IOP Conf. Series: Materials Science and Engineering **324** (2018) 012049 doi:10.1088/1757-899X/324/1/012049

Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model

Weijie Wang [1] and Yanmin Lu [1]

1 CFETS Information Technology (Shanghai) Co., Ltd. China Foreign Exchange Trade System, Shanghai, China
E-mail: wangweijie@chinamoney.com.cn

Abstract. Most existing Collaborative Filtering (CF) algorithms predict a rating as the preference of an active user toward a given item, which is always a decimal fraction. Meanwhile, the actual ratings in most data sets are integers. In this paper, we discuss and demonstrate why rounding can bring different influences to these two metrics; prove that rounding is necessary in post-processing of the predicted ratings, eliminate of model prediction bias, improving the accuracy of the prediction. In addition, we also propose two new rounding approaches based on the predicted rating probability distribution, which can be used to round the predicted rating to an optimal integer rating, and get better prediction accuracy compared to the Basic Rounding approach. Extensive experiments on different data sets validate the correctness of our analysis and the effectiveness of our proposed rounding approaches.

1. Introduction

Collaborative Filtering (CF) systems have attracted lots of research attentions to overcome the information overload problem. Many recent algorithms, such as neighborhood methods, matrix factorization, various probabilistic models, clustering techniques, graph-based approaches and machine learning techniques, based on, e.g., association rule mining [4], have been proposed which aimed to improve the accuracy of prediction rating. For evaluation, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are widely adopted in many recommendation systems to measure the difference between the predicted scores and users' actual ratings, such as Slope One [8], Recommending Based on Rating Frequencies [5] and SVD model [9].

In general, CF algorithms output a decimal number as the predicted preference of an active user toward a given item. However, the actual ratings in most systems are integers. According to our experiment, it is necessary to round the decimal predicted scores to the proper integers, as this not only reduces the MAE values, but also improves the precision and quality of recommendation. Few preliminary works [3] have observed this fact. However, this issue has not been seriously studied. In fact, many recent popular CF algorithms, such as Recommending Based on Rating Frequencies [5], get very inspiring accuracy attributed to the fact that the original predicted ratings are integers.

In this paper, we present a comprehensive study on analyzing and demonstrating how and why rounding influences MAE and RMSE values. We show that the nonuniform distributions of ratings in training sets bias the predicted scores. Based on our empirical findings, we further propose two approaches to round the predicted rating to the most likely integer rating by considering the predicted ratings' probability distribution of each rating level. Extensive experiments show that our proposed approaches achieve inspiring performance in MAE metric.

The rest of this paper is organized as follows. Section 2 briefly reviews some related works. Section 3 analyzes the influence of rounding on MAE and RMSE, and proves rounding is necessary in

IOP Conf. Series: Materials Science and Engineering **324** (2018) 012049 doi:10.1088/1757-899X/324/1/012049

postprocessing of the predicted ratings using classical matrix factorization model. The empirical process of proof and our proposed rounding approaches are presented in Section 4. Section 5 presents the extensive experimental results of our proposed approaches compared to the baseline rounding approach. Section 6 concludes this paper and discusses our potential future works.

2. Related Works

In most CF algorithms, the rounding of predicted scores has seldom been seriously studied. Thomas Hofmann [6] and Rong Jin [7] take the user's rating trend into account and translate the ratings into normalized intervals by employing Gaussian and Decoupling normalization. Both of them avoid the issue of rounding the predicted rating to the integer rating.

A preliminary work [3] has observed that rounding can effectively improve the prediction accuracy and reduce the value of MAE. In [3], a rounding algorithm is proposed by considering the users' trend degree, departure degree and judgment degree. However, the computational complexity of this algorithm is very high when the rating matrix is very sparse (which usually is the case in most recommendation systems) and the number of the prediction rating is very huge. Most importantly, it is not illustrated why rounding can improve the precision of predicted ratings and reduce the MAE values.

The basic rounding approach which rounds the predicted rating to the nearest integer is the simplest solution for rounding the predicted rating to integer; however, it lacks effective rationale and is indemonstrable. In our work, we take it as the baseline rounding approach to validate the effectiveness of our proposed approaches.

In response, in this paper, we analyze and demonstrate how and why rounding predicted ratings to integer can influence the performance of the CF algorithms in terms of MAE and RMSE metrics, and prove that rounding is necessary in post-processing of the predicted ratings and improving the performance and quality of the recommendation. Based on our findings, we further propose two approaches to round the predicted rating to the most likely integer rating. Our experiments show that the proposed approaches get inspiring performance in MAE metric.

