Expert Systems With Applications 53 (2016) 204–218

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

An evolutionary approach for combining results of recommender systems techniques based on collaborative filtering

Edjalma Queiroz da Silva*, Celso G. Camilo-Junior, Luiz Mario L. Pascoal, Thierson C. Rosa

Instituto de Informática, Universidade Federal de Goiás, Goiás 74001-970, Brasil

ARTICLE INFO

Keywords:
Recommender systems
Collaborative filtering
Combining techniques
Genetic algorithms

ABSTRACT

Recommender systems (RS) are often used as guides, helping users to discover products of their interest. Many techniques and approaches to generate an effective recommendation are available for the system designers. On the one hand, this is interesting because different application's scenarios could have a fittest solution but on the other it can also cause some complexity to select the best technique to address at each state of the database. Thus, choose the best technique for each new state becomes too difficult and frequent for manually select. One of big challenges on RS is turn the techniques more useful for real-world scenarios. Therefore, automate or help the design decision is an important task to improve the usability of RS and reduce its cost. Although many works aims to improve the performance of RS for some scenarios, just a few of them try to help the designers on selection or combination of the techniques through applications' state changes. Therefore, this work proposes an evolutionary approach, called Invenire, to automate the choice of techniques used by combining results of different recommendation techniques. This is a new approach that uses a search algorithm to optimize the techniques combination, and can inspire hybrid methods and expert systems on how automate them. To evaluate the proposal, experiments were performed with a dataset from MovieLens and different collaborative filtering approaches. The results obtained show that the Invenire outperforms all collaborative filtering approach separately in all contexts addressed. The improvement achieved varies from 3.6% to 118.99% depending on the combination encountered and the experiment executed. Thus, the proposal was able to increase the accuracy on the generated recommendations and automate the combinations of techniques.

1. Introduction

In order to eliminate doubts in situations where we have to choose among products or items we are faced with, we usually rely on recommendations passed on by others. These recommendations are given to us directly ("word of mouth") (Shardanand & Maes, 1995) or through texts and videos. Film critics, book reviewers, online social networks, and printed newspapers are examples of influencer. A recommender system helps to increase the capacity and effectiveness of transmitting and receiving suggestions, a well known process in the social relationships among human beings (Resnick & Varian, 1997). In a typical system, people provide evaluations of items they have bought or used. These evaluations are usually represented as ratings.[1] The recommender system uses these gradings from some users to suggest the best n items to others. These systems have big challenge to determine the best combination of user expectations and adequate item (products, services or people) to be recommended, i.e., discovering the relationship of interest and options is a major problem (Cacheda, Carneiro, Fernandez, & Formoso, 2011a).

Adomavicius and Tuzhilin (2005) classify recommender systems into three major categories regarding the approach used to generate the recommendations: (i) *content-based* approach, in which similar items to those the user showed preference in the past are recommended; (ii) *collaborative filtering*, which recommends items chosen by people with similar preferences to the user and; (iii) *hybrid* approaches that combine techniques of both previous approaches to attempt to solve some problems inherent to each of them in isolation.

Collaborative filtering (CF) is one of the most used recommendation technologies. This method calculate the similarity between users and uses this information to recommend items not yet tried by the target user (Hu & Pu, 2010). The similarity is based on past reviews of shared items. This similarity is used to generate recommendations of items that were previously evaluated by these

* Corresponding author. Tel.: +556282941869.
E-mail addresses: edjalmasilva@inf.ufg.br, edjalma@ambientinformatica.com.br (E.Q. da Silva), celso@inf.ufg.br (C.G. Camilo-Junior), luizpascoal@inf.ufg.br (L.M.L. Pascoal), thierson@inf.ufg.br (T.C. Rosa).

[1] These ratings are commonly represented as a grade in the range [1, 5] or as a number of "stars" in the same range.

http://dx.doi.org/10.1016/j.eswa.2015.12.050

similar users but that were not yet evaluated by the target user (Herlocker, Konstan, Borchers, & Riedl, 1999; Hu & Pu, 2010).

The CF has quickly become popular in the fields of academia and industry. Companies like Google, Amazon and Netflix make great use of this approach because of its significant competitive advantage. The CF approach basically follows four steps (Cacheda et al., 2011a):

1. Calculate the similarity of each user to the target-user (similarity metrics).
2. Select a subset of h neighbors, i.e., users with highest similarity to the target-user, in order to consider the ratings of these neighbors in the prediction.
3. Normalize ratings and compute the predictions considering the evaluations of neighbors with their weights. The weight in this case is the value of similarity between the neighbor and the target-user.
4. Sort items in decreasing order of predicted scores and present the best n items to the target-user.

