ACADEMIC
World Academic Union

ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 9, No. 3, 2014, *pp*. 181-188

Linear Time Growth Collaborative Filtering Based on Caching Techniques

Waleed M. Al-Adrousy [1], Hesham A. Ali [2], and Taher T. Hamza [3]
[1]Computer Science Department, Computers and information sys Faculty,
Mansoura University, Egypt. Corresponding author Contacts:
waleed_m_m@mans.edu.eg , Fax (Work): +20502223754
[2]Computer Engineering and systems Department, Engineering Faculty,
Mansoura University, Egypt, h_arafat_ali@mans.edu.eg.
[3]Computer Science Department, Computers and information sys Faculty,
Mansoura University, Egypt, taher_hamza@yahoo.com

(*Received October 10, 2013, accepted March 8, 2014*)

Abstract. *One of social networks features is to recommend some items for users suitable and their profiles and needs. One of the used techniques of item recommendation is to calculate the average rating of other users' ratings of that item. The calculated average can be used as a rank for that item. New users' ratings are made within the lifetime of social networks. This dynamic sized set of ratings can lead to a performance problem. Ranking is not fixed since new users are constantly evaluating items in networks. However, updating ranking by traditional averaging can have a quadratic time growth. This paper aims to convert this aggregation method into linear growth instead of quadratic. The suggested algorithm needed extra storage capacity. Caching is suggested to speedup ranking with preserving storage resources. Another target is to have a high hit-ratio with better utilization of small cache size compared to the traditional Greedy- Dual-Size-Frequency algorithm. At the end of this paper an evaluation of the proposed technique performance is provided based on simulated experimental results. The experiment results show that the proposed technique has better performance rather than traditional averaging recommender systems without caching.*

Keywords: Social networks, collaborative filtering, recommender systems, caching techniques.

1. Introduction

One of the most important tools for people to communicate with each other is the Social Networks. According to the free on-line dictionary of computing (http://dictionary.reference.com/browse/social+network), a social network is "a Web site where one connects with those sharing personal or professional interests, place of origin, education at a particular school, etc." [25]. One of users' interests is the need to ask for the best suitable service, product, person or even an article for some special purpose. Social networks have been used as a marketing way. However, not all advertisements are good as claimed or suitable to all users. So, the social networks relationships of friends have been used to guide users to select items suitable for them. This social network component is called a recommender system. Currently, there are two types of recommender systems, collaborative filtering and content-based filtering [7]. Collaborative filtering makes use of users' ratings to items, regardless of its content, while the content-based filtering does the opposite. Content-based filtering systems analyze the semantic content of profiles and pages in social network to suggest suitable items to users [5].

Recommender systems need to have a profile for each user [8]. User's profile can be static or dynamic [14]. A static profile contains user's basic information; preferences given by user explicitly such as preferences and selected friends. A dynamic profile is inferred by analyzing user's actions such as types of daily browsed pages. Both profile types have several modeling techniques, including simple databases and vector representation for user and his/her corresponding interests [7]. Vector Space Modeling (SVM) was the start of a newer technique that focuses on semantic which is named Latent Semantic Analysis [13]. Most recommender systems focus on dynamic profile because of its updated nature. Recommender systems can be evaluated according to accuracy and performance. Performance issues are the main point of interest in this

research. For many Different techniques there is a common step, aggregation of several users' scores based on some criteria. The hosting servers' resources of social networks should be managed optimally to avoid wasting time or cost of hardware. One of the major resource consuming computations is the repeated ranking overtime for items in the social networks. The ranking operation depends on having an average rating for an item given by some users' ratings. The included users in the operation can be all users –in case of global rating, or a subset of users who represent the friends to a specific inquiring user. The second case is used to calculate the user based recommendation which tries to use user's profile to detect how suitable the recommended item can be [23, 9].

The contributions of this is paper can be represented in achieving two main goals: (i) suggesting an enhancement to averaging technique to have a linear order of growth, and; this will be done via a suggested incremental equation that works on a pre-calculated values that indicate the aggregation of the old non-changing ratings. Also (ii) suggesting an enhanced form of Greedy- Dual-Size-Frequency (GDSF) caching. The proposed algorithm aims to achieve better hit-ratio by using a median timestamp of items age in cache with no need for an artificial intelligence training phase. This paper is organized as follows; section 2 presents a background of research domain, section 3 presents the suggested techniques, section 4 presents the testing results, and finally the conclusion is in section 5.

2. Related Work

There are some common problems in researches about recommender systems [18]; prediction accuracy, testing opinions' subjectivity, recommended items suitability rather than being high ranked, effective preferences inference, trust formation of recommended items and usability of recommender systems interface layout. Those problems focus on the *value* of the recommended items rather than the *performance* of recommendation process itself. Recommender systems had many forms over the few past years. Some expert systems such as "Syskill & Webert" and Web Browser Intelligence (WBI) [8] applied intelligence to learn user preferences and patterns of recommendations and decision trees. "Syskill & Webert" system allowed user to make symbolic ratings about items using words like *cold* or *hot*. The symbolic ratings were converted later by the system into numeric ratings [12]. This technique aimed to handle the problem of difference in modeling nature between the human view and machine the view. However, this approach used explicit user likeness actions where user selects the rating for the item. Another technique in recommender systems was modeling of general interests was the analysis of user's navigation actions items pages. Knowledge based filtering was discussed in the work of Bruke et al. [4]. Their work classified knowledge into three areas: the social knowledge about users' database, the individual knowledge about a particular user and content knowledge about the items in the network. They discussed some problems in knowledge based filtering such as the distribution of users to items. In that problem, few items gain the attention of users while some other good items may not have been discovered.

Some techniques of natural language processing and semantic analysis have been used in some researches such as the work of Santos and Boticario [21]. In that research a new concept of Semantic Educational Recommender Systems "SERS" was introduced to join e-learning with semantic analysis. In [19] a survey is introduced to explore the features of some social networks that include e-learning support. One of those discussed networks is "OpenStudy" which was a large social learning community.

Semantic analysis can be complex, so, another technique was to use text mining techniques such as *term frequency* x *inverted document frequency* (TD-IDF) which was used in the work of Middleton et al. [15, 17]. This technique is simple which tries to use a set of keywords that distinguish a content category from another. Each item description is searched for those keywords and the tested to see how frequent each keyword is contained in description text. Of course, selecting the keywords set needs another set of techniques to update that set dynamically and also prevent term duplication. Term duplication can happen when many keywords are added by users but have the same meaning [24]. The focus in this research is on two of the several problems facing the collaborative filtering; *aggregation* used to predict rating for a user to a new item, and *caching* that can be used to enhance that king of computation. More discussion is presented in the following subsections. The content based filtering is beyond the scope of this study.

The first problem in collaborative filtering is the *aggregation* process. The most common case of rating aggregation is applied during averaging users' ratings. One variation of the averaging technique was