The three notable contributions of this paper are:

- First, we analyze and demonstrate why the rounding will bring different influences for MAE value and RMSE value, indeed reduce the MAE value, increase the RMSE values in some specific context.
- Second, we prove that rounding is necessary in post-processing of the predicted ratings and improving the performance and quality of the recommendation.
- Third, we propose two rounding approaches based on the predicted ratings' probability distribution of each rating level, which can be used to round the predicted rating to the most likely integer rating and offer better performance both in MAE and RMSE metrics.

3. The Influence of Rounding Prediction

This section, we demonstrate why and how rounding can affect the performance of recommendation systems in terms of MAE and RMSE metrics. A rating denotes the actual rating given by a user to an item . To distinguish predicted ratings from actual ones, we use the notation to denote the predicted score of user to item . The actual ratings are integers lying in .

For evaluation and analysis, we use 100k-MovieLens rating database, detailed description of the dataset can be seen in Section 5. As CF algorithm, here we employ Matrix Factorization (MF) model to analyze the probability distribution of predicted ratings. MF technique has become a popular choice for implementing CF and a survey can be found [9].

3.1. Evaluation Metrics

Mean Absolute Error (MAE) and RMSE (Root Mean Square Error) have widely been used in evaluating the accuracy of a recommender system, given by:

$$MAE = \frac{\sum_{n=1}^{N}\left|\hat{r}_n - r_n\right|}{N} \qquad RMSE = \sqrt{\frac{\sum_{n=1}^{N}\left(\hat{r}_n - r_n\right)^2}{N}}$$

where $\hat{r}_n$ means the prediction rating; r_n means the true rating in testing data set; N is the number of rating prediction pairs between the testing data and prediction result.

3.2. Performance of Rounding and Analysis
The Basic Rounding operator is defined as follows:

$$BasicRound(x) = \begin{cases} 1 & if\ x < 1 \\ \lfloor x \rfloor & if\ x - \lfloor x \rfloor < 0.5\ \&\ x \in [1, R] \\ \lceil x \rceil & if\ \lceil x \rceil - x \geq 0.5\ \&\ x \in [1, R] \\ R & if\ x > R \end{cases}$$

where $\lfloor x \rfloor$ is the floor integer of x (the largest integer less than or equal to x) and $\lceil x \rceil$ is the ceiling integer of x (the least integer greater than or equal to x).

As depicted in Fig. 1, the MAE values decrease when we round the predicted rating to the nearest neighbor integer with the basic rounding operator. While the RMSE values increase after rounding, especially when the training set ratio is larger than 0.5.

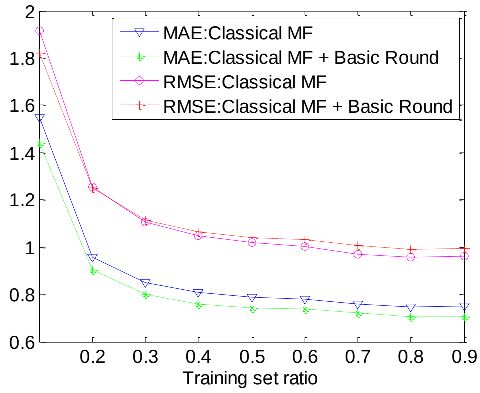

Fig. 1. Rounding the predicted ratings on MovieLens

- Reasons Analysis

In order to demonstrate why MAE values will decrease when predicted ratings, which predicted by classical Matrix Factorization model, rounded to integers, we analyzed the data set and all predicted ratings.

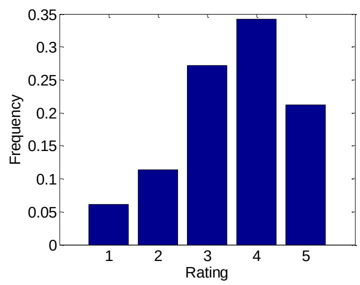

Fig. 2. Actual rating distribution of MovieLens

Fig. 2 illustrates the distribution of ratings in the training set for 5 different ratings (1~5). As can be seen in Fig. 3, the vast majority of the ratings are 3 and 4. In other words, users tend to rate 3 and 4 more frequently. As the result, the ratings of 3 and 4 have more influence on the final prediction model which is learned by all ratings in the training set.
We conduct experiments on MovieLens with the training set ratio of 0.8 and compute the difference between predicted ratings and actual ratings for all training samples. Fig. 3(a) illustrates the probability distribution of the prediction errors. Obviously the distribution follows a unimodal distribution, with a mode at -0.2. This means that the overall predicted ratings are lower than actual ratings by about 0.2.

