POLITECNICO DI MILANO

Knowledge • 0 teams

Recommender System 2015 Challenge Polimi preview

This is the challenge for the recommender system course 2015/2016 held in Politecnico di Milano.

Recommend movie lists using content information

Welcome to the page of the RecSys course: Challenge competition.

Description

Please read carefully till the end of the page!

- In this competition you are required to predict a list of 5 movies for a set of users.
- The original unsplitted dataset includes almost 190K ratings for 15K users and 37K items with 20K features.
- A subset of about 4K users has been selected as test users.
- The goal is to recommend a list of 5 relevant items for each user (consider items with rating >= 8 as relevant).
- MAP@5 is used for evaluation.
- You can use any kind of recommender algorithm you wish (e.g., collaborative-filtering, content-based, hybrid, etc.).

The prize

(in exam points, not euros ...).

Each team will receive the final score according to the quality of recommendations computed on a **private leader-board**:

- the position in the final private leader-board when the competition ends
- positions in the private leader board of every 2 week, during the competition
- improvement in the evaluation metric, during the competition

- quality of recommendation in comparison to 4 baselines

The final score is computed on a private leader-board with the following formula:

final_score = baseline_bonus + activity_bonus + standing_points

Attention: results on the public leader-board are computed on a different subset of the test set, so it may differ from the private one.

Baseline Bonus

You are provided with 4 baselines scores that you have to exceed in order to get the full prize. Such baselines are computed with four different algorithms, namely Random, Top-Popular, Algorithm-A and Algorithm-B, ordered by increasing accuracy (Algorithms A and B are kept secret to avoid helping the teams).

Each algorithm is associated with a bonus score (*b*) that adds to your final score if you exceed it in the final leader-board:

$$b = \begin{cases} \text{Random} & +0 \\ \text{Top-Popular} & +1 \\ \text{Algorithm-A} & +2 \\ \text{Algorithm-B} & +3 \end{cases}$$

Important. These bonuses are non-additive, this means that you will get 1 points of bonus if you exceed Top-Popular!

Important. The baseline bonus is computed only at the final submission and not during the biweekly deadlines.

Activity Bonus

Teams active during the competition will receive extra points. If a team is able to improve the MAP@5 of their last best submission by 0.0005 (in both public and private leader-boards) the team will receive two points of bonus:

$$\delta_i = 2 \times [\text{new} - \text{old} \geq 0.0005]$$

The improvement is evaluated at each biweekly deadline. Activity bonuses are cumulative.

Standing points

According to the standing in the private leader-board, every two weeks points will be assigned to the teams, in the following manner:

$$s_i = \begin{cases} \text{first team} & 28 \\ \text{second team} & 27 \\ \text{top } 20\% & 26 \\ 20\% - 40\% & 22 \\ 40\% - 60\% & 18 \\ 60\% - 80\% & 16 \\ 80\% - 100\% & 14 \end{cases}$$

Final score

The final score is computed with the following formula:

$$\text{final-score} = \frac{\sum_i w_i \cdot s_i}{\sum_i w_i} + b + \sum_i \delta_i$$

where "i" is the i-th biweekly deadline and

$$w_i = \begin{cases} 1 & \text{intermediate deadline} \\ 2 & \text{final deadline} \end{cases}$$

The last deadline weights twice each intermediate deadline.

Attention. Results on the public leader-board are computed on a different subset of the test set, so it may differ from the private one.

Important. If you do not exceed the random baseline, you will get zero points overall.

Deadlines

Deadlines will be every 15 days, on the following dates (at 10.00 CET):

- 1 November
- 15 November
- 1 December
- 15 December
- 15 January
- 1 February (**final deadline**)

Final delivery

Each team must deliver the source code and a 4 pages report in which you describe your algorithm, your experiments and how to fully reproduce them.

You should send your data by one week after the deadline of the competition to this page in a file named TeamName.zip containing:

the composition of the team (Matricola, Name, Surname)
the source code and additional data files you used
the report (please, use this template)
We reserve us the right to change your final standing according to the quality of the code and report that you will deliver!

At the end, each team should also prepare a short public presentation which the team member(s) will explain briefly their algorithm to the class, as well as the story they have tried different algorithms during the competition until they have found their best algorithm.

Now do your best and have fun!