# Collaborative Filtering Recommendation with Fluctuations of User' Preference

Jing Yu
*Search and Recommendation Dept. Datagrand Company Product Technology Center*
Shanghai, China
yujing@datagrand.com

Jingjing Shi
*Search and Recommendation Dept. Datagrand Company Product Technology Center*
Shanghai, China
shijingjing@datagrand.com

Yunwen Chen
*Search and Recommendation Dept. Datagrand Company Product Technology Center*
Shanghai, China
chenyunwen@datagrand.com

Daqi Ji
*Search and Recommendation Dept. Datagrand Company Product Technology Center*
Shanghai, China
jidaqi@datagrand.com

Wenhai Liu
*Search and Recommendation Dept. Datagrand Company Product Technology Center*
Shanghai, China
liuwenhai@datagrand.com

Zhijun Xie
*Search and Recommendation Dept. Datagrand Company Product Technology Center*
Shanghai, China
xiezhijun@datagrand.com

Kai Liu
*Search and Recommendation Dept. Datagrand Company Product Technology Center*
Shanghai, China
liukai@datagrand.com

Xue Feng
*Search and Recommendation Dept. Datagrand Company Product Technology Center*
Shanghai, China
fengxue@datagrand.com

2021 IEEE International Conference on Information Communication and Software Engineering (ICICSE) | 978-1-6654-1919-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICICSE52190.2021.9404120

***Abstract*—Traditional recommendation systems (RSs) recommend users to personal items according to their pre-behaviors to infer users' personal preference. Although researchers have made some improvements, which take time factor into traditional recommendation systems, on catching changes of users' interest, it is difficult to always compute the users' preference fluctuations. Therefore, to solve the mentioned above problems, we add personal preference fluctuations into traditional collaborative filtering systems. In this paper, we will quantify user's preference change. Besides, it also adapts recommending in the opposite direction, that is to say, we will recommend items which dissimilar users have rated to the pointed user. The experiment of fluctuation method run on MovieLens 1M and MovieLens latest small, show that the method including changes of personal preference is better than other methods.**

***Keywords—recommendation systems, user's preference, user's flutuation,***

## I. INTRODUCTION

Recently, with the rapid development of information technologies and world web site, we are always trapped in the information overloading [1]. Aiming to solve mentioned problem, an amount of recommender algorithms have been proposed. Recommender algorithms have been widely used in many e-commerce websites, such as amazon, Netflix and so on, which always recommend products according to browsers pre-behaviors. Opposite to traditional tools, such as search engines (which needs to input keywords) and category navigation (where contents are divided into different categories according to common sense), recommender systems automatically shows the arrangement of the products based on browsers' per-behaviors. Although the recommend systems have made great achievements, the current recommender systems require further improvements to make customers more satisfied.

At present, recommender algorithms are always divided into several categories: collaborative filtering (CF) [2]-[4], content-based methods [5], [6], spectral analysis [7], [8]. Most of above methods, CF is most applied in many fields. A more detailed division: collaborative filtering is divided into user-based method (UCF) [2] and object-based method (OCF) [9], in which the difference between the two categories is similarities between users or between objects respectively. What's more, many researchers adopted physical dynamics, such as random walk and heat conduction.

Besides, the K-nearest neighbors (KNN) methods lying on the intuition that users consume on the same items always have the similar performance, these users we called neighbors in recommender systems [10]. The neighbors based on CF contain three steps. We firstly calculate the similarity among users or items by using various similarity measures. Next, we catch the nearest neighbors of users or items according the above calculated similarity. The last step is that we calculate the ratings of pointed user and item according the known ratings of the nearest neighbors. Some traditional similarity measures, such as cosine similarity, Pearson correlation coefficient and hybrid measure, have been widely used in CF to calculate similarity [11].