In addition, we satiated the distributions of the prediction errors for 5 different rating levels (1~5) in the training set individually. Interestingly, as shown in Fig. 3(b-f), for different rating levels, the distributions of the prediction errors are quite different from each other. Indeed, although each distribution follows a unimodal distribution, their modes are materially different which range from xx to xx. It reflects that most of the modes of these distributions curve are not at the right value (0). One of the explanations for this effect is that among the training rating during the MF model learning, the majority ratings that used to train the model are rating 3 and rating 4. That cause the predicted ratings, which actual ratings is higher than 4, are lower than its actual value. Similarly, the predicted ratings, which actual ratings is lower than 3, are higher than its actual value.

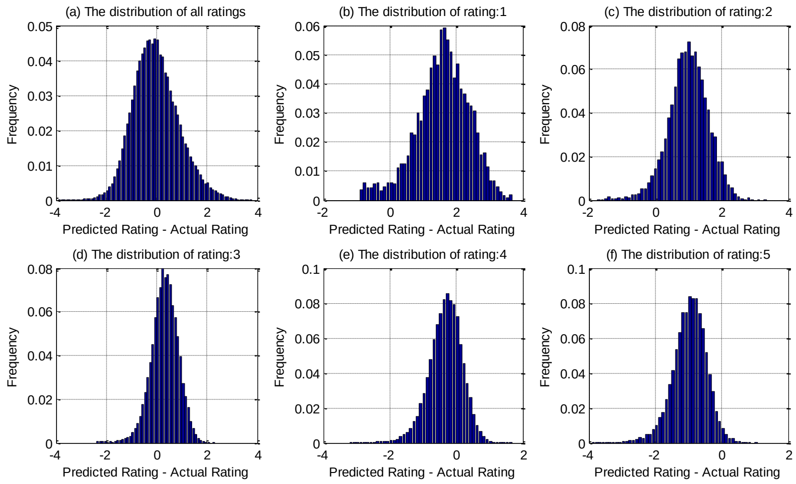

Fig. 3. Distribution of prediction errors in training set on MovieLens

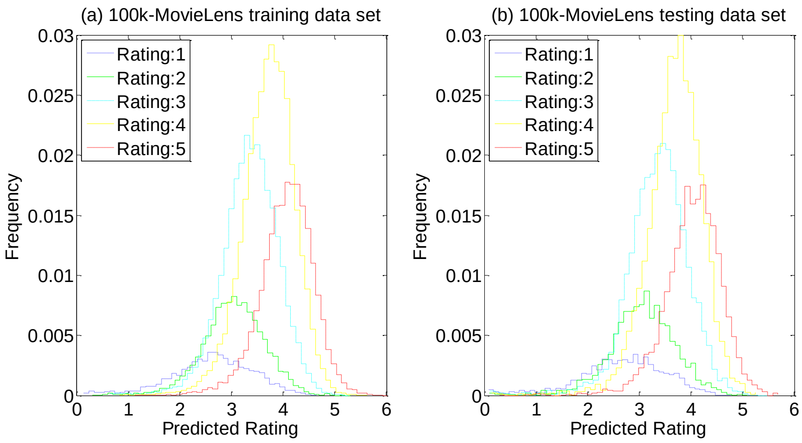

Fig. 4. All predicted ratings' probability distribution

We combine these distributions curve to one Figure as shown in Fig. 4(a). The x-axis represents the predicted scores, and the yaxis represents the percentage of the predicted ratings in training set. Fig. 4(a) shows each distribution of the each rating levels' predicted ratings in training set. For the actual rating of 1, the mode of its distribution is 2.2, which means most of the predicted rating are predicted to be around 2.2. This actually results in quite large prediction errors. Similarly, for the actual ratings of 2, 3, 4, and 5, the modes of their distributions are 2.9, 3.2, 3.8, and 4.3 respectively.

For the probability distribution function, we denote by f_r the probability function of the actual ratings of predicted ratings belonging to integer r. Naturally, $f_r(x)$ is the probability of the actual rating of the predicted rating x belonging to integer r.

According to Fig. 4(a), for a predicted rating x, supposing $\lfloor x \rfloor$ is the nearest integer of x, we have $x \in [\lfloor x \rfloor - 0.5, \lceil x \rceil + 0.5]$, and $f_{\lfloor x \rfloor}(x) = \max_{r \in \{1, \ldots, R\}} \{f_r(x)\}$.
In summary, from the above analysis, we can see rounding can definitely decrease the value of MAE. It should be noted that we assume the distribution of predicted ratings in test set keeps identical with the training set. Fig. 4(b) shows the distribution of predicted ratings on test set. Compared with Fig. 4(a), we can see the distributions of predicted ratings on test set and training are almost the same.

