Towards User Profile Modelling in Recommender System

Djallel Bouneffouf
Department of Computer Science
Telecom SudParis
Djallel.bouneffouf@it-sudparis.eu

Abstract

The notion of profile appeared in the 1970s decade, which was mainly due to the need to create custom applications that could be adapted to the user. In what follows, we treat the different aspects of the user’s profile, defining it, profile, its features and its indicators of interest, and then we describe the different approaches of modelling and acquiring the user’s interests.

1 Introduction

Recommender System (RS) belong to a more general framework of systems called Personalized Access Systems. These systems integrate the user as an information structure, in the process of selecting relevant information to him [Shani et al. 2007].

In a RS, the user’s profile is a kind of query about his interests, describing features that can be shared with a group of individuals. Comparing these features to incoming documents, the system can select those likely to be of interest [Baeza-Yates and Ribeiro-Netel 1999].

1.1 User’s Profile Features

The features that characterize the user’s profile are his acquired knowledge in different subjects (background), his objectives (goals) and his interests [Brusilovsky 2001].

**Background**: The background concerns all the information related to user’s past experience. This includes his profession, experience in fields related to his work, and how the user is familiar with the working environment of the system.

**Objectives**: The objectives are the user’s needs when he searches for information.

**Interests**: The interests are the documents that the user has consulted or notified to the system directly or indirectly through its feedback by indicators of interests.

1.2 Indicators of Interests

The implicit detection of interests is through observable behaviours collected by the system when the user interacts with his environment. In this context several behaviours can be considered, such as:

- Click of the mouse on a document;
- Scrolling a document using the mouse or keyboard;
- Time spent on a document;
- Printing a document;
- Saving a document;
- Copying/pasting all or part of a document;
- Document annotation;
Qualitative evaluation of a document;
- Navigating to reach a document;
- Eye-tracking.

These behaviours during user/system interactions are indicators of the relevance of a document and provide useful evidence to predict interests indirectly expressed by the user.

The structures that store the user’s profile features are modelled as described in what follows.

## Modelling the User’s Profile

User modelling is a research field that concerns the improvement of man-machine interaction by predicting the interests of users [Shani et al. 2007]. Modelling the user’s profile consists of designing a structure for storing all the information which characterizes the user and describes his interests, his background and his objectives [Salton et al. 1975].

There are several ways to model or represent the user’s profile. We describe here the different existing techniques.

### 2.1 Vector Representation

Vector representation is based on the classic vector space model of Salton [Salton et al. 1975], where the profile is represented as an m-dimensional vector, where each dimension corresponds to a distinct term and m is the total number of terms that exist in the user’s profile.

The vector representation has been the first model of the user’s profile exploited. The weighting of terms is usually based on a diagram of the TF/IDF format commonly used in IR [Salton et al. 1975]. The weight associated to each term represents the degree of importance in the user’s profile. Different RS use such representation, like [Mukhopadhyay et al. 2008] an on line newspaper or [Chan et al. 2011] concerning the recommendation of web services.

In addition to its simplicity of implementation, the use of several vectors to represent the profile permits to take into account the different interests and their evolution through time; but the default of this representation is in the lacks of structure and semantic (no connexion between terms).

### 2.2 Connexion Representation

The connexion representation is based on an associative interconnection of the nodes representing the user’s profile.

In this context, the system proposed by [Wibowo et al. 2011] uses the concepts existing in Word-Net to group similar terms. The user’s profile is then represented as a semantic network in which each group of concepts is represented by nodes and arcs.

A similar approach has been used by [Mezghani et al. 2012]. Initially, each semantic network contains a collection of nodes in which each node represents a concept. The nodes contain a single vector of weighted terms. When a new user information is collected, the profile is enriched by integrating weighted terms in the corresponding concepts.

This representation has the advantage of structuring [Wibowo et al. 2011], but the problem of the connexion representation is the absence of hierarchical relation among the concepts of the network, reflecting the semantic generalization / specialization between concepts (e. g., Biology is one specialization of Science).

### 2.3 Ontologies Representation

Ontologies may be used to represent the semantic relations among the informational units that make the user’s profile [Middleton et al. 2004]. This representation allows to overcome the limitations of the connexion representation by presenting the user’s profile in the form of a hierarchy of concepts.