However, many researchers fail to catch the fluctuation of user's interest, in other words, some users may be changeable who watch different types of movies in a short period, the other may be immutable in the short or the term who always watch the same type of movies. For the moment, few researchers take this influential factor affected users' behaviors into consideration. In this paper, the above-mentioned user's own personality determines the hobby of interest fluctuations over a period is called personal performance fluctuations, referred to as fluctuations. Therefore, in this paper, we will add fluctuations into traditional collaborative filtering systems.

The rest of the paper is organized as followed. In Chapter II, traditional collaborative filtering methods are reviewed. Moreover, the paper introduces new concept fluctuation of recommender systems and how to combine fluctuation with traditional collaborative filtering approach. The detailed process will be explained in Chapter III. In Chapter IV, experimental results and comparison analysis are shown. Finally, our conclusions and future work are drawn in Chapter V.

## II. Related Work

We will introduce this chapter from two aspects: collaborative filtering and fluctuation in economics.

### *A. Collaborative Filtering*

At present, collaborative filtering is the one of the most widely used methods in recommender systems. The most important technique in collaborative filtering is how to calculate the similarity between user and user or item and item. Based on this, many researchers have proposed many kinds of similarity calculation methods

Relying on the traditional user-item scoring matrix, the cosine distance was originally used to calculate the similarity. Later, considering the user's own rating preference (some users gave generally higher ratings and some users generally gave lower ratings), Ahn [12] proposed an adjusted cosine similarity (ACOS). Besides, based on Pearson correlation coefficient (PCC) method, Ekstrand [13] et.al presented compute linear correlation between a pair of objects. Moreover, Constrained Pearson correlation coefficient (CPCC) [14] is a variant of PCC, which adapt frequency to measure the user similarity. Bobadilla et. al [15] proposed JMSD method to comprehensive consider mean squared difference (MSD) and Jaccard, which take full advantage of merits of Jaccard and MSD avoiding the co-rated items. Based on JMSD, Bobadilla [16] et al. proposed a measure by combining the six individual similarity measures, including Jaccard, mean squared difference and other methods.

Besides, in order to handle the common problems on recommender systems: cold-start, Ahn [17] proposed PIP (proximity, impact and popularity), considering the three factor of user self. Based on PIP method, NHSM method proposed by Haifeng Liu [18] not only considers the weakness of PIP method, but also takes local context information into consideration.

Considering sparsity of data existing in many recommender systems, BCF proposed by Bidyu [19] et al, which adapt distribution and Jaccard method mixture into similarity computing. Besides, it takes into account of the degree of association between users and users.

Parivash [20] et.al put forward AC-PCC which introduces new weighting schemes that allows us to consider new features in finding similarities between users.

### *B. Fluctuation in Economic Fields*

As a matter of observation, the growth path followed by the economy involves fluctuations in the extent to which. The problem of economic fluctuations is often discussed under heading of the theory of economic cycles [23]. Besides, it involves a certain period of time factors. Therefore, we introduce the fluctuation into recommender system and consider it under cycles beyond time factors.

## III. Fluctuation Method

In this chapter, we proposed fluctuation method to measure the degree of user preference change. First of all, we will quantify the fluctuation according to common sense, and then we will add the value of fluctuation into collaborative filtering. However, opposite to traditional collaborative filtering recommender systems, it not only contains similarity recommendation, but also contains reverse recommendation.

### *A. Concept of Fluctuation*

Fluctuation is common form of material movement that originally belonged to the field of physics. After a period of evolution, the word gradually used in the economic field, derived from economic fluctuations, which refers to the high or low productivity level of an economic phenomenon. Later many researchers further extend the economic cycle fluctuation theory. Drawing on the concept of fluctuation in economics, the fluctuation in this paper represents the degree of user preference change.

How the fluctuation should be defined in recommender system? To measure the degree of user preference change, we introduce fluctuation which originated from economy and physical. Mentioned fluctuation of recommender systems is aiming to measure the degree of user preference change, that is to say, some users are always changing their preference, while some users insist on one or several types of items, the rest are in the middle of the above-mentioned users.

