Research on Rating Error and Quality Metrics for Collaborative Filtering Recommendation Methods

Kun Zhao\(^1\) and Jiaming Pi\(^2,3,*\)

\(^1\) Information School, Yunnan University of Finance and Economics, Kunming, China
\(^2\) Business School, Yunnan University of Finance and Economics, Kunming, China
\(^3\) Zhonghua Vocational College, Yunnan University of Finance and Economics, Kunming, China

*Corresponding author e-mail: 277352342@qq.com

Abstract. The collaborative filtering recommendation system has been widely used in E-commerce as a relatively successful recommendation system. At present, the focus of collaborative filtering recommendation research is mainly on how to improve the accuracy of recommendation by improving the recommendation algorithm. However, in real world, the user's rating behaviour is not perfectly rational. It is no odd that there is deviation of rating to any a given item for a user in real evaluation. In this case, what it means for the improvement of collaborative filtering recommendation methods, and how they performed when we use the commonly used quality metrics to evaluate the collaborative filtering recommendation methods? In view of these problems, this paper presupposes that the user's rating behaviour is a bounded rational behaviour, and based on the normality hypothesis of rating error, introduces a simulated rating experiment innovatively to analyse the effect that rating prediction can achieve in sense of the commonly used quality metrics. This study is of significance for the research and application of collaborative filtering recommendation technology.

1. Introduction

Collaborative Filtering Recommendation (CFR) is one of the most successful and widely used personalized recommendation techniques [1], which is a hot issue in the field of personalized recommendation. The concept of collaborative filtering is firstly proposed by David Goldberg and etc. in 1992 when a mail filtering system called Tapestry [2] was developed. Since then, many CFR systems have been developed and used in various websites such as Amazon [3], Google [4] and Yahoo [5].

One of the research focuses of CFR is to increase the accuracy and effectiveness of recommendation result by improving the recommendation algorithm. For this purpose, many research energy have been exerted and a variety techniques have been applied, such as using matrix reducing technique [6-8], improving similarity calculation models [9-10], improving nearest neighbour selection methods [11], populating user-item rating matrix [12-14], establishing user reliability model [15-16], using social network analysis (SNA) techniques [17], and etc.

However, in the actual use of the recommendation system, there is a fact that the user's rating behaviour is not perfectly rational. For example, it is no odd that a user may rate an item with 4 at a time and 3, 5 or other score at another time due to the "rational constraints" of the user. Therefore, there is no way to ensure that each rating prediction for the user is completely accurate, no matter whatever the CFR algorithms are used and how they are "accurate". For this concern, a reasonable expectation for the prediction is that: the deviation of user's rating shows a distribution pattern with certain characteristics under the condition that user’ rating is a sort of bounded-rationality behaviour.
Therefore, based on the bounded rationality assumption for users in their rating behaviours, we further suppose that the rating deviation follows the normal distribution, and then, we innovatively introduced a simulated rating experiment to study the following two issues: (1) how far we can go in the sense of improving the effect of CFR methods, what aspects make main blocks for the improvement, and what are the goals and focus for the improvement? (2) What kind of effect can be anticipated by using the commonly used quality metrics to evaluate the CFR methods? Such as absolute mean error (MAE), mean square error (MSE), Precision, Recall, and Coverage [18]. The study on these issues is of significance to the research and application of CFR techniques.

2. Overview of Related Theory
This section gives a brief description about the normal distribution and commonly used quality metrics in the evaluation of CFR methods.

2.1. Characteristics of Normal Distribution
The normal distribution has an extremely wide range of practical backgrounds. A large number of phenomena in nature and human society present their normal distribution characteristics. For example, in the medical field, some medical phenomena, such as the height of Human beings, the number of red blood cells, etc., all appear normal or approximately normal distribution. In the production field, under the same production conditions, the calibre, length, etc. of the product appear normal distribution. In the field of scientific experiments, some observations and experimental results, such as the weight of the same object, the velocity component of the ideal gas molecule, etc., all appear as normal or approximately normal distribution. Since the rating deviation generated by the user’s irrational behaviours is also random, this paper assumes that the rating deviation follows a normal distribution.