Each class in the hierarchy represents the knowledge of an area of the user’s interests. The relationship (generalization / specification) between the elements of the hierarchy reflects a more realistic interest of the user.

The representation of the profile based on ontologies creates some problems related to the hetero-
geneity and diversity of the user’s interests, for-instance, the users may have different perceptions of the same concept, which leads to inaccurate representations Godoy and Amandi [2005].

2.4 Multidimensional Representation

The user’s profile can contain several types of information such as demographics, interests, purpose, history and other information Niederée et al. [2004]. In Lakiotaki et al. [2011], the authors represent the contents of the user’s profile by a structured model with a predefined category called dimensions: personal data, data source, data delivery, behavioural data and security data. This model was proposed in the development of a digital library service.

In the same context, authors in Kostadinov et al. [2007] propose a set of open dimensions that can contain most of the information that characterizes the user. The authors propose eight dimensions: personal data, the focus, the domain ontology, the expected quality, customization, security, preferences, miscellaneous information.

The multidimensional representation has the advantage of providing a better interpretation of the semantics of the user’s profile.

3 Acquiring the User’s Profile

In the recommendation process, the crucial step is the construction of the user’s profile that truly reflects his interests. However, this step is not an easy task because the user may not be sure on his interests Bila et al. [2008], on one hand and often does not want or even can not make efforts for its creation, on the other hand.

In this context, several approaches have been proposed for the acquisition of the user’s profile.

3.1 Simplistic Approach

This approach promotes the description of the user’s profile through a set of keywords explicitly provided by the user Baeza-Yates and Ribeiro-Neto [1999]. The method requires much from the user because, if he is not familiar with the system and the vocabulary of incoming documents, it becomes difficult for him to provide the proper keywords that describe his interests.

3.2 Dynamic Approach

An alternative approach is to build the user’s profile by dynamically collecting information about his preferences. At the first step, the user must provide a set of keywords describing preferences to initialize the profile; then, at each arrival of a new document, the system uses the user’s profile to select the documents potentially fitting his interests, and displays them to the user. The user indicates one relevant document and one irrelevant document to him. This information is used to adjust the description of the user’s profile in order to reflect his new preferences Baeza-Yates and Ribeiro-Neto [1999].

3.3 Machine Learning Approach

Machine learning is an essential step in designing an automatic RS Lopez-Lopez et al. [2009]. As it is not reasonable to ask the user a set of keywords describing his preferences, the idea is to observe the user’s behaviour through his interactions with the system to learn his profile. Most systems construct the user’s profile by learning from consulted documents. This profile is generally based on the vector model and different indicators of interests are used, such as the movements and mouse clicks, for example. The terms weight adjustment is often used with learning techniques such as neural networks, genetic algorithms and others Anderson [2002].

4 Conclusion and Discussion

We present now a synthesis of the different approaches to represent and acquire the user’s profile discussed above.
Table 1: Representing the user’s profile

| Representation technique | Advantage | Disadvantage |
|--------------------------|-----------|--------------|
| Vector representation    | Takes into account the diversity of interests and their evolution through time and is easy to implement | The lack of structure of the data |
| Connectionist representation | Semantic relationships between interests | The absence of hierarchical relations between the network’s concepts |
| Ontology representation  | Hierarchical relations between the interests of the user’s profile | Problems related to the heterogeneity and the diversity of the user’s interests |
| Multidimensional representation | Better interpretation of the semantics of the user’s profile | Ambiguity in interpreting the roles of each dimension |

Table 1 summarizes the advantages and disadvantages of different approaches on user’s profile modelling. From the table, we note that the connectionist representation solves the shortcomings of the vector representation by establishing relationships between interests in the user’s profile. Moreover, the ontology representation allows to expand the limits of connexion representation by including a hierarchy between interests. Finally, the multidimensional representation allows a better interpretation of the semantics of the user’s profile.