In order to clearly see the fluctuations in the preferences of each user, we compute the variance of the pointed user watching the cate-movie in different time period. Moreover, it is obviously that the larger the calculated variance is, the value of the fluctuation for each user preference is smaller, on the other hand, the smaller the variance value is, the greater the fluctuation of the user's preference. For the time period how to divide, according to the common sense in life, we form week, month, quarter, etc.at different levels of statistic calculations. At last, we take average value is the variance value, and finally achieve the user's preference fluctuations.

### *B. Modeling of Fluctuation*

TABLE I. NOTATIONS RELATED TO FLUCTUATION

| Notations | Description |
|---|---|
| $T$ | The set of time-partitioned |
| $E$ | The set of which day in a week |
| $W$ | The set of weekly division of time |
| $M$ | The set of monthly division of time |
| $S$ | The set of seasonal division of time |
| $I_u^t$ | The set of items rated by user $u$ at time $t$ |
| $r_{ui}$ | The rating of user $u$ on item $i$ |
| $F_{uj}^t$ | The weighted frequency of $j$ classes rated by user $u$ during time period $t$ |
| $B_{ij}$ | Whether item $i$ belongs to class $j$ (if it belongs to1, otherwise 0) |
| $F_u$ | Fluctuation of preference on user $u$ |
| $v_u^t$ | The variance of all item categories rated by user $u$ during time period |
| $v_u$ | The variance of all item categories rated by user $u$ |

Drawing on the implications of fluctuation in economics, it is necessary to give a certain period of time in order to calculate the fluctuation value on the user's preference. In order to fully consider the analysis of time, we define set $T=\{E,W,M,S\}$ to fully consider the time, where $E$ is a discrete variable, and $W$, $M$, $S$ are continuous time variables. Besides, $E$ is the set of Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday, and W is set of a weekly division of time , $M$ is the monthly division of time, $S$ is the seasonal division of time. In order to better introduce the fluctuation quantification, the common notations of related fluctuation are given in Table I.

Using the method of calculating fluctuation in economics, we consider the season as the maximum dimension of short-term time, the minimum dimension is one day. Besides, considering statistics sense of real life, we map the minimum time dimension to which day in a week. So we divided time into four categories ($E$, $W$, $M$, $S$), according to these categories doing statistics, then calculate the variance in the period of time, and finally calculate the average as variance value for user's preference. The specific formula is calculated as shown in Eq. 123.

$$F_{uj}^{t} = \sum_{i\in I^{t}} B_{ij} r_{ui} \tag{1}$$

$$v_{u}^{t} = \frac{1}{|C|}\sum_{j\in|C|}\Big(F_{uj}^{t} - \frac{1}{|C|}\sum_{j\in|C|} F_{uj}^{t}\Big)^{2} \tag{2}$$

$$v_{u} = \frac{1}{4}\sum_{P\in T}\frac{1}{P}\sum_{t\in P} v_{u}^{t} \tag{3}$$

Intuitively, if the variance of preference based on user $u$ is larger, the fluctuation of preference based on user $u$ is smaller, otherwise the variance of preference based on user $u$ is smaller, then the fluctuation of user $u$ is greater. In other words, $F_u$ and $v_u$ are inversely proportional. The inverse relationship is expressed mathematically as followed:

$$F_u \propto \frac{1}{v_u} \tag{4}$$

In order to better visualize the inverse relationship, we formula the above relationship as followed in Eq.5:

$$F_u = e^{-v_u} \tag{5}$$

Through the formula 1, 2, 3, and 5, we can easily quantify the fluctuations in user preference. Next, we incorporated the fluctuation value into traditional collaborative filtering recommender system, and also combine the reverse recommendation method. At last, fluctuation method can recommend personal items according to their fluctuation of preference on user u.