Normal distribution is a kind of probability distribution of continuous random variables, it is also known as Gaussian distribution. In general, normal distribution describes the characteristics that a random variable X (X is a one-dimensional variable) obeys a probability distribution with an expected value of μ and a standard deviation of σ, and its probability density function is as formula (1).

\[
f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)
\]

The general shape of a normal curve is shown in figure 1. It is a bell-shaped curve with its apex in the middle, symmetrically slowing down on both sides.

-6σ -2σ -1σ μ +1σ +2σ +3σ

The probability at different intervals is also shown in the diagram. 

Figure 1. Normal distribution diagram.
2.1.1. Curve Characteristics. As shown in Figure 1, the curve of normal distribution has characteristics of concentration, symmetry, and uniform variability. It presents different patterns depending on the parameters μ and σ.

μ is positional parameter that indicates the point where the distribution will be concentrated on. The probability law is that the peak of the normal curve is located in the centre, the greater the probability of those points closer to μ, and smaller of those far from μ. On both sides of X=μ, the curve of a normal distribution are completely symmetrical. The expectation, mean, median, and mode of a normal distribution are all the same, and equal to μ.

σ describes the data dispersion degree of a normal distribution. The greater the σ, the more dispersed the data distribution, and the smaller the σ, the more concentrated. σ is also called the shape parameter of normal distribution. The greater σ is, the flatter the curve is. Conversely, the smaller σ is, the higher the curve.

Starting from the point where the mean value indicates, the curve of a normal distribution gradually decreases to the left and right sides. The two ends of the curve never intersect with the horizontal axis, and its probability will be of 0 at positive and negative infinity. The curve shows its two inflection points at μ±σ.

2.1.2. Area characteristics. As shown in Fig. 1, the area between the horizontal axis and the normal curve within a certain interval indicates the probability which the value of the variable falls within the interval (probability distribution). The total area between the curve and the horizontal axis is equal to 1, corresponding to the integral value of the probability density function from positive infinity to negative infinity; this means the sum of the frequencies is 100%.

In each interval of the distribution normal distribution has the following characteristics:

1) The area within the horizontal axis interval (μ-σ, μ+σ) is 68.268949%, that is, 
P(|X-μ|<σ)=2Φ(1)-1=0.6826;

2) The area within the horizontal axis interval (μ-1.96σ, μ+1.96σ) is 95.449974%, that is, 
P(|X-μ|<2σ)=2Φ(2)-1=0.9544;

3) The area within the horizontal axis interval (μ-2.58σ, μ+2.58σ) is 99.730020%, that is, 
P(|X-μ|<3σ)=2Φ(3)-1=0.9974.

In short, the above characteristics show that in a normal distribution, the ratio within one standard deviation is about 68% of the total value, the ratios within the two standard deviations is about 95%, and the ratio within the three standard deviations is about 99%. On the contrary, in practical applications, if the assumption on a random variable that obeys a normal distribution is correct, the area characteristics of the distribution should be consistent with these characteristics. Therefore, these features are usually summarized as "68-95-99.7 rule" or "rule of thumb" [19].

In addition, in practical applications, the interval (μ-3σ, μ+3σ) is basically regarded as the actual possible value interval of the random variable X, that is, the so-called "3σ" principle of normal distribution. It is generally considered that a small probability event is almost impossible to occur in one experiment. The so-called "small probability event" refers to an event in which the probability of occurrence is less than 5%. It can be seen that the probability that X falls outside (μ-3σ, μ+3σ) is less than three thousandths. Therefore, it is considered that the corresponding event would not occur in real world.

2.2. Quality Metrics of CFR

In the examination of recommending techniques, it is general practice to separate the experimental samples into training set and test set. The training set is used to generate or establish relevant information and models required for the rating prediction, such as the rating expectation and similarity between each items. The test set is used to predict the score of the item which is actually evaluated by the user, and then we measure the quality of the technique in concern by using metrics such as MAE, MSE, Precision, Recall, F1, Coverage and others.
MAE (Mean Absolute Error) is used to measure the total deviation of the predicted score from the actual score in the test sample. The smaller the value, the higher the accuracy of the algorithm, where $N$ is the number of predicted items in the system, $p_i$ and $q_i$ are the predicted score and the actual scores separately.

$$MAE = \frac{\sum_{i=1}^{N} |p_i - q_i|}{N}$$ (2)

MSE (Mean Squared Error) measures the total deviation of the predicted rating from the actual rating in the test sample by the square of each paired rating error. The smaller the value, the higher the accuracy of the algorithm is. It is more sensitive than MAE for the evaluation of the prediction quality. MSE is computed by formula (3), where $N$ is the number of predicted items in the system, $p_i$ and $q_i$ are the predicted score and the actual scores separately.

