Advanced Engineering Informatics 29 (2015) 830–839

Contents lists available at ScienceDirect

Advanced Engineering Informatics

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

Applying artificial immune systems to collaborative filtering for movie recommendation

Meng-Hui Chen [a], Chin-Hung Teng [b,c], Pei-Chann Chang [a,c,*]

[a] Department of Information Management, Yuan Ze University, Chung-Li 32023, Taiwan
[b] Department of Information Communication, Yuan Ze University, Chung-Li 32023, Taiwan
[c] Innovation Center for Big Data and Digital Convergence, Yuan Ze University, Chung-Li 32023, Taiwan

ARTICLE INFO

Article history:
Received 24 August 2014
Received in revised form 16 April 2015
Accepted 20 April 2015
Available online 4 May 2015

Keywords:
Recommendation system
Collaborative filtering
Artificial immune system

ABSTRACT

Collaborative filtering is a widely used recommendation technique and many collaborative filtering techniques have been developed, each with its own merits and drawbacks. In this study, we apply an artificial immune network to collaborative filtering for movie recommendation. We propose new formulas in calculating the affinity between an antigen and an antibody and the affinity of an antigen to an immune network. In addition, a modified similarity estimation formula based on the Pearson correlation coefficient is also developed. A series of experiments based on MovieLens and EachMovie datasets are conducted, and the results are very encouraging.

1. Introduction

Recommender systems have been an important research topic over the past twenty years. These systems learn users' preferences in order to make recommendations to them. As e-commerce continues to grow, more and more products are being purchased online, and there is an increasing customer demand for the large number of items available on websites to be filtered so that they can more easily find specific items that actually interest them. Recommender systems were developed for this purpose, and can predict user preference for an item. Companies that successfully employ recommender systems in their e-commerce business include Amazon and Netflix. They learn user needs and try to provide the necessary information to users via the recommender system, thus increasing sales and profits.

Personalized recommender systems obtain useful information from historical data, such as a user's interests and purchasing behavior, in order to recommend relevant information or products to that user. Personalized recommendation was first proposed by Robert Armstrong [1] in 1995, and in the same year Henry Lieberman [2] presented an intelligent personalized navigation system for web browsing. Yahoo announced a personalized web entry point called My Yahoo the following year, and since then many novel personalized recommendation concepts have been proposed, and many recommendation algorithms have been developed.

Recommender systems can be roughly classified into four types: content-based recommendation, knowledge-based recommendation, collaborative filtering recommendation, and hybrid recommendation. In this research, we focus on collaborative filtering recommendation because it is a widely used algorithm in this field. Collaborative filtering algorithm was introduced in the 1990s, and has been effectively used in many recommender systems [3,4]. It predicts a rating for a user based on the rating preferences of similar users. Most collaborative filtering algorithms operate by finding similar users and then predicting a rating of an item based on the preferences and previous ratings of those users.

Collaborative filtering techniques can be further categorized into two types, model-based and memory-based, depending on how the data are processed. Model-based collaborative filtering techniques aim at building a model to represent user rating data, and use that model to predict user preference for a specific item. On the other hand, memory-based algorithms employ all user rating data to predict a missed user rating of an item. Memory-based techniques can also be classified as user-based and item-based collaborative filtering. User-based collaborative filtering is the first automatic collaborative filtering method [5,6]. It works by finding other users with rating preferences similar to those of the target user, and uses their ratings to predict the target user's rating of the item in question. In contrast to user-based collaborative filtering, item-based collaborative filtering is developed from the perspective of the item. Item-based collaborative filtering was first introduced by Sarwar et al. [7], and has been used by Amazon.com [8].

The dataset applied in this research is MovieLens dataset and it was created in 1997 by the GroupLens. GroupLens is a research lab

* Corresponding author at: Department of Information Management, Yuan Ze University, Chung-Li 32023, Taiwan.
E-mail address: iepchang@saturn.yzu.edu.tw (P.-C. Chang).

http://dx.doi.org/10.1016/j.aei.2015.04.005
1474-0346/

in the Department of Computer Science and Engineering at the University of Minnesota. There are five important fields in the lab, one of those is recommendation. Liu et al. [28] proposed a collaborative filtering based recommendation system. One of the three datasets in the paper is MovieLens. Their research results show that the collaborative filtering framework established by user interest could make the forecast accuracy higher. Sarwar et al. [20] focused on the development of different correlation coefficients. The improved correlation coefficient derives the higher accuracy in MovieLen datasets than others.

Artificial immune system (AIS) is a technique that simulates the mechanism of a biological immune system fighting foreign pathogens. It has been successfully used in optimization problems and scheduling [9]. AIS has also been used in collaborative filtering for recommending items [10,11]. For example, Acilar and Arslan [10] employed AIS to solve data sparsity and scalability problems. However, as a model-based approach, it has limited accuracy; lower than some state of the art techniques.