In order to incorporate quantitative fluctuation into the recommendation system, we now incorporate this variable into traditional collaborative filtering. Figure 1 shows the specific steps of adding fluctuations into traditional collaborative filtering. Different from traditional collaborative filtering in the past, the fluctuation value of each user is different, as a result of which, each weigh on each user is different. Hence, we computed predicted ratings according to the following formula.

$$p_{ui} = \overline{r}_u + F_u \frac{\sum_{v\in U1} \mathrm{cov}(u,v)(r_{vi}-\overline{r}_v)}{\sum_{v\in U1} |\mathrm{cov}(u,v)|} + (1-F_u)\frac{\sum_{v\in U2} \mathrm{cov}(u,v)(r_{vi}-\overline{r}_v)}{\sum_{v\in U2} |\mathrm{cov}(u,v)|} \tag{6}$$

In the above formula, $U_1$ is the most similar collection of 20 users, and $U_2$ is the most dissimilar set of 20 users.

## C. *Process of Modeling of Fluctuation*

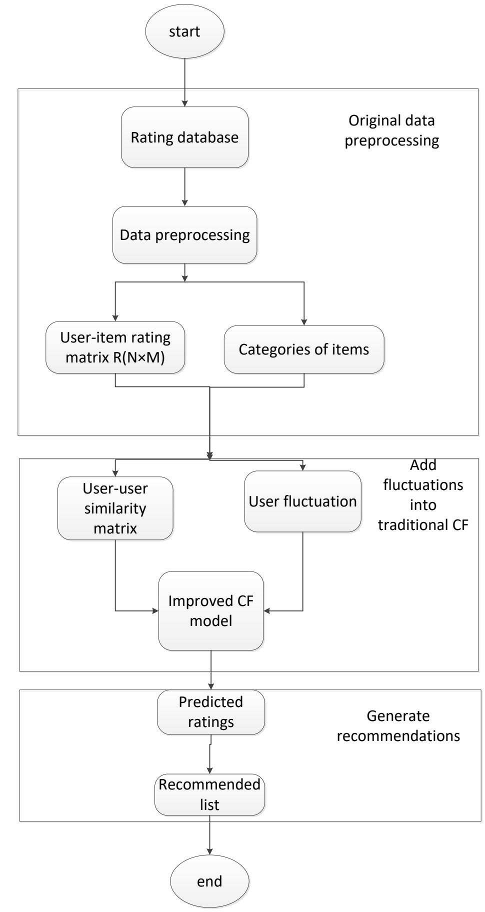


Fig. 1. process of adding fluctuation into traditional CF.

# IV. EXPERIMENTS

To demonstrate the performance of our experimental experiments, we experiments, we run the experiments on the MovieLens dataset, and compare it with the existing methods of collaborative filtering and the method of introducing time with context. We also implement the other CFs based on PCC [13], CPCC [14], PIP [17], MJD [16], JMSD [15], AC-PCC [20] , NHSM [18] and BCF [19].

## A. *Dataset Description*

The data set in this paper is widely used in the field of recommendation system, and comes from the real world, with some authoritative public datasets MovieLens 1M and ML-latest-small. A specific description of the two datasets is given in Table II.

TABLE II. DESCRIPTION OF THE DATASETS USED IN THE EXPERIMENTS.

| **Dataset** | **Purpose** | **#User(M)** | **#Item(N)** | **#Rating(R)** | **Rating domain** |
|---|---|---|---|---|---|
| ML-Latest-Small | MOVIE | 706 | 8570 | 100023 | {1,2,3,4,5} |
| ML-1M | Movie | 6040 | 3952 | 1M | {1,2,3,4,5} |

Each of the datasets is divided into two parts. 80% of ratings from the dataset were randomly selected as the training datasets. The 20% remaining data was used as the testing set.

## B. *Evaluation Methods*

In assessing performance on user fluctuations, we used the evaluated methods commonly used in recommender systems. In this paper, we adapt *MAE* [21], *NMAE* [21] and *F1*-measure [22] to prove that introducing user fluctuation into current recommender system perform better.

