Available online at www.sciencedirect.com
ScienceDirect
Knowledge-Based Systems 21 (2008) 305–320
Knowledge-Based SYSTEMS
www.elsevier.com/locate/knosys

A flexible semantic inference methodology to reason about user preferences in knowledge-based recommender systems ☆

Yolanda Blanco-Fernández [a,*], José J. Pazos-Arias [a], Alberto Gil-Solla [a], Manuel Ramos-Cabrer [a], Martín López-Nores [a], Jorge García-Duque [a], Ana Fernández-Vilas [a], Rebeca P. Díaz-Redondo [a], Jesús Bermejo-Muñoz [b]

[a] *ETSE Telecomunicación, Campus Universitario, Vigo 36310, Spain*
[b] *Telvent, Ronda de Tamargillo, 29, Sevilla 41006, Spain*

Received 22 December 2006; received in revised form 18 June 2007; accepted 28 July 2007
Available online 9 August 2007

Abstract

Recommender systems arose with the goal of helping users search in overloaded information domains (like e-commerce, e-learning or Digital TV). These tools automatically select items (commercial products, educational courses, TV programs, etc.) that may be appealing to each user taking into account his/her personal preferences. The personalization strategies used to compare these preferences with the available items suffer from well-known deficiencies that reduce the quality of the recommendations. Most of the limitations arise from using syntactic matching techniques because they miss a lot of useful knowledge during the recommendation process. In this paper, we propose a personalization strategy that overcomes these drawbacks by applying inference techniques borrowed from the Semantic Web. Our approach reasons about the semantics of items and user preferences to discover complex associations between them. These semantic associations provide additional knowledge about the user preferences, and permit the recommender system to compare them with the available items in a more effective way. The proposed strategy is flexible enough to be applied in many recommender systems, regardless of their application domain. Here, we illustrate its use in AVATAR, a tool that selects appealing audiovisual programs from among the myriad available in Digital TV.

Keywords: Recommender systems; Semantic Web; Ontologies; Semantic reasoning; Content-based filtering

1. Introduction

In recent years, we have witnessed an exponential growth in the amount of information available in diverse domains (like e-commerce, e-learning or Digital TV). This overload is hard to digest for users, who have difficulties to find really relevant items (e.g. commercial products, educational courses, TV programs). Recommender systems arose in the middle 1990s to provide assistance in these searching tasks; to this aim, they automatically select those items that may be appealing to each user considering the preferences defined in his/her personal profile.

The first recommendation strategies were the so-called *content-based methods* [2,44]. This technique suggests items similar to the ones that each user liked in the past. The similarity metrics employed limit the quality of the offered recommendations, because they are based on more or less sophisticated matching techniques that miss a lot of knowledge during the personalization process. In fact, due to their syntactic nature, the existing metrics only detect

☆ Work supported by the Spanish Ministry of Education and Science Projects TSI2004-03677 and TSI2007-61599, and by the Xunta de Galicia project PGIDIT05PXI32204PN.
* Corresponding author. Tel.: +34 986813967; fax: +34 986812100.
E-mail addresses: yolanda@det.uvigo.es (Y. Blanco-Fernández), jose@det.uvigo.es (J.J. Pazos-Arias), agil@det.uvigo.es (A. Gil-Solla), mramos@det.uvigo.es (M. Ramos-Cabrer), mlnores@det.uvigo.es (M. López-Nores), jgd@det.uvigo.es (J. García-Duque), avilas@det.uvigo.es (A. Fernández-Vilas), rebeca@det.uvigo.es (R.P. Díaz-Redondo), jesus.bermejo@telvent.abengoa.com (J. Bermejo-Muñoz).

0950-7051/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.knosys.2007.07.004

similarity between items that share the *same* attributes or features. This causes overspecialized recommendations that only include items very similar to those the user already knows.

