Assignment 1: evaluation and non-personalized algorithms

Denis Kotkov

Tasks

1. Explore the dataset
2. Implement popularity and randomized algorithms
3. Evaluate the algorithms
4. Explore generated recommendations

All the tasks need to be implemented from scratch. The use of dedicated libraries for recommendation, such as recbole or lenskit is prohibited. However, feel free to use other libraries, such as numpy and pandas.

Task 1: dataset exploration

- Count ratings, unique users and unique items
- Calculate average numbers of ratings per user and per item, and sparsity of the dataset (fraction of missing values)
- What do you think about these numbers?
- Plot three distributions: number of users per number of ratings, number of items per number of ratings and rating values per number of ratings. What patterns do you notice?

Task 2: non-personalized algorithms

- Implement the following two algorithms:
  - Popularity algorithm. For each user, the algorithm orders items based on the number of ratings it received in the training set.
  - Randomized algorithm. For each user, the algorithm orders items randomly.

Task 3: evaluation of algorithms

- Split the dataset by users into training and test in proportion of 80/20. For each user, 80% of ratings should belong to the training set, while 20% of ratings – to the test set.
- The recommendations should be generated based on only the training set.
- Measure and report NDCG@5 (average for all the users) for each algorithm. When evaluting an algorithm, uniformly sample 100 negative examples for each user.
- Which algoritm has the highest performance and why?

Task 4: recommendation exploration

- Have a look at top-10 recommendations provided by the popularity algorithm. Do you think these movies are popular?

Submission

- Please implement the following structure of your deliberables:
  - “code” – folder containg programming code of the assignment (please do not include the dataset)
  - “report.pdf” – pdf file containing answers to the questions
- Submit your assignment via the corresponding return box in Moodle by the deadline
- Optionally, you can also indicate how long it took you to complete the assignment