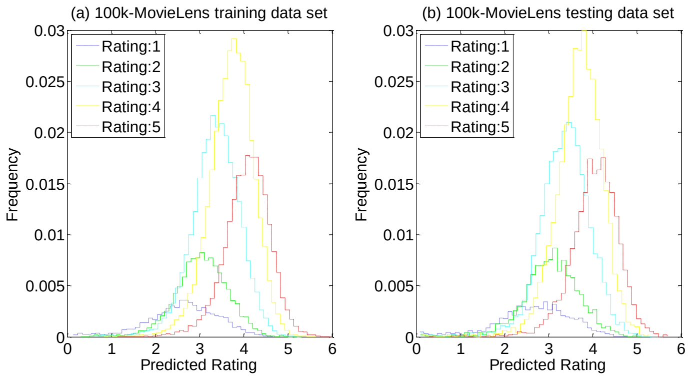

Figure 5. All predicted ratings' probability distribution

4. Proving and Our Approaches

In this section, we present the detailed analysis and discussion on the performance of MAE and RMSE in the rounding classical MF model. First, we theoretically prove why rounding bring different influences for MAE and RMSE. Second, we further propose two new rounding approaches to improve the performance of recommendation systems in term of MAE metric.

4.1. MAE Performance in Rounding Model

There exist two ways to round the predicted ratings to proper integers with the purpose of decreasing the MAE. One way is to round a rating to its neighbor, i.e. rounding a predicted rating x to the floor integer $\lfloor x \rfloor$ or the ceiling integer $\lceil x \rceil$. The other way is rounding a rating to the optimal integer. This allows a rating to be rounded to any integer, as long as this integer could decrease MAE the most. In the following, we discuss these two possible rounding approaches respectively.

- **Rounding to the Neighbor Integers**

Here, we first consider rounding the predicted ratings to their neighbor integers. It should be noted that we discuss not a signal predicted rating, but the all predicted rating which has same value.
Theorem 1: Given a predicted rating x, rounding x to at most one of its two neighbor integers ($\lfloor x \rfloor$ or $\lceil x \rceil$) can reduce the value of MAE after rounding.

Proof: For any predicted rating x, we have $x \in [\lfloor x \rfloor, \lceil x \rceil]$. If we round all predicted ratings with the value of x to $\lfloor x \rfloor$, the decrement of MAE is calculated as follows:

$$Dec_{MAE}(x, \lfloor x \rfloor) = \frac{\sum_{r=1}^{r=R} f_r(x) * |x - r| - \sum_{r=1}^{r=R} f_r(x) * |\lfloor x \rfloor - r|}{\sum_{r=1}^{r=R} f_r(x)} = \frac{\left[\sum_{r=1}^{r=\lfloor x \rfloor} f_r(x) - \sum_{r=\lceil x \rceil}^{r=R} f_r(x)\right] * (x - \lfloor x \rfloor)}{\sum_{r=1}^{r=R} f_r(x)} \tag{1}$$

Similarly, if we round these predicted ratings to $\lceil x \rceil$, the decrement of MAE will be:

$$Dec_{MAE}(x, \lceil x \rceil) = \frac{\left[\sum_{r=\lceil x \rceil}^{r=R} f_r(x) - \sum_{r=1}^{r=\lfloor x \rfloor} f_r(x)\right] * (\lceil x \rceil - x)}{\sum_{r=1}^{r=R} f_r(x)} \tag{2}$$

It is important to note that the sign of Equation (1) (2) depends on the sign of $\sum_{r=1}^{r=\lfloor x \rfloor} f_r(x) - \sum_{r=\lceil x \rceil}^{r=R} f_r(x)$, for both $(x - \lfloor x \rfloor) \geq 0$ and $(\lceil x \rceil - x) \geq 0$. Hence, only one direction of rounding can reduce the value of MAE.

Theorem 2: Given a predicted rating x, only rounding x to an integer can get the highest decrement of MAE compared with other numbers in $[\lfloor x \rfloor, \lceil x \rceil]$.

Proof: According to Equation (1) (2), as the signs of $Dec_{MAE}(x, \lfloor x \rfloor)$, $Dec_{MAE}(x, \lceil x \rceil)$ are all determined by the sign of $\sum_{r=1}^{r=\lfloor x \rfloor} f_r(x) - \sum_{r=\lceil x \rceil}^{r=R} f_r(x)$,the MAE value could be reduced by modifying the rating in only one direction, i.e. either increasing or decreasing x. In other words, if $\sum_{r=1}^{r=\lfloor x \rfloor} f_r(x) > \sum_{r=\lceil x \rceil}^{r=R} f_r(x)$, to decrease the MAE value, we should modify the original predicted rating to a candidate number x_0, where $x_0 \in (x, \lceil x \rceil)$. Therefore we can get the decrement on MAE value

$$Dec_{MAE}(x, x_0) = \frac{\sum_{r=1}^{r=R} f_r(x) * |x - r| - \sum_{r=1}^{r=R} f_r(x) * |x_0 - r|}{\sum_{r=1}^{r=R} f_r(x)} = \frac{\left[\sum_{r=1}^{r=\lfloor x \rfloor} f_i(x) - \sum_{r=\lceil x \rceil}^{r=R} f_r(x)\right] * (x - x_0)}{\sum_{r=1}^{r=R} f_r(x)} \tag{3}$$

Obviously, $Dec_{MAE}(x, x_0) < Dec_{MAE}(x, \lfloor x \rfloor)$. As can be seen in Equation (3), when x_0 is an integer, we get the largest decrement of MAE.