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (p_i - q_i)^2$$ (3)

In the TOP-N-based recommendation system, Precision and Recall are usually used to comprehensively measure the accuracy of recommendation. In the calculation of these two metrics, the item set is usually divided into two categories: positive and negative according to the actual rating and the predicted rating according by a certain threshold $\eta$. The items with scores greater than $\eta$ belong to Positive set, and those smaller than $\eta$ belong to Negative set. Precision is used to measure the prediction accuracy of the items whose prediction rating are Positive while Recall is used to measure prediction accuracy of the items whose actual rating are Positive. The larger the value, the better the performance of the algorithm is.

Precision is computed as the ratio of the user’s preferred items in the recommendation list to all recommended items, which is shown in formula (4). The larger the value, the better the recommendation algorithm works.

$$R = \frac{N_{r_u}}{N_s}$$ (4)

Recall is computed as the ratio of user’s preferred items in the recommendation list to all users’ preferred items in the system, which is shown in formula (5). The larger the value, the better the recommendation algorithm works.

$$R = \frac{N_{r_u}}{N_r}$$ (5)

Where $N_{r_u}$ represents the number of user’s preferred items in the recommendation list, $N_s$ represents the number of all recommended items, and $N_r$ represents the number of all users’ preferred items.

In our experiments, because the items in MovieLens dataset are with higher scores, we take 4 as the thresholds to distinguish between Positive and Negative, that is, $\eta=4$, and the maximum number for recommending list is TOP-N=10.

When using Precision and Recall to test the quality of predictive rating, they would usually conflict mutually, that is, one is high while the other is low, which makes it difficult to compare the prediction quality among different methods. In order to solve this problem, the $F_\beta$ measure is used [20].

In general, $F_\beta$ measure is defined in formula (6)

$$F_\beta = (1 + \beta^2) \times \frac{\text{Precision} \times \text{Recall}}{(\beta^2 \times \text{Precision}) + \text{Recall}}$$ (6)

In experiment, we usually use $F_1$ measure, that is, we set $\beta=1$ to take Precision and Recall as of equal importance.
In general, Coverage is the ratio of recommended items to total items. In the algorithm evaluation, the recommendation system generates a recommending list \( LP(u) \) of length \( N \) for each user \( u \in U_p \) (recommended according to the predicted rating), where \( U_p \) is the user set and \( I_p \) is the item set. The Coverage of the system is computed by formula equation (7):

\[
Coverage = \frac{\text{card}(\bigcup_{u \in U_p} LP(u))}{\text{card}(I_p)}
\]  

In the formula, \( \text{card}() \) is the function used to computes the number of items in a set. It can be inferred from the formula that Coverage may be low for systems based on user ratings, because recommendations are made depending on the user's actual preference to the items. Items that are not preferred by users would not be recommended that makes Coverage less than 100%. Of course, it is preferable for a sound recommendation system both to meet user satisfaction in high degree and of high Coverage.

3. Experiment Design

3.1. Purposes and Steps of Experiment

We assume that the user's rating may deviate from its “true value” to some degree and that the deviation obeys the normal distribution. Based on the assumption, we further examine how accuracy a rating prediction can be made in different deviation cases and how these quality metrics perform, so as to understand how rating deviation would put their influence on the predicting accuracy and the performance of the quality metrics. We believe this understanding is of helpful in lying down a reliable foundation for CFR research. For this reason, we design three experiment schemes, free normal error, controlled normal error, and contingent normal error. In our experiments, we used MovieLens-1M dataset as the experimental dataset, which uses a 5-rating scale. The experimental steps are as follows:

1. Based on the three experiment schemes, generate the corresponding "predicted rating" separately for each user in the experimental sample according to their actual score.
2. Use MAE, MSE, Precision, Recall, and Coverage to test the accuracy of the "prediction".