The collaborative filtering algorithms can be classified into two types: *memory-based* algorithms and *model-based* algorithms. They basically differ in how they process the matrix of ratings (User χ Item). The *model-based* algorithms have two distinct phases. In the first stage, the algorithm handles the matrix of ratings to generate an efficient model that represents the original matrix. In the second step, this generated model is used as the input matrix for the calculation of prediction rating for target user (Cacheda et al., 2011a; Resnick, Iacovou, Suchak, Bergstrom, & Riedl, 1994; Shardanand & Maes, 1995). The *memory-based* collaborative filtering uses the entire matrix to calculate its prediction. First, it use any similarity measure to select users (or items) that are similar to the target-user. Then, the prediction ratings of target user are calculated from the ratings of his neighbors. Otherwise, the *memory-based* method is divided into two other algorithms. The first one is called *user-based* algorithms, where the method for obtaining neighbors is based on the user (Shardanand & Maes, 1995). And the second one is called *item-based* algorithms, where neighbors are based on item (Sarwar, Karypis, Konstan, & Riedl, 2000).

Because the algorithms are still inefficient in some cases, the development of new collaborative filtering algorithms has focused mainly on how to provide accurate recommendations (Goldberg, Nichols, Oki, & Terry, 1992). One of the analyzed points is how one can well calculate the similarity between users. Various techniques, e.g. Euclidean, Tanimoto and Pearson correlation, are presented in the literature to do this. On the one hand, the large amount of options is important because gives custom and good solutions for different domains. However, may cause doubt of which technique to choose for recommendation process. Each of these approaches have particularities and its performance depend on the context to be applied, therefore each case must be carefully analyzed before choosing which technique to adopt.

This process has high cost because the designer spends much time running experiments to decide the best technique. Furthermore, even that he has chosen a good one, when significant modifications happen in the database he should repeat the process because the algorithms' performance is highly dependent of rating matrix (created from database). Therefore, we propose the method that automatically combines some rankings of recommendations, resulting from the *memory-based* techniques, to get better result than any of them alone. Our key insight is that combining results one can get the best of each addressed technique without the complexity of choose or turn them to hybrid.

As the task of discovering a good combination manually is a difficult task, it is desired that the combination be automated. For such a matter, the work proposes a genetic algorithm(GA; Holland, 1975; Goldberg & Holland, 1988) able to automate the combination of results of different *memory-based* similarity techniques. Although many hybrid approaches were done (Adomavicius & Tuzhilin, 2005; Burke, 2002; 2007; Lu, Wu, Mao, Wang, & Zhang, 2015), the use of search algorithm for combine techniques and automate the designers' decision was not explored and has a great potential.

The GA was chosen because it is widely used in the literature. Moreover, GAs are known for their flexibility, ease of implementation, and effectiveness in performing global search in adverse environments. In this approach, the GA should be able to generate a list (L) of n items to be recommended. These items are selected from the ranking of techniques used in the combination. Therefore, the list formed by the GA depends on the performance of each technique. The techniques that achieve lower error (RMSE) will have more items among the n finals. An example of the composition of this list, in case of $\|L\| = 10$, would be: 3 items coming from the rank of technique A, 3 items coming from the rank of technique B, and 4 items arising from the rank of technique C, totaling 10 items in the final list proposed by the GA.

Four experiments was designed to test the proposal. The experiment one (Section 5.1) shows that proposal GA (specialist model) outperformed the base techniques in a minimum of 9.028% and a maximum of 48.21%. The experiment two (Section 5.2) shows the results for generalist model constructed by proposal. The idea is measure the impact of generalization in the scenarios. Although the effectiveness is worse than specialist models and then a few base techniques, the efficiency of generalist model is better than the specialist ones. So, it is an option for who wants low computational cost. The experiment three (Section 5.3) shows the performance of generalist model for different states of a database. This model outperformed six from 10 scenarios for a database ($M1$) and obtained the second best averages on all databases, very close to the best ones. Finally, the experiment four (Section 5.4) shows the results from the comparison experiments between generalist models, specialist models and base techniques on different stages of a database. The specialist model outperformed others on all scenarios followed by generalist models. The base techniques were worse in average.

Thus, this work has the following main contributions:

1. An automated and effective approach to select a good combination of recommendation techniques' results.
2. The cost reduction for recommender systems design.
3. A flexible method for different application scenarios.
4. A customizable method able to combine many techniques' results, including some from different paradigms.

The rest of the work is organized as follows: In Section 2 we briefly review some of the research literature related to our work. In Section 3 we present the main theoretical concepts needed to develop this work. In Section 4 we present our proposal. In Section 5 we present experiments and results. In Section 6, conclusions and future work are shown.

2. Related work

The first system created using the collaborative filtering (CF) approach was the *Tapestry* (Goldberg et al., 1992; Resnick & Varian, 1997), which was a system with complete capabilities of filtering electronic documents. For instance, a user can create a filtering rules for *e-mail* such as "Show me all documents answered by other members of my research group". However, this system required the users to determine the relevant predictive relationships. Thus, it were only valuable in small closed communities where everyone was aware of the interests and duties of other users.

From this, many others works were done to improve the CF systems. There are studies that make comparisons between traditional