- **Rounding to the Optimal Integers**

In Section 4.1.1, we only consider rounding a predicted rating to its neighbor integer. However, sometimes this is not the optimal choice. For instance, in Fig. 3(b), for those samples with actual ratings of 1, the predicted ratings range from 0 to 5 with a mode of about 2.2. In this case, by only rounding the predicted rating to the neighbor integer (which is 2 for most samples) cannot achieve the best performance.
For this purpose, we modify Equation (1) (2) to calculate the decrement of MAE when rounding a given predicted rating x to each integer r_0, after derivation we get Equation (4) as follows:

$$Dec_{MAE}(x, r_0) = \begin{cases} \dfrac{\left[\left(\sum_{r=1}^{r=r_0} f_r(x) - \sum_{r=\lceil x \rceil}^{r=R} f_r(x)\right) * (x - r_0) + \sum_{i=r_0+1}^{i=\lfloor x \rfloor} \left(f_i(x) * (x - 2i + r_0)\right)\right]}{\sum_{r=1}^{r=R} f_r(x)}, & if\ r_0 \le \lfloor x \rfloor \\ \dfrac{\left[\left(\sum_{r=r_0}^{r=R} f_r(x) - \sum_{r=1}^{r=\lfloor x \rfloor} f_r(x)\right) * (r_0 - x) + \sum_{i=\lceil x \rceil}^{i=r_0-1} \left(f_i(x) * (2i - x - r_0)\right)\right]}{\sum_{r=1}^{r=R} f_r(x)}, & if\ r_0 \ge \lceil x \rceil \end{cases} \tag{4}$$

Theorem 3: Rounding a predicted rating to the neighbor integer is not always the optimal rounding approach.

Proof: Here, we take following as an example: given predicted rating: 3.2, and its probability distribution of each rating levels are:
$f_1(3.2) = 5/140$, $f_2(3.2) = 4/140$, $f_3(3.2) = 3/140$, $f_4(3.2) = 1/140$, $f_5(3.2) = 1/140$. According to Equation (4), $Dec_{MAE}(3.2, 3) = 2/14$, $Dec_{MAE}(3.2, 2) = 6/14$, $Dec_{MAE}(3.2, 1) = 2/14$ we also get: $Dec_{MAE}(3.2, 4) = -8/14$, $Dec_{MAE}(3.2, 5) = -20/14$ and it's no doubt that we will make the biggest decrement of MAE when we round 3.2 to 2, rather than round to 3. Thus, it is necessary to consider rounding the predicted ratings to other integers besides its neighbors.

Assume the optimal rounding integer of predicted rating x is r_0, and $\sum_{r=1}^{r=\lfloor x \rfloor} f_r(x) > \sum_{r=\lceil x \rceil}^{r=R} f_r(x)$. We can calculate the decrement of MAE as follows:

$$Dec_{MAE}(x, r_0) = \left[\left(\sum_{r=1}^{r=r_0} f_r(x) - \sum_{r=\lceil x \rceil}^{r=R} f_r(x)\right) * (x - r_0) + \sum_{i=r_0+1}^{i=\lfloor x \rfloor} \left(f_i(x) * (x - 2i + r_0)\right)\right] \Big/ \sum_{r=1}^{r=R} f_r(x)$$

$$Dec_{MAE}(x, \lfloor x \rfloor) = \left(\sum_{r=1}^{r=\lfloor x \rfloor} f_r(x) - \sum_{r=\lceil x \rceil}^{r=R} f_r(x)\right) * (x - \lfloor x \rfloor) \Big/ \sum_{r=1}^{r=R} f_r(x)$$

Finally, we get the function to judge which integer is the optimal to the predicted rating x, the biggest value of Φ, the optimal integer of the predicted rating x is.

$$\Phi_{floor}(r_0, x) = \left[\sum_{r=1}^{r=r_0} f_r(x) - \sum_{r=\lceil x \rceil}^{r=R} f_r(x)\right] * (\lfloor x \rfloor - r_0) + \sum_{i=r_0+1}^{i=\lfloor x \rfloor} \left[f_i(x) * (\lfloor x \rfloor + r_0 - 2i)\right],\ if\ r_0 \in \{1, \mathrm{L}\ , \lfloor x \rfloor\} \tag{5}$$

4.2. RMSE Performance in Rounding Model

Compared to MAE, RMSE disproportionately penalizes large errors, so that, given a test set with four hidden items RMSE would prefer a algorithm that makes an error of 1 on three ratings and 0 on the fourth to one that makes an error of 2 on one rating and 0 on all three others, while MAE would prefer the second algorithm, Obviously, it is difficult to strictly define which metric is better to evaluate the accuracy of the predicted ratings.