*MAE* is commonly used in measuring the accuracy of a recommender system. It is computed the absolute average between real ratings and predicted rating values. It is formula as followed:

$$MAE = \frac{1}{N}\sum_{u=1}^{N}\frac{1}{M}\sum_{i=1}^{M}\left|r_{ui} - p_{ui}\right| \tag{7}$$

*NMAE* is normalized *MAE*, which is formula as followed:

$$NMAE = \frac{1}{\mathrm{N}(r_{\max} - r_{\min})}\sum_{u=1}^{N}\frac{1}{M}\sum_{i=1}^{M}\left|r_{ui} - p_{ui}\right| \tag{8}$$

*F1*-measure is to evaluate the recall rate and precision of the recommended system, so in order to better introduce *F1*-measure, firstly giving formulas of recall and precision.

$$\mathrm{Pr}\,ecision = \frac{1}{N}\sum_{u=1}^{N}\frac{\left|IR_{up} \cap IR_{ua}\right|}{\left|IR_{up}\right|} \tag{9}$$

$$\mathrm{Re}\,call = \frac{1}{N}\sum_{u=1}^{N}\frac{\left|IR_{up} \cap IR_{ua}\right|}{\left|IR_{ua}\right|} \tag{10}$$

$$F1 - measure = \frac{2 \times Precision \times \mathrm{Re}\,call}{\mathrm{Pr}\,ecision + \mathrm{Re}\,call} \tag{11}$$

In the above formula, $IR_{up}$ represents the recommender systems to user recommended list of unrated items according to predicted ratings, and $IR_{ua}$ is the set of recommended list based on real ratings.

## C. *Experimental Results and Analysis*

According to mentioned above, we will compare performance of user fluctuations from *MAE*, *NMAE* and *F1*-measure. In Figure 2, we can conclude that when the more the number of items, the better the performance of the method.

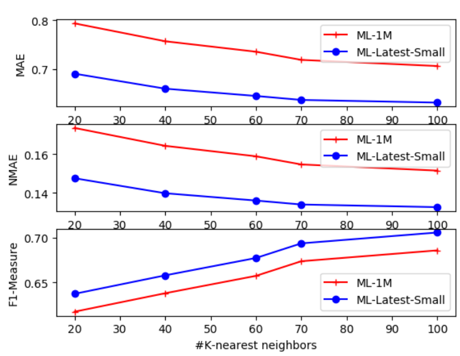


Fig. 2. *MAE*, *NMAE* and *F1*-measure with different K-nearest neighbors on ML-1M and ML-small-latest.

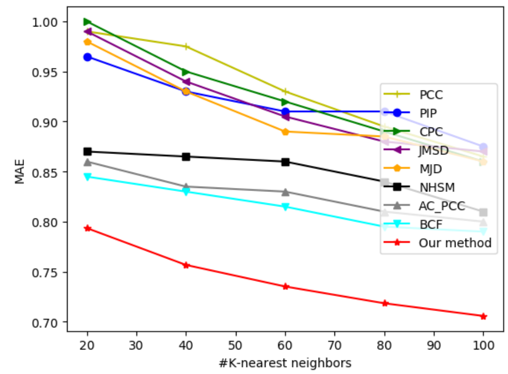


Fig. 3. *MAE* with different K-nearest neighbors on ML-1M.

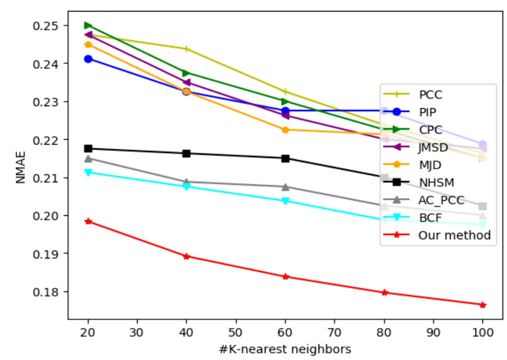


Fig. 4. *NMAE* with different K-nearest neighbors on ML-1M.