It is interesting to note that most CFR methods producing predicting values with decimal based on some sort of a weighted formula though the real ratings are integer values in MovieLens database. To consistent with this situation, our experiments also produce predicting values with decimal so as to make our experiments of practical significance.

3.2. Experiment I: Free Normal Error

The purpose of our experiments is to exam what effect the rating prediction may produce under the influence of some irrational random rating behaviour in the actual rating situation.

Our experiments generate "predicted rating" obeying \( N(\mu, \sigma^2) \) which \( \sigma \) is set to a suitable value and use each actual score as the expected value \( \mu \). The specific steps as follows:

Step1: Classified the experimental sample according to its actual rating, and stored them in arrays \( A_1 \sim A_5 \) respectively. For example, if the rating is 1, it is stored in \( A_1 \), the rating 2 is stored in \( A_2 \), ..., and the rating 5 is stored in \( A_5 \).

Step2: The distribution for each integer rating in \([1, 5]\), except for the distribution at the point where \( \mu=3 \) (midpoint of the interval), the others are not normal distributed in the strict sense, especially at the two ends of the interval, which are actually normal unilateral distributed. Therefore, we treat the predicting rating actually as follows.

1. Generate an error array obeying the normal distribution \( N(0, \sigma^2) \) for each of the rating arrays \( A_1 \sim A_5 \) respectively, and denote them by \( E_1 \sim E_5 \). For example, \( E_1(i) \) is the "predicting error" of actual rating \( A_1(i) \), and the positive and negative errors contained therein are substantially symmetrically distributed.

2. For each rating array, according to their position of the rating in the interval, the following processes are respectively performed:
For the item with an actual rating of 1, we take "1+absolute value of error" as the “predicting rating” of the actual rating A1(i) and store them B1.

For the item with an actual rating of 5, we take "5-absolute value of error" as the "predicting rating" of the actual rating A5(i) and store them in B5. E.g., B2(i)=2+abs(E2(i)); for rating 4, if B4(i)>5, then take B4(i)=4-abs(E4(i)). The function abs() is the absolute value of the element.

We perform two experiments with standard deviations σ=0.8 and σ=1.0, respectively.

3.3. Experiment II: Controlled Normal Error
This experiment simulates the situation that excludes irregularly randomness in scoring, i.e., the scoring bias is only caused randomly under more rigor condition. The specific approach is as follows:

Step1: Classify the sample and store them in A1~A5 respectively as in the Step1 of section 3.2.

Step2: For each rating, generate a rating error that obeys the standard normal distribution N(0,1) and store it in its corresponding array E1~E5.

Step3: Normalize the error arrays E1~E5 by using Ei=Ei/max(abs(Ei)). Because the distribution interval of the normal distribution is (-∞, +∞) in theory. According to the characteristics (σ = 1.0) of the standard normal distribution, the value exceeding 1 accounts for about 32% of the total rating. For the easiness of the processing, we normalize it to [0,1]. Considering that there is still about 1% of the rating error value more than 3, in order to avoid a large number of large rating error values having a greater impact on normalization, according to the principle of "small probability event", we set the value of the error that is more than 3 to 3.

Step4: Use a parameter K to control the scope of rating error, and according to the restriction for each rating value, generate the "predicting rating" as follows.

For the item with an actual rating of 1, use B1=1+K*abs(E1) to compute the "predicting rating" and store them in array B1.

For the item with an actual rating of 5, use B5=5-K*abs(E5)to compute the "predicting rating" and store them B5.

For the rating of 2~4, use B2=2+K*E2, B3=3+K*E3, B4=4+K*E4 to generate their corresponding “predicting rating”, and store them in arrays B2~B4 respectively.

We performed three experiments with K=1, 2, and 3, respectively. It is worth to note that the error distribution for rating 2 and 4 are not symmetry in the sense of normal distribution through above processing. However, this would not make much influence on the rating errors; therefore, we do not take further treatment to them.

3.4. Experiment III: Contingent normal error
In the rating prediction, there is usually such a "rating on contingency" phenomenon in which the error of items may be higher for items with lower ratings than those with higher ratings. The purpose of this experiment is to analyse what effect the score prediction can achieve in this contingent rating situation.

