Far East Journal of Electronics and Communications

http://www.pphmj.com
http://dx.doi.org/10.17654/EC017030539
Volume 17, Number 3, 2017, Pages 539-544 **ISSN: 0973-7006**

IMPROVING DIVERSITY USING BANDWAGON EFFECT FOR DEVELOPING RECOMMENDATION SYSTEM

Suk-Kyoon Kang[1], Kiejin Park[2,*] and Limei Peng[1]

[1]Department of Industrial Engineering
Ajou University
South Korea

[2]Department of Integrative Systems Engineering
Ajou University
South Korea

Abstract

The recommendation system using collaborative filtering (CF) methods is widely used. However, it is short of recommending only similar items that are popular with users. To break this limitation of CF method, we design the recommending system based on the psychology concept of bandwagon effect. Generally, consumers decide what they are going to purchase based on what others have purchased. This is called bandwagon effect. To design the recommendation system based on bandwagon effect, we use the matrix factorization (MF) based alternating least square (ALS). Moreover, to store big data and computing, we construct a cluster based on in-memory framework spark and accomplish the development

Received: March 13, 2017; Accepted: May 1, 2017

Keywords and phrases: bandwagon effect, alternating least squares, recommendation system, collaborative filtering, spark.

A preliminary version of this paper was presented at the ICIDB-2016, Seoul, South Korea.

*Corresponding author

Communicated by Gyanendra Prasad Joshi

and computing of recommendation system. In order for improving the recommendation diversity, we compare the recommendation list from the existing recommendation system and our proposed recommendation system and it showed that our proposed system indicated better diversity during recommendation.

1. Introduction

Collaborative filtering (CF) uses users' evaluation grades, connects similar users, uses the results to predict the objective evaluation grades and provides users the most appropriate item. However, CF suffers from the following problems: (1) It suffers from cold-start problem for users with insufficient information. (2) It lacks the scalability problem to process big data [1, 2]. On the other hand, recent CF recommendation system uses the past evaluation grade as an objective to predict the users' interest and thus it is quite possible to recommend a similar item of interest in the past recommended item. This would prevent us to recommend diverse items for users.

In this paper, since matrix factorization (MF) based ALS algorithm is very effective against solving the sparsity and scalability problems of CF, in order to extend the recommendation variety, we propose the recommendation system based on a psychology concept called *bandwagon effect*. Besides, for computing big data, we use the open-source Apache Spark cluster to increase the processing speed, which is based on Hadoop and distributed in-memory computing framework [3, 4].

2. Related Works

(a) MF (matrix factorization) and ALS (alternating least square)

MF first separates matrix into two feature vectors. Based on this, it traces the column that is near to the original column and predicts an evaluation vector that does not exist. Equation (1) shows the method of evaluating data using MF method [5]:

$$\min_{PQ}|R - PQ|^2 = \min_{p,q} \sum_{u,i \in k} (r_{u,i} - p_u q_i)^2, \tag{1}$$

R is the rating matrix representing the user's preference $(r_{u,i})$ for the item. P is a user factor matrix that consists of vectors of a user's (p_u) affinity to factors and Q is an item factor matrix that consists of vectors with an item's (q_i) relation to the factor. The key of MF evaluation method is to decide these factors. In fact, the number of factors is small and thus is called "*low-rank approximation*". However, to decide the exact value of factors, we need an optimization process. The equation of MF, say equation (1), has the problem of "non-convex" and thus we need to convert it to the status of "convex". An appropriate algorithm for the problem is the ALS which is suitable for distributed processing environments.

3. Bandwagon Effect and Recommending System

Bandwagon effect means the phenomenon that an item is popular with most of the people. Generally, consumers decide what they are going to purchase based on what others have purchased. This is called *bandwagon effect*.

(a) Repeatability and sustainability

To react to the bandwagon effect, we need to generate the popular item (PI). Equations (2) and (3) are used to define the PI. Repeatability indicates that how frequently people find the items and sustainability means that how long people are interested in the selected items, which are used to trace the popularity of an item:

$$c_i \text{ (repeatability)}, \tag{2}$$

$$s_i = t_l - t_f \text{ (sustainability)}, \tag{3}$$

c_i is defined by the purchasing frequency of a specific item and s_i is defined by the time period between the time point an item is first purchased and the time point it is last purchased.

