

Name: - Umang Kundaria

ucf_id: - 5326402

Assignment-5

Machine Learning for Recommender Systems

1.Recommender systems are a hot topic. Recommendation systems can be formulated as a task of matrix completion in machine learning. Recommender systems aim to predict the rating that a user will give for an item (e.g., a restaurant, a movie, a product).

2. Download the movie rating dataset from: <https://www.kaggle.com/rounakbanik/the-movies-dataset>. These files contain metadata for all 45,000 movies listed in the Full Movie Lens Dataset. The dataset consists of movies released on or before July 2017. Data points include cast, crew, plot keywords, budget, revenue, posters, release dates, languages, production companies, countries, TMDb vote counts and vote averages. This dataset also has files containing 26 million ratings from 270,000 users for all 45,000 movies. Ratings are on a scale of 1-5 and have been obtained from the official GroupLens website.

3. Building a small recommender system with the matrix data: "ratings.csv". You can use the recommender system library: Surprise (<http://surpriselib.com>), use other recommender system libraries, or implement from scratches.

- a. Read data from "ratings.csv" with line format: 'userID movieID rating timestamp'.
- b. MAE and RMSE are two famous metrics for evaluating the performances of a recommender system. The definition of MAE can be found via: https://en.wikipedia.org/wiki/Mean_absolute_error. The definition of RMSE can be found via: https://en.wikipedia.org/wiki/Root-mean-square_deviation
- c. Compute the average MAE and RMSE of the Probabilistic Matrix Factorization (PMF), User based Collaborative Filtering, Item based Collaborative Filtering, under the 5-folds cross-validation
- d. Compare the average (mean) performances of User-based collaborative filtering, item-based collaborative filtering, PMF with respect to RMSE and MAE. Which ML model is the best in the movie rating data?

Ans: - PMF (ML model) is the best in the movie rating data because The PMF had the lowest RMSE and MAE

- e. Examine how the cosine, MSD (Mean Squared Difference), and Pearson similarities impact the performances of User based Collaborative Filtering and Item based Collaborative Filtering. Plot your results. Is the impact of the three metrics on User based Collaborative Filtering consistent with the impact of the three metrics on Item based Collaborative Filtering?

Ans: The impact of the three metrics on User based Collaborative Filtering is consistent with the impact of the three metrics on Item based Collaborative Filtering and both impacts can be shown in the codes and graph.

- f. Examine how the number of neighbours impacts the performances of User based Collaborative Filtering and Item based Collaborative Filtering? Plot your results.
- g. Identify the best number of neighbour (denoted by K) for User/Item based collaborative filtering in terms of RMSE. Is the best K of User based collaborative filtering the same with the best K of Item based collaborative filtering?

Ans: - The best number of neighbour (denoted by K) for User/Item based collaborative filtering in terms of RMSE is 15 and 68. The best K of User based collaborative filtering is not the same with the best K of Item based collaborative filtering.

Links: -

<https://colab.research.google.com/drive/1fgfK6jFFNu1-t4dcjygu5DpqoWeVasqs?usp=sharing>

https://colab.research.google.com/gist/umang1111/ddc8d77b4ffd6e29f2d732b4878629cd/hw_5-ml.ipynb

[https://github.com/umang1111/hw5/blob/main/hw_5\(ML\).ipynb](https://github.com/umang1111/hw5/blob/main/hw_5(ML).ipynb)