In tradition, the user's rating in collaborative filtering is calculated based on Pearson Coefficient only. However, the predicted results of user's rating should not be confined by this rule only. In this research, a model based approach by combing AIS and collaborative filtering is proposed and applied to predict the user's rating instead. Using this model, the immune network of antibodies will be employed as a classification rule to predict the user's rating. Even the user in the same class but his rating can be calculated from different immune networks which will be more accurately to reflect the user's interest.

The remainder of this paper is organized as follows. First, we will give a simple review of related work in Section 2. Following this, Section 3 introduces our AIS algorithm for user or item classification, and describes our new AIS-based collaborative filtering system. The experiments are given and their results are discussed in Section 4. Finally, we conclude in Section 5.

2. Literature review

In this paper, we mainly focus on collaborative filtering techniques, thus detailed research work on collaborative filtering is surveyed in the following:

2.1. Collaborative filtering

A typical collaborative filtering scenario is one in which, given a set of users and a set of items, the users may rate a subset of items, while the system must predict a missing user rating for an item. Based on their information processing approaches, collaborative filtering techniques can be classified into two types: model-based and memory-based. Memory-based approaches predict the missed rating from a group of users or items with similar profiles. In this type of approaches, similarity calculation is a critical step for finding a group of similar users or items.

Well-known similarity metrics include the Pearson correlation coefficient [5], constrained Pearson [15], weighted Pearson correlation [16] and cosine similarity [17]. Although these similarity measures have been used in many collaborative filtering algorithms, some researchers are dissatisfied with their performance. As a result, novel similarity models are continuously being put forward [18,19]. For example, Liu et al. [19] proposed a similarity model taking the local context information of user ratings and the global preference of user behavior into account. They claim that this new model is more effective, especially in under cold user conditions.

Because memory-based collaborative filtering techniques achieve recommendation based on a group of similar users or items, they are also called neighbor-based methods. Neighbor-based approaches can be further classified unto two types, user-based [5,6] and item-based [7], according to their similarity calculation methods. User-based approaches filter information based on a group of similar users, while item-based approaches compute the similarity between items instead of user similarity.

In contrast to neighbor-based approaches, model-based approaches achieve item recommendation by first constructing a model and then predict user ratings based on this model. Many model-based approaches, such as SVD [20], factor analysis [21], neural networks [22], PCA [23], Bayes networks [17] and latent class models [24,25] have been proposed. Typically, model-based approaches tend to have lower prediction time than neighbor-based methods. However, many models are very complex, with large number of parameters to be estimated, and thus require a long time to learn the models. On the other hand, neighbor-based approaches are typically much simpler and easy to implement, and can produce reasonable accuracy if sufficient user rating information is available. Therefore, this research will adopt the neighbor-based approach by applying AIS in collaborative filtering for movie recommendation.

2.2. Artificial immune system

Artificial immune system is derived from the mechanism of a biological immune system fighting foreign pathogens. By using the adaptive immune response, this algorithm can be used to search for the solution to an optimization problem. Immune network theory was first proposed by Jerne [12] in 1974.

In 2003, Dasgupta et al. [13] explained that an immune system is a complex system. It has a strong information processing characteristics, such as feature selection, pattern recognition, learning and memory recall. Three major immunological principles are typically applied in an immune system. They are immune network theory, negative selection and clonal selection. In 2005, Alatas and Akin [14] proposed using an artificial immune system algorithm to mine fuzzy classification rules in order to improve classification accuracy. They used the AIS algorithm to find the best classification rule in that category, and saved it in the database to improve the classification accuracy. AIS has also been used in collaborative filtering [10,11]. As described in Section 1, its advantages lie in data reduction, and the reported accuracy cannot therefore compare with many state of the art collaborative filtering techniques.

In this research, we apply AIS in collaborative filtering for movie recommendation. The differences of our approach among other earlier approaches are 1. We propose new formulas in calculating the affinity between an antigen and an antibody and the affinity of an antigen to an immune network. This new formula can be applied either in user-based or item-based approaches. 2. A modified similarity estimation formula based on the Pearson correlation coefficient is also developed. 3. We also derive a prediction formula suitable for the resulting classification. From experimental results on the MovieLens dataset, we found that our system is able to produce prediction accuracy comparable to state of art techniques in terms of mean absolute error. In addition, the precision and recall rate of our system are also very high.

3. Proposed system

The main idea in this research is to develop a rating model by applying AIS in collaborative filtering for movie recommendation. As mentioned above, there are two types of neighbor-based collaborative filtering approaches; one is user-based [5,6], and the other is item-based [7]. The rating model can be used in both approaches. In this study, we call this approach as user- or