Instead of fighting the *overspecialization* of content-based methods, researchers proposed new personalization strategies, such as collaborative filtering [21,19,37] and hybrid approaches mixing both techniques [4,9,40]. To diversify the offered recommendations, collaborative recommender systems suggest to each user items that were appealing to other users with similar tastes. Regardless of its success in many application domains, collaborative filtering has two serious drawbacks. On the one hand, it requires knowing many user profiles in order to elaborate accurate recommendations for a given user. On the other, its applicability and quality are limited by the so-called *sparsity problem*, which occurs when the available data are insufficient for identifying similar users[1] [9].

In order to develop an approach that provides high-quality recommendations (even when data are sparse), this paper rescues the content-based methods by developing effective mechanisms against *overspecialization*. Specifically, our approach overcomes this problem – due to the syntactic nature of the existing similarity metrics – by borrowing inference techniques from the Semantic Web. This initiative is based on describing the Web resources by metadata, so that computers can understand their semantics and infer relationships between them. Such a reasoning[2] process requires formalizing the semantic descriptions of the resources in a commonly agreed and reusable way; for that purpose, the Semantic Web community resorts to *ontologies*, i.e. conceptualizations that identify typical concepts and relationships in a specific application domain. Concepts are identified by means of classes and relationships are represented as properties (both hierarchically organized), and besides, specific instances of classes and properties are included in this formal knowledge base.[3]

Bearing in mind the results achieved in the Semantic Web, our proposal includes reasoning about the semantic descriptions of the items available in the recommender system (formalized in a domain ontology), and inferring implicit semantic relationships between them. Such relationships diversify the recommendations because they allow establishing correspondences between the user preferences and other items appealing to him/her that do not necessarily share the same features. In other words, our content-based strategy suggests items *semantically related* to those the user liked in the past, not just items *similar* to his/her preferences.

In order to discover as much knowledge about the user preferences as possible, our approach looks for relationships hidden in the system ontology. Even though such relationships are already defined in the Semantic Web community, they have never been included in a personalization environment like a recommender system. For that reason, we develop a new inference methodology to take advantage of the benefits of semantic reasoning in our recommendation strategy. The main features of this methodology are the following ones:

- Firstly, it explores extensively the knowledge base and discovers the hidden relationships from the entities (classes and their instances), properties and hierarchical links explicitly formalized in it. Such relationships provide the knowledge missed by the current approaches and permit to compare the user preferences with the available items in a more effective way, beyond the traditional syntactic similarity metrics.
- Secondly, our reasoning methodology ensures that the inferred relationships are relevant for the user, and that they adapt in a flexible way as his/her preferences evolve over time. Thereby, our content-based strategy guarantees the diversity of the recommendations without risking their personalized nature.
- Finally, our methodology is not exclusively joined to a specific domain, as it is flexible enough to be used in diverse personalization applications. In this paper, we illustrate its use in AVATAR, a tool that selects appealing audiovisual programs from among the myriad available in Digital TV.

This paper is organized as follows: Section 2 reviews the most well-known matching techniques used in traditional content-based approaches, as well as the proposals defined in the Semantic Web for querying ontologies and inferring knowledge from them. Section 3 describes the two key elements of our semantic reasoning methodology, applied in the AVATAR system: the ontology about the TV domain and the user profiles that store their personal preferences. The semantic relationships adopted in our approach and the inference mechanism from the system knowledge base are detailed in Sections 4 and 5, respectively. Section 6 shows an example of the reasoning-based recommendations elaborated by AVATAR. Section 7 compares our proposal with other similar approaches defined in the literature. Finally, Section 8 draws some conclusions and describes our future work.

2. Background

2.1. Conventional matching techniques

As we mentioned in the previous section, content-based recommender systems resort to matching techniques to compare the user preferences to the available items. The

[1] As the amount of products available in the collaborative recommender system increases, it is less likely that two users rate the same products in their profiles. Since there is no overlapping among their profiles, it is hard to find users with similar tastes.

[2] We take *reasoning* as a synonym for *inference*.

[3] The terms *ontology* and *knowledge base* are used without distinction.