (PAPI, 2002) and IMS Learner Information Package (LIP) (IMS, 2001).

The *vector space model* (VSM) is an algebraic model used for information filtering, information retrieval, indexing and relevancy rankings. It resembles the attribute-value pairs, but it has a more mathematical structure, in the sense that each element (term, or generally attribute) has a corresponding value or weight representing it and the vector has length and direction, both used, for example, in a similarity metric. The space of all vectors is often called vector domain or domain model. It has been extensively used in documents retrieval and indexing (Salton, Wong, & Yang, 1975). This representation approach has also been followed in a variety of personalization systems (Billsus & Pazzani, 2000; Lawrence, Almasi, Kotlyar, Viveros, & Duri, 2001; Ricci, Arslan, Mirzadeh, & Venturini, 2002).

One of the earlier representation approaches in user modeling has been the use of *stereotypes*. Stereotyping consists of creating a set of prototypical user profiles that represent the features of classes of similar users (Rich, 1979). Instead of keeping an individual model for each user, users are classified into the stereotypical description that best matches their individual characteristics, from which they inherit additional properties and rules.

The need to automatically learn user profiles has given rise to the use of more complicated representation methods such as the *classifier-based models*. These are based on decision trees, neural networks, inducted rules and Bayesian networks. Decision trees classify instances by sorting them down the tree from the root to some leaf node, which provides the classification of the instance. Each node in the tree specifies a test of some attribute of the instance and each branch corresponds to one of the possible values for this attribute (Cho, Kim, & Kim, 2002). In contrast to the limited decision trees representation range, artificial neural networks can represent real-valued, discrete-valued and vector-valued functions. The classifier-based models often take as input the *usage history and ratings*. The usage history is a log of the user transaction or interaction with the personalization system, which can be seen as a form of implicit user profile. It is a very practical model used in learning and adapting the user profile (Kang, Lim, & Kim, 2005).

Finally, the recent emerge of the Semantic Web technologies has led to *ontology-based representation* in user profiling. The Semantic Web vision of a next generation Web provides the mechanisms to identify those resources that better satisfy the requests not only on the basis of descriptive keywords but also on the basis of knowledge. The most common ways of representing semantic user profiles are the ontology-based and description logic based representations (Baldoni, Baroglio, & Henze, 2005). In recent work, semantic Web languages, such as Resource Description Framework (RDF), Ontology Web Language (OWL) are used to represent users and their semantic preferences. Gauch, Chaffee and Pretschner (2003) exploit hierarchical structures in ontologies to imply generalizations of user preferences upward in topic hierarchies (e.g., interest in football implies interest in sports).

Related Content