Information Systems 58 (2016) 87–104

Contents lists available at ScienceDirect

Information Systems

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

Dealing with the new user cold-start problem in recommender systems: A comparative review

Le Hoang Son *,[1]

VNU University of Science, Vietnam National University, Vietnam

ARTICLE INFO

Article history:
Received 15 September 2014
Received in revised form
6 October 2014
Accepted 7 October 2014
Recommended by D. Shasha
Available online 24 December 2014

Keywords:
Collaborative filtering
NHSM
New user cold start
Recommender systems

ABSTRACT

The Recommender System (RS) is an efficient tool for decision makers that assists in the selection of appropriate items according to their preferences and interests. This system has been applied to various domains to personalize applications by recommending items such as books, movies, songs, restaurants, news articles and jokes, among others. An important issue for the RS that has greatly captured the attention of researchers is the new user cold-start problem, which occurs when there is a new user that has been registered to the system and no prior rating of this user is found in the rating table. In this paper, we first present a classification that divides the relevant studies addressing the new user cold-start problem into three major groups and summarize their advantages and disadvantages in a tabular format. Next, some typical algorithms of these groups, such as MIPFGWC-CS, NHSM, FARAMS and HU–FCF, are described. Finally, these algorithms are implemented and validated on some benchmark RS datasets under various settings of the new user cold start. The experimental results indicate that NHSM achieves better accuracy and computational time than the relevant methods.

Contents

* Tel.: +84 904171284; fax: +84 438623938.
E-mail addresses: sonlh@vnu.edu.vn, chinhson2002@gmail.com
[1] Official address: 334 Nguyen Trai, Thanh Xuan, Hanoi, Vietnam.

http://dx.doi.org/10.1016/j.is.2014.10.001
0306-4379/

1. Introduction

The growing development of content-based systems that provide a large amount of data, such as videos, images, blogs, multimedia, and wikis, brings great challenges for analysts attempting to extract useful knowledge and capture meaningful events from the massive data. Machine learning tools should indeed be oriented to what users intend to do and how they want the results to be returned in a given format. An efficient tool that assists decision makers to choose appropriate items according to their preferences and interests and that is currently widely used is the *Recommender System* (RS). Ricci et al. [31] defined the RS as a special type of information system that (i) helps to make choices without sufficient personal experience of the alternatives, (ii) suggests products to customers, and (iii) provides consumers with information to help them decide which products to purchase. The RS is based on a number of technologies, such as information filtering, classification learning, user modeling and adaptive hypermedia, and it is applied to various domains to personalize applications by recommending items such as books, movies, songs, restaurants, news articles and jokes, among others. It has been applied to e-commerce to learn from a customer and recommend products that he or she will find most valuable from among the available products, thus helping the customer find suitable products to purchase. Some e-commerce RSs are named as follows [37,22]. For example, Amazon.com is the most famous e-commerce RS, structured with an information page for each book while providing details of the text and purchase information. Two recommendations are found herein, including books frequently purchased by customers who purchased the selected book and authors whose books are frequently purchased. eBay.com is another example that provides the Feedback Profile feature, which allows both buyers and sellers to contribute to the feedback profiles of other customers with whom they have done business. The feedback consists of a satisfaction rating and a specific comment about the other customer. On Moviefinder.com, customers can locate movies with a similar "mood, theme, genre or cast" through Match Maker or by their previously indicated interests through WePredict. We clearly recognize that RSs are becoming important and with increasing influence on various practical applications.

An important issue for RSs that has greatly captured the attention of researchers is *the cold-start problem*. This problem has two variants: the *new user cold-start* problem and the *new item cold-start* problem. The new item cold-start problem occurs when there is a new item that has been transferred to the system. Because it is a new product, it has no user ratings (or the number of ratings is less than a threshold as defined in some equivalent papers) and is therefore ranked at the bottom of the recommended items list. Moreover, this problem can be partially handled by staff members of the system providing prior ratings to the new item. Thus, the concentration of the cold-start problem is dedicated to the new user cold-start problem when no prior rating could be made due to the privacy and security of the system. It is difficult to give the prediction to a specific item for the new user cold-start problem because the basic filtering methods in RSs, such as collaborative filtering and content-based filtering, require the historic rating of this user to calculate the similarities for the determination of the neighborhood. For this reason, the *new user cold-start problem* can negatively affect the recommender performance due to the inability of the system to produce meaningful recommendations [33]. Addressing this problem has been the primary focus of various studies in recent years.

The aim of this paper is to provide a comparative review of those studies that could answer our research question "which (group of) algorithm is the most effective among all?". For this purpose, we first provide a classification that divides the relevant studies into three groups: (i) makes use of additional data sources; (ii) selects the most prominent groups of analogous users; and (iii) enhances the prediction using hybrid methods. A table that summarizes the advantages and disadvantages of all groups of methods is presented. Second, some typical algorithms of the groups of methods, such as MIPFGWC-CS [46] (the first group), NHSM [20] (the second group), FARAMS [17] and HU–FCF [42] (the third group), are described in detail. Finally, these algorithms are implemented and validated on some benchmark RS datasets, such as MovieLens [23] and Jester [12], under various settings of the new user cold start. The experimental results could reveal the answer for our research question stated above.

The remainder of the paper is organized as follows. In Section 2, we present a literature review of the relevant studies according to the three aforementioned groups. Section 3 elaborates on the four typical methods, namely, MIPFGWC-CS [46], NHSM [20], FARAMS [17] and HU–FCF [42]. Section 4 presents the comparative experiments of these algorithms involving benchmark RS datasets. Finally, Section 5 draws conclusions and delineates the future research directions.

2. Literature review

The beginning of this section starts with an example that clearly demonstrates the new user cold-start problem.

Example 1. We have a RS that includes three tables: the users' demographic data (Table 1), the movies' information (Table 2) and the rating (Table 3). This type of system is able to predict the user rating of a movie, which is expressed in Table 3. Nonetheless, the new user cold-start problem occurs with a new user, e.g., Kim (User ID: 6) in Table 1, who has no prior rating such that it is difficult to provide a prediction for the first movie, e.g., Titanic (ID: 1).