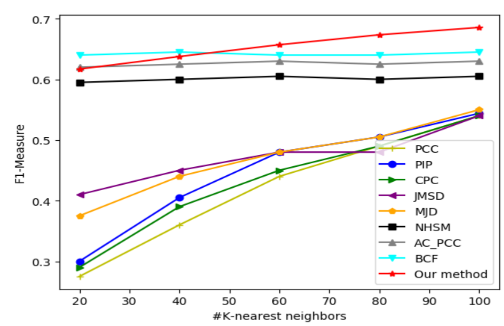


Fig. 5. *F1*-measure with different K-nearest neighbors on ML-1M.

As we all know, the smaller *MAE* and *NMAE*, the better the performance of recommender system; the larger *F1*-measure, the better the performance of the recommender systems. So it is obviously that the performance will be better when the number of items increases. Besides, it also caters to the trend of future's development, the number of users is basically remain unchanged while the number of items will continue to increase.

Next, we will compare various methods based on ML-1M, in Figure 3, 4, 5, we adapt *MAE*, *NMAE* and *F1*-meansure to evaluate the performance of our method.

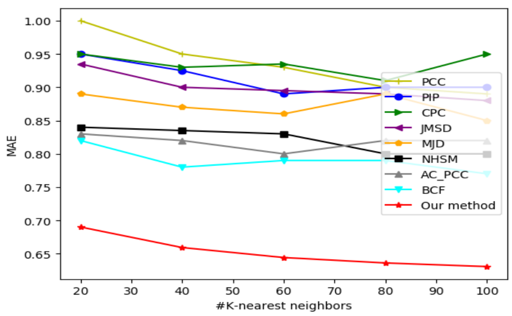


Fig. 6. *MAE* with different K-nearest neighbors on ML-small-latest.

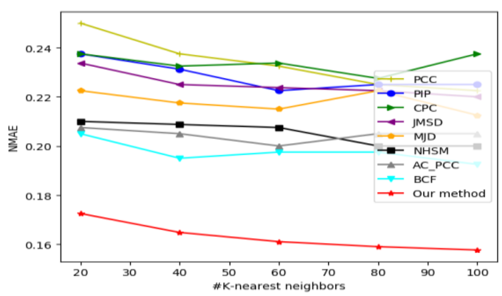


Fig. 7. *NMAE* with different K-nearest neighbors on ML-small-latest.

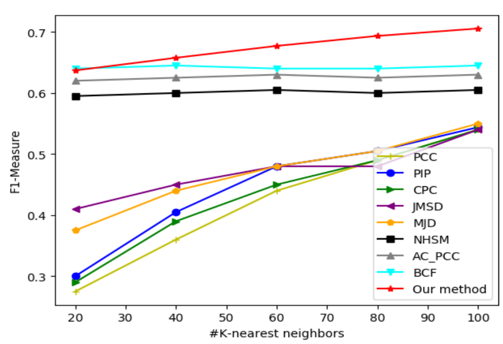


Fig. 8. *F1*-measure with different K-nearest neighbors on ML-small-latest.

From the figures above, we can see that our method performs better than the other methods.

## ACKNOWLEDGMENT

In this paper, we introduce user fluctuations to measure the degree of change in user preference. We count the four time periods and we quantify this user fluctuation value according to common sense. Moreover, we computed predicted rating values by adding quantifying fluctuation user values into traditional collaborative filtering. Experimental evidence shows that adding fluctuations to traditional collaborative filtering not only takes degree of user preference for dynamic changes into consideration, but also future optimizes collaborative filtering. However, we still have some flaws in quantifying user fluctuations. In the future, we will try our best to optimize the quantitative user fluctuations.

## REFERENCES


[1] A. Broder, R. Kumar, F. Moghoul, P. Raghavan, S. Rajagopalan, R. Stata, A.Tomkins. Fab: Large-scale structure of a notion-wide production network . J. Wiener, Comput. Netw. 33 (2000) 309.

[2] D. Goldberg, D. Nichols, B.M. Oki, D. Terry. Fab: Using collaborative filtering to weave an information tapestry. Commun . ACM 35 (1992) 61.