Table 2: Acquisition of the user’s profile

| Representation technique | Advantage | Disadvantage |
|--------------------------|-----------|--------------|
| Simplistic approach      | The system is sure on the user’s interests | Hard for the user to provide the appropriate keywords, the profile is static and the system is user-dependent |
| Dynamic approach         | The system is sure on the user’s interests, it allows the system to change of the profile | Hard for the user to provide the appropriate keywords and the system is user-dependent |
| Machine learning approach | Learning is not restricted to the keywords already entered by the user | The problem of cold start (no recommendation when the user profile is empty) |

Table 2 summarizes the advantages and disadvantages of different user’s profile acquisitions approaches. Another observation is that the learning approach and the dynamic approach are complementary, in the sense that the learning approach solves the user dependence found in the dynamic approach, and the dynamic approach solves the cold start problem that exists in the learning approach. In short, the best approaches to model and acquire the user’s profile are respectively a multidimensional approach and a hybrid between the dynamic and the learning approaches.

The representation and acquisition of the user’s profile is a part of the recommendation process.

References

Corin Ross Anderson. *A machine learning approach to web personalization*. PhD thesis, 2002. AAI3053472.

Ricardo A. Baeza-Yates and Berthier Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1999.

Nilton Bila, Jin Cao, Robert Dinoff, Tin Kam Ho, Richard Hull, Bharat Kumar, and Paulo Santos. Mobile user profile acquisition through network observables and explicit user queries. In *Proceedings of the The Ninth International Conference on Mobile Data Management*, MDM ’08, pages 98–107, Washington, DC, USA, 2008. IEEE Computer Society.

Peter Brusilovsky. Adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 11(1-2):87–110, March 2001.
Nguyen Ngoc Chan, Walid Gaaloul, and Samir Tata. A web service recommender system using vector space model and latent semantic indexing. In Proceedings of the 2011 IEEE International Conference on Advanced Information Networking and Applications, AINA ’11, pages 602–609, Washington, DC, USA, 2011. IEEE Computer Society.

Daniela Godoy and Analía Amandi. User profiling in personal information agents: a survey. Knowledge Eng. Review, 20(4):329–361, 2005.

Dimitre Kostadinov, Mokrane Bouzeghoub, and Stéphane Lopes. Query rewriting based on user’s profile knowledge. In BDA, 2007.

Kleanthi Lakiotaki, Nikolaos F. Matsatsinis, and Alexis Tsoukias. Multicriteria user modeling in recommender systems. IEEE Intelligent Systems, 26(2):64–76, 2011.

L. M. Lopez-Lopez, J. J. Castro-Schez, D. Vallejo-Fernandez, and J. Albusac. A recommender system based on a machine learning algorithm for b2c portals. In Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Volume 01, WI-IAT ’09, pages 524–531, Washington, DC, USA, 2009. IEEE Computer Society.

Manel Mezghani, Corinne Amel Zayani, Ikram Amous, and Faiez Gargouri. A user profile modelling using social annotations: a survey. In Proceedings of the 21st international conference companion on World Wide Web, WWW ’12 Companion, pages 969–976, New York, NY, USA, 2012. ACM.

Stuart E. Middleton, Nigel R. Shadbolt, and David C. De Roure. Ontological user profiling in recommender systems. ACM Trans. Inf. Syst., 22(1):54–88, January 2004.

Debajyoti Mukhopadhyay, Ruma Dutta, Anirban Kundu, and Rana Dattagupta. A product recommendation system using vector space model and association rule. In Proceedings of the 2008 International Conference on Information Technology, ICIT ’08, pages 279–282, Washington, DC, USA, 2008. IEEE Computer Society.

Claudia Niederée, Avaré Stewart, Bhaskar Mehta, and M. Hemmje. A Multi-Dimensional, Unified User Model for Cross-System Personalization. In E4PIA Workshop 2004, 2004.

G. Salton, A. Wong, and C. S. Yang. A vector space model for automatic indexing. Commun. ACM, 18(11):613–620, November 1975.

Guy Shani, Lior Rokach, Amnon Meisles, Lihi Naamani, Nischal Piratla, and David Ben-shimon. Establishing user profiles in the mediascout recommender system, 2007.

Adi Wibowo, Andreas Handoko, and Albert Halim. Application of Topic Based Vector Space Model with WordNet. In International Conference on Uncertainty Reasoning and Knowledge Engineering, 2011.