Assignment 2: kNN

Denis Kotkov

Tasks

1. Implement kNN (user-based or item-based) collaborative filtering, popularity and randomized algorithms

2. Tune the chosen kNN collaborative filtering algorithm

3. Evaluate the algorithms

4. Explore generated recommendations

In this assignment, please use the data split provided in Moodle (data_split(updated).zip)

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: algorithm implementation

- Implement the following three algorithms:
  - kNN collaborative filtering. Feel free to choose user-based or item-based collaborative filtering for this assignment. For each user, the algorithm orders items based on the predicted rating.
  - 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 2: tuning the kNN algorithm

- Use training and validation sets to infer optimal implementation decisions:
  - Similarity measures: cosine similarity, euclidean distance, Pearson correlation…
  - Selecting neighbors: top k neighbors (picking the optimal k), similarity threshold (picking threshold), clustering
  - Rating normalization
- Report implementation decisions that you tested along with their performance in terms of NDCG@5
- It is up to you what and how many implementation decisions to investigate

Task 3: evaluation of algorithms

- Use training and validation sets combined for training and test set for testing

- Measure and report NDCG@5 (average for all the users) for each algorithm. When evaluting algorithms, uniformly sample 100 negative examples for each user

- For the given data split, NDCG@5 for the popularity baseline should be around 0.75. Perfomance of the random baseline can vary, but is lower than that of the popularity baseline

Task 4: recommendation exploration

- Pick a user at random from the dataset, have a look at movies this user watched and top-5 recommendations provided by kNN collaborative filtering. Do you think this user would enjoy the recommendations?

- For a few users (or movies), pick a few top similar users (or movies). Do you think these users (or items) are similar to the target ones?

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