Movies recommender system

The Pussycats: Alex Draghi, Laura Falciani, Nicola Gennaro

2018

Table of Contents

Measures

1. Precision
2. Recall
3. Sps (Short-term Prediction Successes)

The Movielens dataset

The dataset contains 1,000,209 anonymous ratings of approximately 3,900 movies made by 6,040 MovieLens users since 2000 to 2003.

Train-Test split

We divided the users in two groups: train set (0.8) and test set (0.2), having care that the number of ratings in the two sets was divided according to the same ratio.
We used this partition throughout the whole analysis.

Top n predictions

We used two different approaches:

1. Suggest most rated movies
2. Suggest highest Laplace smoothing score movies

KNN

We used the cosine similarity as distance measure between users.

$$d_{i,u} = \frac{\mathcal{R}_i \cdot \mathcal{R}_u}{\|\mathcal{R}_i\| \|\mathcal{R}_u\|}$$

Then, using the k closest users to user i a score is computed between the user i and an item j:

$$s_{i,j} = \sum_{u \in \mathcal{N}_k(i)} d_{i,u} \mathbb{1}(j \in \mathcal{R}_u)$$