Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 35 (2008) 1386–1399
Expert Systems with Applications
www.elsevier.com/locate/eswa

Fuzzy-genetic approach to recommender systems based on a novel hybrid user model

Mohammad Yahya H. Al-Shamri , Kamal K. Bharadwaj

School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi 110 067, India

Abstract

The main strengths of collaborative filtering (CF), the most successful and widely used filtering technique for recommender systems, are its cross-genre or 'outside the box' recommendation ability and that it is completely independent of any machine-readable representation of the items being recommended. However, CF suffers from sparsity, scalability, and loss of neighbor transitivity. CF techniques are either memory-based or model-based. While the former is more accurate, its scalability compared to model-based is poor. An important contribution of this paper is a hybrid fuzzy-genetic approach to recommender systems that retains the accuracy of memory-based CF and the scalability of model-based CF. Using hybrid features, a novel user model is built that helped in achieving significant reduction in system complexity, sparsity, and made the neighbor transitivity relationship hold. The user model is employed to find a set of like-minded users within which a memory-based search is carried out. This set is much smaller than the entire set, thus improving system's scalability. Besides our proposed approaches are scalable and compact in size, computational results reveal that they outperform the classical approach.

Keywords: Recommender systems; Collaborative filtering; Web personalization; User model; Fuzzy sets

1. Introduction

The popular use of web as a global information system has flooded us with a tremendous amount of data and information. The end users are overloaded with a huge amount of information from myriad resources. This explosive growth of data has generated an urgent need for powerful automated web personalization tools that can intelligently assist us in transforming the vast amount of data into useful information and knowledge. In other words, these tools ensure that the right information is delivered to the right people at the right time (Adomavicius & Tuzhilin, 2005; Eirinaki & Vazirgiannis, 2003). Web recommender systems (RS), the most successful example of Web personalization tools, tailor information access, trim down the information overload, and efficiently guide the user in a personalized manner to interesting items within a very large space of possible options (Burke, 2002). Typically RS recommend information (URLs, netnews articles), entertainment (books, movies, restaurants), or individuals (experts). Amazon.com (Linden, Smith, & York, 2003) and MovieLens.org (Miller, Albert, Lam, Konstan, & Riedl, 2003) are two well-known examples of RS on the web.

Recommender systems employ four information filtering techniques, demographic filtering (DMF) (Krulwich, 1997), content-based filtering (CBF) (Lang, 1995), collaborative filtering (Resnick, Iakovou, Sushak, Bergstrom, & Riedl, 1994; Shardanand & Maes, 1995), and hybrid filtering techniques (Balabanovic & Shoham, 1997; Pazzani, 1999; Shahabi, Banaei-Kashani, Chen, & McLeod, 2001). DMF categorizes the user based on the user personal attributes and makes recommendations based on demographic classes while the CBF suggests items similar to the ones the user preferred in the past.

* Corresponding author. Mobile: +91 (11) 9810196636; fax: +91 (11) 26717528.
E-mail addresses: mohamad.alshamri@gmail.com (M.Y.H. Al-Shamri), kbharadwaj@gmail.com (K.K. Bharadwaj).

0957-4174/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2007.08.016

The most familiar and most widely implemented filtering is the collaborative filtering. The initial CF was introduced by Tapestry (Goldberg, Nichols, Oki, & Terry, 1992), and was automated by GroupLens (Resnick et al., 1994) and Ringo (Shardanand & Maes, 1995). Typically, CF explores similar users (neighbors), recognizes commonalities between the user and his neighbors on the basis of their ratings, and then accordingly generates new recommendations based on inter-user comparisons (Adomavicius & Tuzhilin, 2005; Eirinaki & Vazirgiannis, 2003). Further, CF and DMF have the unique capacity to identify cross-genre niches and can entice users to jump outside of the familiar 'outside the box' (Burke, 2002).

Breese, Heckerman, and Kadie (1998) classified CF algorithms as either memory-based (Resnick et al., 1994; Shardanand & Maes, 1995), the entire data is used for recommendations, or model-based (Shahabi et al., 2001) in which a model is derived offline from the data to be used for online recommendations. While the former is more accurate, its scalability compared to model-based is poor. Practically, CF faces two fundamental challenges, namely accuracy and scalability. Although memory-based algorithms are simple, provide high accuracy recommendations, and admit easy addition of new data, but they are computationally expensive as the size of the input dataset increases. Eventually, the user will leave the Web site before the processing completes. On the other hand, applying the model-based algorithm alone on such sparse data, though reduces the online processing cost, often comes at the cost of recommendation accuracy. One common threat in current RS research is the need to combine recommendation techniques to achieve peak performance because each technique has its own pros and cons.

Essentially, RS keep a profile for each user. Without any information about the user, the RS are not able to assist the user. The user profile contains raw information about the user. The terms user profile and user model are often used as synonyms. However, recent researchers (Froschl, 2005; Koch, 2000) differentiate between them according to the level of sophistication. The user profile is a collection of raw personal information represents preferences, background, personal details, and interactions with the system. Depending on the user profile, a user can be modeled. Thus, the user profile is used to retrieve the needed information to build up a model for the user. Koch (2000) describes the user model as the representation of the system's beliefs about the user and the user profile as a simple user model.

Some efforts have been made towards introducing fuzziness in RS. Nasraoui and Petenes (2003) used fuzzy approximate reasoning to develop a general framework for the recommendation process while Suryavanshi, Shiri, and Mudur (2005) used relational fuzzy subtractive clustering. Shahabi et al. (2001) proposed a Yoda RS that softly classifying active user based on typical patterns of users and then generating soft recommendations for him. The user profile includes many features that can be described as fuzzy. But at the item level it is difficult to fuzzify the profile because it would require prohibitively large space and long processing time.

Our work in this paper is an attempt towards introducing hybridization at four different levels, namely feature-level, model-level, CF algorithm-level, and approach-level. Firstly, we proposed hybrid features that exploit both user ratings for high rated items and some content descriptions of the items. At the model-level we built a user model from the set of hybrid features and DMF profile. Hybridization between model-based and memory-based algorithms of CF is done at the CF algorithm-level. The user model is used to find a set of like-minded users within which a memory-based search is carried out. This set is much smaller in size than the original set, thus making the technique scalable. Finally, we developed a hybrid fuzzy-genetic RS by employing GA to evolve appropriate weights for each feature of the user model and proposing a novel fuzzy distance metric to match users.

The contributions of this paper are three-fold:

- A novel user model is built that enables hybrid filtering, reduces the system complexity and the computational time by roughly a factor of six.
- A novel fuzzy distance function is proposed for users matching.
- A hybrid fuzzy-genetic approach to recommender systems is developed.

The rest of this paper is organized as follows: some background on the RS is given in the next section. In Section 3, the proposed user model is presented, while the fuzzy and hybrid fuzzy-genetic approaches are given in Section 4. The experimental results of the proposed approaches and classical approach are discussed in Section 5. Finally, in the last section we conclude our work with a review of our contributions along with some future research directions.

2. Background

2.1. Recommender systems

Recommender systems have gained an increasing importance since the early work on CF in the mid-1990s when researchers started focusing on RS that explicitly rely on the ratings structure (Adomavicius & Tuzhilin, 2005). Normally, explicit ratings from users are binary ratings (like/dislike) or follow a specified numerical scale indicating the degree of preference (e.g., 1 – bad to MAX – excellent, where MAX is the maximum possible rating for a given system). Usually, one of the following information filtering techniques is employed to generate recommendations:

- *Demographic filtering (DMF)*: The user will be recommended items similar to the ones other people with same