The process for generating the predicting rating is the same as in experiment I. However, we make a contingency treatment for each of the ratings from 1 to 5, that is, take different standard deviations σ to generate normal distribution errors according to the actual rating. We have performed three experiments. The standard deviation of rating 1~5 in each experiment are σ1=[1.4,1.0,0.8,0.6,0.6], σ2=[1.8,1.4,1.0,0.6,0.6 ] and σ3 = [2.0, 1.5, 1.0, 0.6, 0.6]. Our purpose is to observe how the rating error on the lower ratings (rating 1~2) influence the quality metrics under condition that the standard deviation for the higher rating (rating 3~5) are not changed. We denote the above three sets of σ1, σ2, and σ3 values as Σ=1, Σ=2, and Σ=3, respectively.
4. Experiment Result and Discussions

We implement the above three experiments on MovieLens-1M database. Tables 1−2 give the calculation results of the metrics for each of the three experiments. Table 1 is the MAE and MSE on each rating, and Table 2 summarizes the six metrics values obtained in each experiment, i.e. MAE, MSE, Precision, Recall, F1 and Coverage.

### Table 1. The MAE and MSE on each rating

| Experiment | Parameter Values | MAE   | MSE   |   |   |   |   |   |   |   |
|------------|------------------|-------|-------|---|---|---|---|---|---|---|
|            |                  | 1     | 2     | 3 | 4 | 5 | 1  | 2  | 3  | 4  | 5  |
| Exp. I(σ)  | 0.8              | 0.629 | 0.638 | 0.634| 0.641| 0.637| 0.624| 0.634| 0.625| 0.644| 0.637 |
|            | 1.0              | 0.796 | 0.797 | 0.783| 0.797| 0.792| 0.991| 0.999| 0.923| 0.994| 0.990 |
| Exp. II(K) | 1.0              | 0.265 | 0.262 | 0.265| 0.263| 0.263| 0.109| 0.107| 0.109| 0.108| 0.107 |
|            | 2.0              | 0.514 | 0.515 | 0.532| 0.509| 0.524| 0.415| 0.394| 0.438| 0.387| 0.426 |
|            | 3.0              | 0.793 | 0.711 | 0.779| 0.712| 0.791| 0.971| 0.748| 0.913| 0.745| 0.972 |
| Exp. III(Σ)| 1                | 1.100 | 0.797 | 0.634| 0.481| 0.478| 1.897| 0.985| 0.625| 0.362| 0.358 |
|            | 2                | 1.426 | 1.107 | 0.780| 0.482| 0.478| 3.110| 1.863| 0.920| 0.365| 0.358 |
|            | 3                | 1.560 | 1.169 | 0.783| 0.479| 0.475| 3.673| 2.068| 0.923| 0.360| 0.357 |

### Table 2. Summary of the Quality Metrics (TOP-N=10, η=4)

| Experiment | Parameter Values | MAE   | MSE   | Precision (%) | Recall (%) | F1 (%) | Coverage (%) |
|------------|------------------|-------|-------|---------------|------------|--------|--------------|
|            |                  |       |       |               |            |        |              |
| Exp. I(σ)  | 0.8              | 0.637 | 0.635 | 61.05         | 63.37      | 62.19  | 88.99        |
|            | 1.0              | 0.792 | 0.975 | 51.52         | 59.18      | 55.09  | 89.58        |
| Exp. II(K) | 1.0              | 0.264 | 0.108 | 80.12         | 70.30      | 74.89  | 85.83        |
|            | 2.0              | 0.519 | 0.412 | 67.19         | 66.98      | 67.09  | 87.29        |
|            | 3.0              | 0.752 | 0.853 | 53.70         | 62.10      | 57.59  | 89.17        |
| Exp. III(Σ)| 1                | 0.589 | 0.584 | 58.31         | 65.94      | 61.89  | 88.43        |
|            | 2                | 0.680 | 0.824 | 45.58         | 61.94      | 52.52  | 89.58        |
|            | 3                | 0.693 | 0.877 | 43.56         | 61.41      | 50.97  | 90.13        |

Taking the user’s rating behaviour into account that is of the nature being bounded rationality, it is impossible for the predicting accurate achieve the accurate of 100%. As regard to this, we naturally want to know how far we can go in attempting to increase the predicting accurate by improving CFR methods. However, this question is hard to answer. Therefore, our experiments have significant in providing some sort of yardsticks to measure the quality of rating predictions. From the above experimental results, we can make observation as follows.

