



Тушканова Ольга Николаевна




Входные данные

Набор **ml-100k**

user, item, rating, timestamp

62	498	4	879373848
62	382	3	879375537
28	209	4	881961214
135	23	4	879857765
32	294	3	883709863
90	382	5	891383835
286	208	4	877531942
293	685	3	888905170
216	144	4	880234639
166	328	5	886397722
250	496	4	878090499
271	132	5	885848672
160	174	5	876860807
265	118	4	875320714
198	498	3	884207492
42	96	5	881107178
168	151	5	884288058
110	307	4	886987260
58	144	4	884304936
90	648	4	891384754
271	346	4	885844430

Задание

Для заданного пользователя вывести **топ-5** рекомендаций **с названиями, датой выхода и рейтингом**.

Пример вывода:

User 88

1240	('Ghost in the Shell (1995)', '12-Apr-1996')	4.51
1368	('Mina Tannenbaum (1994)', '01-Jan-1994')	4.403
1449	('Pather Panchali (1955)', '22-Mar-1996')	4.338
1131	('Safe (1995)', '01-Jan-1995')	4.214
922	('Dead Man (1995)', '10-May-1996')	4.186

Примечания

- Для расчета оценок использовать подход user-based коллаборативной фильтрации, метод kNN, **k = 4**.
- Для расчета оценки \hat{r}_{ui} пользователя u и для фильма i использовать формулу

$$\hat{r}_{ui} = \bar{r}_u + \frac{\sum_{v=1}^k sim_{vu} \cdot (r_{vi} - \bar{r}_v)}{\sum_{v=1}^k |sim_{vu}|}$$

где \hat{r}_{ui} - рассчитываемая оценка,
 \bar{r}_u - средняя оценка у пользователя u ,
 \bar{r}_v - средняя оценка у пользователя v ,
 r_{vi} - оценка пользователя v для фильма i ,
 sim_{vu} - значение метрики сходства для пользователей u и v

Примечания

- В качестве метрики сходства использовать метрику косинуса, при этом пользователи должны иметь **не менее 5** общих фильмов:

$$sim_{u,v} = \frac{\sum_{i=1}^m u_i v_i}{\sqrt{\sum_{l=1}^m u_l^2} \sqrt{\sum_{l=1}^m v_l^2}}$$

где u_i - оценка пользователя u для фильма i ,
 v_i - оценка пользователя v для фильма i ,
 m - количество фильмов, для которых у обоих пользователей есть оценка

Примечания

- ~/.surprise_data/ml-100k/ml-100k/README – описание формата данных
- ~/.surprise_data/ml-100k/ml-100k/u.item – информация о фильмах

Конфигурация метрики сходства

https://surprise.readthedocs.io/en/stable/prediction_algorithms.html#similarity-measures-configuration

Часто задаваемые вопросы

<https://surprise.readthedocs.io/en/stable/FAQ.html#>

Спасибо за внимание!