École Polytechnique Fédérale de Lausanne

CS 449 : Systems for Data Science

# Milestone 1 Report

Karim Hadidane
SCIPER: 271018

# Part 3

## Question 3.1.1 :

**Compute and report the global average rating ($\bar{r}_{\bullet,\bullet}$). Do ratings, on average, coincide with the middle of the rating scale (3 from the scale {1, 2, 3, 4, 5})? If not, are they higher or lower on average? By how much?**

*Answer :*
The global average rating is $\bar{r}_{\bullet,\bullet} = 3.5298$
Hence, the ratings do not coincide with the middle of the rating scale 3. They are higher on average by 0.5298.

## Question 3.1.2 :

**Compute the average rating for each user ($\bar{r}_{u,\bullet}$). Do all users rate, on average, close to the global average? Check min and max for user average and assume a difference less than 0.5 is small. Do most users rate, on average, close to the global average? Calculate and report the ratio of users with average ratings that deviate with less than 0.5 from the global average.**

*Answer :*
The minimum user average rating is **1.4919** and the maximum user average rating is **4.8695**. These values reflect the fact that some users rate, on average, significantly lower or higher than the global average. Therefore, not all users rate, on average, close to the global average.
The ratio of users that rate close to the global average is **0.7465** . Therefore, most of the users rate, on average, close to the global average.

## Question 3.1.3 :

**Compute the average rating for each item ($\bar{r}_{\bullet,i}$). Are all items rated, on average, close to the global average? Check min and max for item average and assume a difference less than 0.5 is small. Are most items rated, on average, close to the global average?**

**Calculate and report the ratio of items with average ratings that deviate with less than 0.5 from the global average.**

*Answer :*
Not all items are rated, on average, close to the global average reported in the previous question.
The minimum item average rating is **1** and the maximum item average rating is **5**.
The ratio of items that are rated close to the global average is **0.4898** . Therefore, not most of the items are rated, on average, close to the global average.

## Question 3.1.4 :

**Compare the prediction accuracy (average MAE on ml-100k/u1.test) of the previous methods ($\bar{r}_{\bullet,\bullet}$, $\bar{r}_{u,\bullet}$, $\bar{r}_{\bullet,i}$) to the proposed baseline ($p_{u,i}$, Eq. 5). Report the results you obtained in a table. Discuss the difference(s) you observed and why you think they occur. For items that do not have ratings in the training set, use the global average ($\bar{r}_{\bullet,\bullet}$), instead of the item-specific average ($\bar{r}_{\bullet,i}$).**

*Answer :*
The Mean Absolute Error is used here to assess the accuracy of the prediction methods. The results are given in the table below.

| **Prediction method** | Global average | User average | Item average | Baseline |
|---|---|---|---|---|
| **MAE** | 0.9680 | 0.8501 | 0.8275 | 0.7669 |

We notice that the global average approach yields the highest MAE and equivalently the worst accuracy. On the other hand, the baseline method has the best accuracy. The per-user average and per-item average methods have close MAE values.
The first method generalizes the global average rating to predict all ratings. It does not consider the user rating behavior or the item's rating history, which explains its poor accuracy.
The second approach, the user average, uses information about the user rating behavior (user average) to predict its rating for a new item i. Hence, its accuracy is better than the global average method.
The third prediction method, based on item average, does slightly better than the previous one. The improvement of the accuracy implies that the

item rating average is more meaningful in predicting the actual rating than the user rating average. In other words, highly rated (highly appreciated) items in the training set are more likely to have a high rating in the test set.
The baseline method does better than the three previous methods. It combines knowledge from the user behavior (user's average rating) and the item's rating history (global average deviation).

## Question 3.1.5 :

**Measure the time required for computing predictions for all ratings in the test set (ml-100k/u1.text) with all four methods by recording the current time before and after(ex: with System.nanoTime() in Scala). The duration is the difference between the two. For all four methods, perform ten measurements and report in a table the min, max, average, and standard-deviation. Report also the technical specifications (model, CPU speed, RAM, OS, language version) of the machine on which you ran the tests. Which of the four prediction methods is the most expensive to compute? How much more compared to using the global average rating ($\bar{r}_{\bullet,\bullet}$)? Calculate and report the ratio between the average time for computing the baseline (Eq. 5) and the average time for computing the global average.**

*Answer :*

The time measurements results are given in the table below.

| | min | max | average | standard deviation |
|---|---|---|---|---|
| **global average** | 52000 | 112000 | 71100.0 | 17438.1765 |
| **user average** | 874000 | 1120000 | 988200.0 | 89724.9129 |
| **item average** | 770000 | 856000 | 807700.0 | 29836.3871 |
| **baseline** | 3233000 | 4590000 | 3551500.0 | 372786.8694 |

• Technical specifications of the machine:
Model : Apple MacBook Pro 2018
CPU speed : 2.3 GHz Intel Core i5
RAM : 8 GB 2133 MHz LPDDR3
OS : macOS High Sierra
Language version : scala 2.12.13
• The baseline prediction is the most expensive among the four methods. When compared with the global average approach, its average time cost is almost 50 times higher. (49.9507)

# Part 4

## Question 4.1.1 :

**Report your personal top 5 recommendations using the baseline predictor (Eq. 5), including the movie identifier, the movie title, and the prediction score. If additional recommendations have the same predicted value as the top 5 recommendations, prioritize the movies with the small- est identifiers in your top 5 (ex: if the top 8 recommendations all have predicted scores of 5.0, choose the top 5 with the smallest ids.) so your results do not depend on the sorting behaviour. Are these movies you have actually liked (but did not rate) or would like to see in the future?**

*Answer :*
Top 5 recommendations are:

814, "Great Day in Harlem", 5.0
1122, "They Made Me a Criminal (1939)", 5.0
1189, "Prefontaine (1997)", 5.0
1201, "Marlene Dietrich: Shadow and Light (1996)", 5.0
1467, "Saint of Fort Washington", 5.0

Yes those movies are indeed what I liked and did not rate.

## Question 4.1.2 :

**How could you modify the predictions to favour more popular movies, e.g. by smoothly decreasing the prediction score of movies with few ratings while keeping the prediction score of those with many ratings almost identical? Provide the equation(s) of your modifications, which equations of this document they are intended to replace, and the new top5 recommendations you obtain for yourself (movie identifier, movie title, prediction score).**

*Answer :*

One way to do so is by changing Eq 5 in the following way:

$$q_{u,i} = \bar{r}_{u,\bullet} + \bar{\hat{r}}_{\bullet,i} * scale((\bar{r}_{u,\bullet} + \bar{\hat{r}}_{\bullet,i}), \bar{r}_{u,\bullet}) - 0.5 + tanh(\frac{\mid U(i) \mid}{max \mid U(i) \mid} * \mid U(i) \mid)$$

where $\mid U(i) \mid$denotes the number of ratings for item i.

The values that violate the rating interval [1,5] are mapped to the corresponding boundaries (e.g. if the output is less than 1, it is mapped to 1).
The tanh function argument is more significant for popular movies, and therefore, its value would be close to 1, which boosts the prediction of the baseline method. For non-popular items, the tanh function's value would be close to 0, and their baseline prediction would decrease by at most 0.5.

New top 5 recommendations are:

408, Close Shave, 4.8569
318,Schindler's List (1993), 4.8546
169, Wrong Trousers, 4.8292
483, Casablanca (1942), 4.8236
64, Shawshank Redemption, 4.7978