As shown by Fig. 1, after rounding the predicted ratings, the values of RMSE increase for all rating level especially when training set ratio is larger than 0.5. In this section, we discuss the effect on RMSE values by rounding the predicted ratings. Similar to Section 4.1, two different rounding methods are considered.

- **Rounding to the Neighbor Integers**

First we consider rounding a predicted rating to the neighbor integers. According to our discussion above, for any predicted rating x, $x \in [\lfloor x \rfloor, \lceil x \rceil]$, if we round all predicted ratings with values of x to $\lfloor x \rfloor$, the decrement of RMSE will be:

$$Dec_{RMSE}(x, \lfloor x \rfloor) = \left(\sqrt{\sum_{r=1}^{r=R} f_r(x) * (x-r)^2} - \sqrt{\sum_{r=1}^{r=R} f_r(x) * (\lfloor x \rfloor - r)^2} \right) \Big/ \sqrt{\sum_{r=1}^{r=R} f_r(x)} \tag{6}$$

Similarly, if we round these predicted ratings to $\lceil x \rceil$, the decrement of RMSE will be:

$$Dec_{RMSE}(x, \lceil x \rceil) = \left(\sqrt{\sum_{r=1}^{r=R} f_r(x) * (x-r)^2} - \sqrt{\sum_{r=1}^{r=R} f_r(x) * (\lceil x \rceil - r)^2} \right) \Big/ \sqrt{\sum_{r=1}^{r=R} f_r(x)} \tag{7}$$

As the signs of Equations (6) (7) are determined by the sign of their numerators, the value of RMSE may not be reduced by either rounding x to $\lceil x \rceil$ or $\lfloor x \rfloor$.

- **Rounding to the Optimal Integers**

Second, we consider rounding the predicted rating to the optimal integers in order to reduce the value of RMSE. According to the formulas above, we calculate the decrement of RMSE when rounding the predicted rating to each candidate integer as follows:

$$Dec_{RMSE}(x, r_0) = \left(\sqrt{\sum_{r=1}^{r=R} f_r(x) * (x-r)^2} - \sqrt{\sum_{r=1}^{r=R} f_r(x) * (r_0 - r)^2} \right) \Big/ \sqrt{\sum_{r=1}^{r=R} f_r(x)}$$

The sign of $Dec_{RMSE}(x, r_0)$ is determined by the sign of its square root of the numerators. Thus, any modification may not reduce RMSE by rounding x to any integers.

4.3. Our Proposed Approach: PDPR_Neighbor

According to Theorems 1 and 2, to decrease the MAE value, we should round a predicted rating to either $\lfloor x \rfloor$ or $\lceil x \rceil$(Note that here we only consider the possible values in $[\lfloor x \rfloor, \lceil x \rceil]$). More specifically, we should round the predicted rating x to the integer $\lfloor x \rfloor$ when $\sum_{r=1}^{r=\lfloor x \rfloor} f_r(x) > \sum_{r=\lceil x \rceil}^{r=R} f_r(x)$, and to the integer $\lceil x \rceil$ when $\sum_{r=1}^{r=\lfloor x \rfloor} f_r(x) \le \sum_{r=\lceil x \rceil}^{r=R} f_r(x)$. Based on this finding, we propose our first rounding algorithm which only rounds a predicted rating to its floor or ceiling integer, called *PDPR_Neighbor*, as follows:

$$PDPR_{Neighbor}(x) = \begin{cases} BasicRound(x)\ , \\ \quad if\ \ x \notin [1,R] \vee |Diff(x)| \le Threshold \\ \left[\lfloor x \rfloor * (1 + sgn(Diff(x))) + \lceil x \rceil * (1 - sgn(Diff(x))) \right] \big/ 2\ , \\ \quad if\ \ x \in [1,R] \wedge |Diff(x)| > Threshold \end{cases}$$

$$Diff(x) = \sum_{r=1}^{r=\lfloor x \rfloor} f_r(x) - \sum_{r=\lceil x \rceil}^{r=R} f_r(x)$$

$$Threshold = \min\left(\sum_{r=1}^{r=\lfloor x \rfloor} f_r(x), \sum_{r=\lceil x \rceil}^{r=R} f_r(x)\right) \Big/ 10$$

where $sgn(x)$ is sign function, as our rounding approach is under the assumption that the distribution of predicted rating in testing set keeps identical with the training set. So here, a threshold is set in our approach in order to avoid overfitting.