(b) Expectancy

Expectancy is defined by comparing the similarity between the selected item and the rest items. Using item feature vectors in (1), the similarity between the popular item group and other items is calculated and the expectancy of each item can be defined as equation (4):

$$e_{g,i} = sim(i_g, i_j) = \frac{\sum_{k=1}^{n} w_{g,k} \cdot w_{j,k}}{\sqrt{\sum_{k=1}^{n} w_{g,k}^2 \cdot w_{j,k}^2}}, \tag{4}$$

where items $w_{g,k}$ and $w_{j,k}$ define the *n* feature vector for the popular item and the other items, respectively.

Similarity between the extracted popular item group feature vector $(w_{g,k})$ and other item $(w_{j,k})$ is given according to equation (4) using cosine similarity and it is called *expectancy* $e_{g,i}$, which is shown in equation (5):

$$p_{u,i} = r_{u,i} + \frac{e_{g,i}}{\bar{e}}, \tag{5}$$

where $\bar{e}$ is the average value of all items' expectancy; $e_{g,i}$ is the expectancy calculated according to the similarity between the popular item group and other items; $r_{u,i}$ is the original evaluation score and $p_{u,i}$ is the evaluation score by considering the expectancy.

4. Performance Evaluation

(a) Evaluation metrics

Performance evaluation is based on equation (6) which compares the extracted list created by existing recommendation system and the extracted list created by our proposed recommendation system:

$$Diversity(L_A, L_B, N) = \frac{|L_A - L_B|}{N}, \tag{6}$$

where L_A means the items with evaluation scores by considering nothing and L_B means the items selected by our proposed recommendation system. In addition, we compare the identity between the two recommendation lists according to equation (7):

$$Identity = \frac{EQ_{A,B}}{N}, \tag{7}$$

where $EQ_{A,B}$ means the same recommendation items between the existing recommendation system.

(b) Experimental results

Figure 1 shows the variety of the recommendation lists by comparing the evaluation score when considering and not considering bandwagon effect. When recommending the top-10 list, the diversity is 0.26623 which shows about 26% difference, and when recommending the top-30 list, the difference is about 21%. It also shows the identity of recommendation lists generated by recommendation system with and without considering bandwagon effect, respectively. When extracting the top-10 data, about 78% identity is shown and when extracting the top-30, 89% identity is shown.

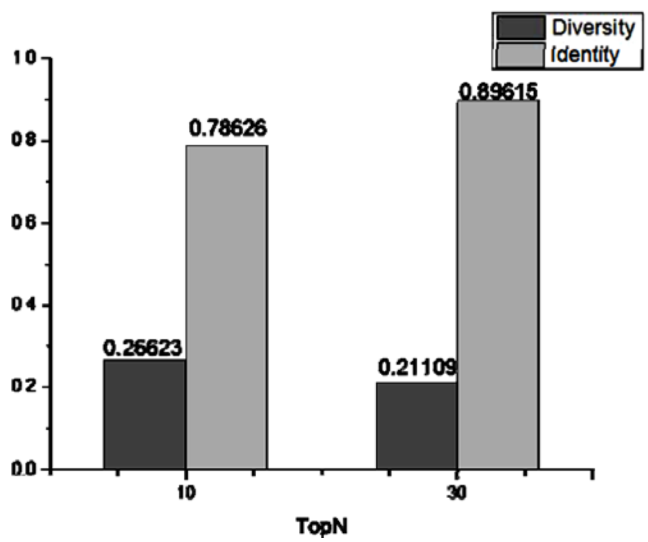

Figure 1. The results of diversity and identity metrics.

5. Conclusions

In this paper, in order to develop the recommendation system taking the bandwagon effect into consideration, we used the MF-based ALS method. In addition, to store and compute big data, we use the spark-based cluster where Hadoop and distributed in-memory processing framework were used. In order for improving the recommendation diversity, we compared the recommendation list from the existing recommendation system and our proposed recommendation system and it showed that our proposed system indicated better diversity during recommendation.

References

[1] B. Lika, K. Kolomvatsos and S. Hadjiefthymiades, Facing the cold start problem in recommender systems, Expert Systems with Applications: An International Journal 41 (2014), 2065-2073.

[2] G. Adomavicius and A. Tuzhilin, Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions, IEEE Transactions on Knowledge and Data Engineering 17 (2005), 734-749.

[3] K. Park and L. Peng, A fast social-user reaction analysis using Hadoop and SPARK platform, International Journal of Applied Engineering Research 11(18) (2016), 9345-9349.

[4] K. Park, C. Baek and L. Peng, A development of streaming big data analysis system using in-memory cluster computing framework: spark, Advanced Multimedia and Ubiquitous Engineering 393 (2016), 157-163.

[5] Y. Koren, R. Bell and C. Volinsky, Matrix factorization techniques for recommender systems, IEEE Computer 42 (2009), 30-37.