1. In Experiment II, when k=1, the rating prediction obtains excellent effect with MAE=0.264, MSE=0.108, F1=74.89%, which is difficult to achieve in real rating predictions. This result can be used as a benchmark to exam how well a CFR method performed in real applications. In Experiment II, when K=2 and 3, sound MAE and MSE values are also obtained, but F1 is decreased to 67.09% and 57.59%, respectively. In the two experiments of Experiment I, F1 is 62.19% and 55.09%, respectively. In the three experiments of Experiment III, F1 is 61.89%, 52.55% and 50.97%, respectively. Though their MAE and MSE are all reach a reasonable level in these experiments, the values of Precision, Recall and F1 are not as good as expected.

2. There is contradiction between Precision and Recall. In some experiments, Precision is higher than Recall, while others are reversed. In Experiment III when Σ=3, the difference between the two is the largest, which is about 16%. In Experiment II, the Precision is higher than Recall when K value is
1.0, and the Precision is lower than Recall when K value is 3.0. In the three experiments in Experiment III, the values of Precision are all lower than Recall. Therefore, it is preferable to use $F_1$ to analyse the comprehensive performance of the two.

(3) As to the aspect of quality metrics, as pointed out in many studies, the performance of a recommending method cannot be accurately reflected by using only one metric. It can be seen from the above experiments that even if some experiments can get better MAE and MSE, the $F_1$ value does not necessarily keep the same trend. For example, in Experiment III when take $\Sigma=2$, the MAE value is 0.680 and MSE value is 0.824, they are both better than Experiment II when take $K=3.0$, which the MAE value is 0.752 and the value MSE is 0.853, but the $F_1$ value of the former is worse than the latter (52.52% VS 57.59%). Therefore, it is desirable to conduct comprehensively examination on the performance of CFR method by using multiple metrics.

(4) The contingent rating phenomenon is an important factor that affects the accuracy of prediction. In the three experiments of Experiment III, which the error distribution of the higher ratings keep the same, but the errors on the lower ratings have a large influence on MAE and MSE. Although the experiment does not prove the authenticity of the contingent rating phenomenon, in real rating activities, the phenomenon does exist that we pay more attention to the valuable items while deliver a lower rating to items with less value at will. Therefore, our experiments on contingent rating phenomenon are of realistic foundation. At the same time, the experimental results show the importance of improving the predicting accuracy on lower ratings.

(5) In the above experiments, the values of Coverage are all above 85%. The lowest Coverage is 85.83% which is in Experiment II when takes $K=1.0$ (i.e. the maximum rating deviation is 1). The highest Coverage is 90.13% which is in Experiment III when takes $\Sigma=3$ (the rating deviations on the lower ratings are greater than those of higher ratings). All experiments present such an interesting phenomenon: the larger the range of the rating deviation, the higher the coverage. Therefore, if we want to use Coverage to measure the accuracy of the recommendation, it is not reasonable enough to some extent. In fact, for the experimental sample we used, even if the prediction error is 0, the Coverage is only 83.61%.

5. Conclusion
Based on the premise of bounded rationality of users in rating and the hypothesis that the rating error obeys the normal distribution, we have designed three experiment schemes to simulate how the predicting errors would present, free normal error, controlled normal error, and contingent normal error. We use the common quality metrics in the examination of CFR methods to evaluate the simulation results. These results can be served as some sort of yardsticks to measure the quality of rating prediction in real applications.

From our experiments we suggest that in case of predicting error being inevitable, it is important to pay attention to the nature of user’s rating behaviour in CFR methods research and it is preferable to base the research on a reasonable foundation about user’s rating behaviour. Therefore, we call for researches of this kind. Though our work has shortcomings in some way, for example, the experiments we have designed may be not well mirror the real situations that may occur in user’s rating activities, it initial this attempt and have achieved some useful conclusions that throw light on CFR method research.

6. Acknowledgments
This work was financially supported by the National Natural Science Foundation Program of China [71462036].