4.4. Our Proposed Approach: PDPR_Optimal

According to the example of Theorem 3's proof, the optimal integer of a predicted rating (3.2) can be determined by Equation (5). Thus, we propose our second algorithm which rounds a predicted rating to the optimal integer that makes largest decrement of MAE, called *PDPR_Optimal*, as follows:

$$PDPR_{Optimal}(x) = \begin{cases} BasicRound(x)\ ,\ if\ x \notin [1,R] \vee |Diff(x)| \le Threshold \\ \underset{r_0 \in PossibleIntsSet(x)}{\arg\max} \left(\Phi_{floor}(r_0,x) * (1 + sgn(Diff(x)) + \Phi_{ceiling}(r_0,x) * (1 + sgn(-Diff(x))\right)\ , \\ \qquad if\ x \in [1,R] \wedge |Diff(x)| > Threshold \end{cases}$$

$$PossibleIntsSet(x) = \begin{cases} \{ r \in Z \,|\, 1 \le r \le \lfloor x \rfloor \}\ ,\ if\ Diff(x) > 0 \\ \{ r \in Z \,|\, \lceil x \rceil \le r \le R \},\ if\ Diff(x) < 0 \end{cases}$$

A threshold is also set to avoid overfitting as our rounding approach is under the assumption that the distribution of predicted ratings in testing set must keep identical with the one in training set, while the distribution of predicted ratings can't completely consistent indeed.

The core part of above two round algorithms (*PDPR_Neighbor* and *PDPR_Optimal*) are all the calculation of each rating level's probabilistic distribution. We assume that the distribution of predicted ratings for each rating level follows Gaussian distribution. Compared with the algorithm *PDPR_Neighbor*, *PDPR_Optimal* rounds the predicted rating to the optimal integer in theory.

5. EXPERIMENT

Our approaches are tested on two different movie domains. The first one is the 100k-MovieLens rating data set with 100,000 ratings by 943 users on 1,682 movies. The ratings are integers ranging from 1 to 5. The second one is the EachMovie dataset with 2,811,983 numeric ratings by 72,916 users on 1,628 movies. The ratings range from 0.0 to 1.0 with a step size of 0.2. For EachMovie, we choose 100,118 ratings by the first 2936 users on 1,628 movies. We multiply all ratings by 5, and then plus one under the amplified ratings. Thus, the resulting ratings become integers ranging from 1 to 6.

5.1. The Necessity of Rounding Predicted Ratings

We compare the following algorithms with Basic Rounding approach: user-based kNN (using Pearson similarity and the neighborhood size of 30 which is suggested as the optimal value in literature [2]), Slope One [8], and Classical MF [9]. Besides, we also take the MAE of Rating Frequencies algorithm [5] as a reference, which predicted ratings are integers, to evaluate the rounding performance.