[3] J.L. Herlocker, J.A. Konstan, A. Borchers, J. Riedl.J: An algorithmic framework for performing collaborative filtering..Twenty-Second Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Berkeley, 1999, p. 230.

[4] J. B. Schafer, D. Frankowski, J. Herlocker, S. Sen, Collaborative filtering recommender systems, Lect. Notes Comput. Sci. 4321, (2007) 29

[5] M. Balabanovi, Y. Shoham, Fab: content-based, collaborative recommendation. Commun. ACM,40(3):66-72,1997.

[6] Pazzani, M.J.: A framework for collaborative, content-based and demographic filtering. Artif. Intell. Rev. 13(5–6), 393–408 (1999).

[7] Maslov, S., Zhang, Y.C.: Extracting Hidden Information from Knowledge Networks. Phys. Rev. Lett. 87, 248701 (2001)

[8] J.Ren, T.Zhou, and Y.-C. Zhang, InformationFiltering via Self-Consistent Refinement, EPL 82, 58007 (2008).

[9] B. Sarwar, G. Karypis, J. Konstan and J. Riedl, "Item-Based Collaborative Filtering Recommendation Algorithms," WWW '01: Proceedings of the 10th International Conference on World Wide Web, New York, 1-5 May 2001, pp. 285-295.

[10] B.K. Patra , R. Launonen , V. Ollikainen , S. Nandi , Exploiting Bhattacharyya similarity measure to diminish user cold-start problem in sparse data, in: Proceedings 17th International Conference Discovery Science (DS 2014), 2014, pp. 252–263 .

[11] C. Desrosiers , G. Karypis , A comprehensive survey of neighborhood-based recommendation methods, in: Recommender Systems Handbook, 2011,pp. 107–144

[12] H.J Ahn , A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem, Inf. Sci. 178 (1) (2008) 37–51 .

[13] M.D. Ekstrand , J.T. Riedl , J.A. Konstan , Collaborative filtering recommender systems, Found. Trends Hum.–Comput. Interact 4 (2) (2011) 81–173 .

[14] U. Shardanand , P. Maes , Social information filtering: algorithms for automating ''Word of Mouth'', in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1995, pp. 210–217 .

[15] J. Bobadilla , F. Serradilla , J. Bernal , A new collaborative filtering metric that improves the behavior of recommender systems, Knowl. Based Syst. 23 (6) (2010) 520–528 .

[16] J. Bobadilla , F. Ortega , A. Hernando , J. Bernal , A collaborative filtering approach to mitigate the new user cold start problem, Knowl. Based Syst. 26(2012) 225–238 .

[17] H.J Ahn , A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem, Inf. Sci. 178 (1) (2008) 37–51 .

[18] H. Liu , Z. Hu , A. Mian , H. Tian , X. Zhu , A new user similarity model to improve the accuracy of collaborative filtering, Knowl. Based Syst. 56 (2014) 156–166 .

[19] B.K. Patra , R. Launonen , V. Ollikainen , et al. , A new similarity measure using Bhattacharyya coefficient for collaborative filtering in sparse data, Knowl. Based Syst. 82 (2015) 163–177 .

[20] P. Pirasteh , D. Hwang , J. Jung , Weighted similarity schemes for high scalability in user-based collaborative filtering, Mobile Netw. Appl. 20 (2015)497–507 .

[21] X. Su, T.M. Khoshgoftaar, A survey of collaborative filtering techniques, Adv. Artif. Intell. 2009 (2009) 1–19.

[22] L. Jin, A Bayesian approach toward active learning for collaborative filtering, Presented at the Proceeding UAI '04 Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence (2004) 278–285.

[23] D.C.Rowan , Fluctuations in Economic Activity, in :output, inflation and Growth (1983)358-381.

[24] Beaudry, P., Portier, F. Stock prices, news, and economic fluctuations. American Economic Review 94, (2006)1293–1307.