7. References
[1] HERLOCKER J,KONSTAN J, TERVEEN L and et al2004Evaluating collaborative filtering recommend systems.J. ACM Transon Information Systems.22(1)5-53.
[2] Goldberg D, Nichols D and D. Terry1992Using collaborative filtering to weave an information tapestry. J. Commun.ACM. vol. 35 61-70.
[3] Linden G, Smith B and York J 2003 Amazon.com recommendations: Item-to-item collaborative filtering. *J. IEEE Internet Computing*. 7(1) 76–80.

[4] Das A, Datar M and Garg A, 2007 Google news personalization: Scalable online collaborative filtering. *C. Proceeding of the WWW 2007/ T rack: Industrial Practice and Experience. Banff, Alberta, Canada*. 271–280.

[5] Park S, and Pennock D. 2007 Applying collaborative filtering techniques to movie search for better ranking and browsing. *C. Proceeding of the 13thAssociation for Computing Machinery Special Interest Group on Knowledge Disco very in Data. San Jo se, California, USA*. 550–559.

[6] XinGuan, Chang-TsunLi and YuGuan 2017 Matrix Factorization With Rating Completion: An Enhanced SVD Model for Collaborative Filtering Recommender Systems. *J. IEEE Access*. 27668-27678.

[7] QianPeng 2017 Collaborative filtering recommendation based on k-SVD. *D. Beijing University of Technolog.*

[8] Lin Ziyang 2017 Research and design of collaborative filtering recommendation engine based on similarity modeling and SVD optimization. *D. Xidian University.*

[9] Huete J F, Fernández-Luna J M, de Campos L M and et al 2012 Using past-prediction accuracy in recommender systems. *J. Information Sciences*. 199 78-92.

[10] Liu Yi, Feng Wei, WeiTongtong, Chen Zhifei, XuHuan and Zhang Lixia 2017 An Improved Collaborative Filtering Recommendation Algorithm. *J. Computer and Modernization. (1)* 1-4, 12.

[11] Zhang Jinglong, Huang Mengxing, Zhang Yu and Wu Qingzhou 2018 Collaborative filtering recommendation algorithm based on tag optimization. *J. Journal of Computer Applications*. (10) 2916-2919.

[12] Jiang Zongli, Wang Wei and Lu Chen 2017 Improvement of Collaborative Filtering Recommendation Algorithm Based on Mean Estimation. *J. Computer Technology and Development*. (5) 1-5.

[13] Zhang Weimin, Li Yulu and Li Yongli 2017 Collaborative Filtering Recommendation Algorithm Based on Social Relationship and Conditional Completion. *J. Journal of Jilin University(Science Edition)*. (5) 1244-48.

[14] Song Meiqing 2016 A collaborative filtering recommendation algorithm based on multidimensional data filling. *C. Proc of the 2nd IEEE International Conference on Computer and Communications*. 175-179.

[15] Li Yuchen, Chen Li, Shi Chenchen and Lan Xiaoyan 2018 Enhanced collaborative filtering algorithm adopting trust network. *J. Computer Application Research*. (1) 116-120.

[16] Li Liang, Dong Yuxin, Zhao Chunhui and Cheng Weijie 2017 Collaborative Filtering Recommendation Algorithm Based on User Trust. *J. Microcomputer Systems*. (5) 951-955.

[17] Kim 2014 Minsung; Im, Il. Resolving the ‘Gray sheep’ Problem Using Social Network Analysis (SNA) in Collaborative Filtering (CF) Recommender Systems. *J. Journal of Intelligence and Information Systems*. Vol. 20 (No. 2) 137-148.

[18] Suganeshwari G and Ibrahim Syed S P 2016 A Survey on Collaborative Filtering Based Recommendation System. *J. Proceedings of the 3rd International Symposium on Big Data and Cloud Computing Challenges (ISBCC–16’).* 503-518.

[19] Liang Changhong, Li Long, Shi Xiaowei and et al 2003 Concise closed form of standard normal distribution. *J. Journal of Xidian University(Natural Science Edition)*. 30(3) 289-292.

[20] Sun Chen, Xi Hongsheng and Gao Rong 2015 A Recommendation-Support Model Using Neighborhood-Based Linear Least Squares Fitting. *J. Journal of Xi’an Jiaotong University*. 49(6) 77-83.