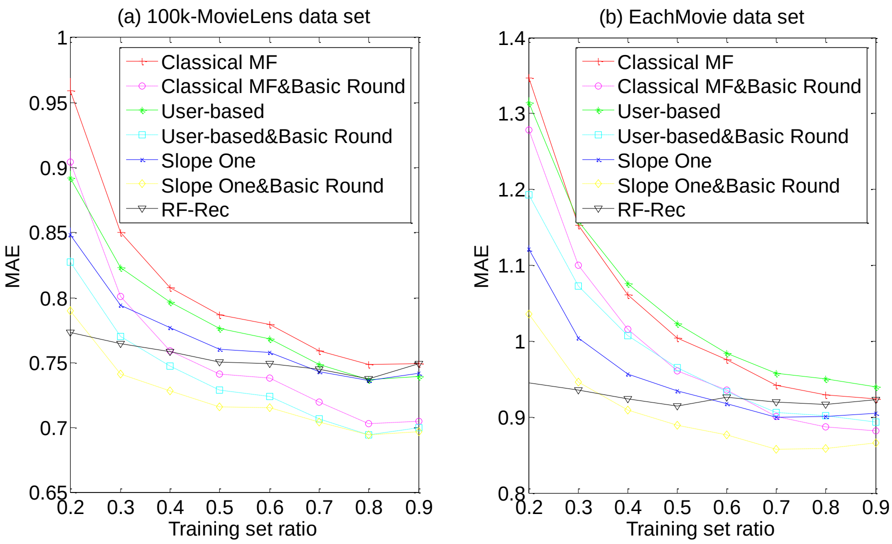

Fig. 6. The MAE values on MovieLens / EachMovie

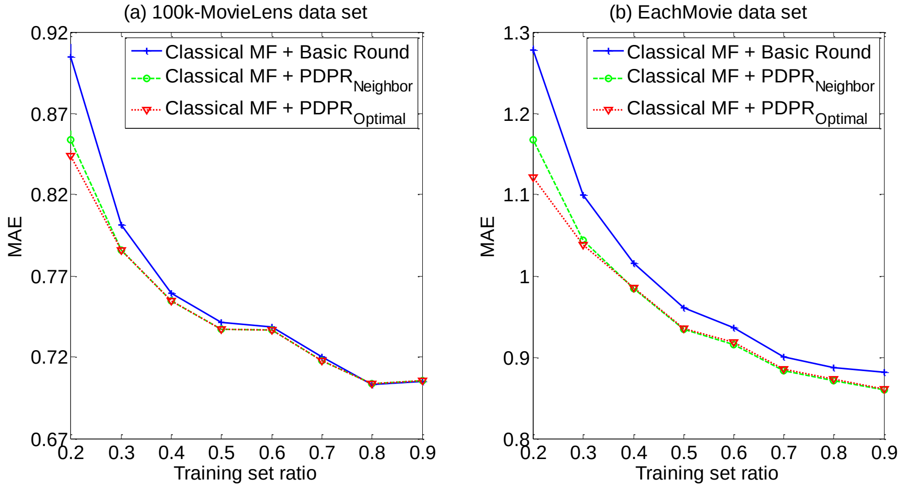

Fig. 7. The MAE values on two different data sets

Fig. 6(a), 7(b) shows the MAE values for different training set ratio from 0.1 to 0.9 on the MovieLens and EachMovie datasets respectively. As can be seen in Fig. 6(a), 7(b), for all different settings, the MAE values are significantly decreased after rounding the predicted ratings by Basic Round approach. This proves that rounding the predicted rating can decrease the MAE values for all tested recommendation algorithms. In particular, the decrement of MAE values is more inspiring on EachMovie data set.

5.2. The Effectiveness of Our Approaches

In this section, we conduct a series of experiments to compare the performance of our approach by changing the training ratio from 0.2 to 0.9 on MovieLens data set and EachMovie data set. From Fig. 7, it can be observed our proposed approaches achieve better performance on different training set ratios. Specifically, the MAE values those using our approaches are competitive with Basic Rounding. As large majority of absolute difference between original predicted ratings and actual ratings are less than 0.5, there even have not a pinpoint of difference between PDPR_Neighbor and PDPR_Optimal, while PDPR_Optimal have a higher robustness under some specified context.

6. Conclusion

In this work, we demonstrate the reason why rounding can bring different influences for MAE and RMSE metrics, and prove that rounding is necessary in post-processing of the predicted ratings and improving the performance and quality of the recommendation. Additionally, we also propose two rounding approaches based on the predicted ratings' probability distribution of each rating level, which can be used to round the predicted rating to the right integer rating and offer better predict accuracy. Extensive experiments on different data sets confirm the correctness of our analysis and effectiveness of our proposed approaches.

In the future, we will reform our rounding approach in order to improve the recall and precision of Top-K recommendation. And it is hoped that the rounding of predicted rating presented here will prove useful to the CF community as a reference scheme.

7. References

[1] Johns. Breese, David Heckerman and Carl Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, 43-52, San Francisco, 1998.

[2] Jon Herlocker, Joseph A. Konstan, and John Riedl. An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms. Information Retrieval, 5(4):287-310, 2002.

[3] Li Yong, Xu De-zhi, Zhang Yong and Xing Chun-xiao. MVC-based Incremental Reengineering Approach. Journal of Chinese Computer Systems, 29(3):469-472, 2008.

[4] Gediminas Adomavicius and Alexander 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(6):734-749, 2005.
[5] Fatih Gedikli, Dietmar Jannach. Recommending Based on Rating Frequencies. ACM Conference on Recommender Systems, 233-236, Barcelona, 2010.
[6] Thomas Hofmann. Collaborative filtering via Gaussian probabilistic latent semantic analysis. ACM Special Interest Group on Information Retrieval, 259-266, Toronto, 2003.
[7] Rong Jin, Luo Si, ChengXiang Zhai and Jamie Callan. Collaborative Filtering with Decoupled Models for Preferences and Ratings. ACM International Conference on Information and Knowledge Management, 309-316, New Orleans, 2003.
[8] Daniel Lemire and Anna Maclachlan. Slope One Predictors for Online Rating-Based Collaborative Filtering. In Proceedings of the 5th SIAM International Conference on Data Mining, 471-475, Newport Beach, 2005.
[9] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. IEEE Computer, 42(8):30–37, 2009.

Acknowledgments
Acknowledging assistance and encouragement from Data Platform Project Group (Door) , special work by technical staff or financial support from CFETS Information Technology (Shanghai) Co., Ltd.