

WSDM - KKBox's Music Recommendation Challenge

Василий Рубцов

20 января 2018

Данные

Информация о первом прослушивании композиции пользователем

- msno — пользователь
- song_id — композиция
- source_system_tab
- source_screen_name
- source_type
- target — повторное прослушивание в течении месяца

Данные о пользователе

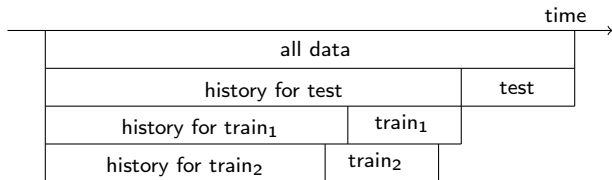
- city
- bd — возраст
- gender
- registered_via — метод регистрации
- registration_init_time — время регистрации
- expiration_date — окончание подписки

Данные о композиции

- song_length
- genre_ids
- artist_name
- composer
- lyricist
- language

Метрика: ROC AUC

Обучающая выборка



Признаки

Основная стратегия — группировать по категориальным признакам (парам и тройкам) и применять некоторую функцию.

Supervised features

- Mean (by categorical features, pairs and triples)

- Linear regression

- 1) msno
- 2) msno, artist_name
- 3) msno, genre_ids
- 4) msno, source_type
- 5) msno, composer
- 6) msno, language
- 7) song_id

- Matrix factorization (LightFM)

$$r_{uic} = b_{uc} + b_u + b_i + \langle p_{uc} + p_u, p_i \rangle$$

Unsupervised features

- Count (by categorical features)
- Time from the previous listening (by categorical features)
- Time before the next listening (by categorical features)
- Count from future (by categorical features)
- Count from past (by categorical features)
- Last time difference (by categorical features)
- Share of unique songs
- Time from test period
- Part of a song listened
- Song length
- User age
- Registration date
- Expiration date

Top 20 features

Feature	Gain
Matrix factorization	4116.54
Mean (genre_ids, source_screen_name, source_type)	1897.92
Last time difference (artist_name, language, msno)	1502.82
Last time difference (artist_name, genre_ids, msno)	1187.89
Mean (genre_ids, lyricist)	1071.63
Time to the next listening (msno, source_type)	636.76
Mean (source_system_tab, source_type)	588.12
Mean (gender, song_id)	581.72
Mean (city, song_id)	572.86
Mean (artist_name, composer, lyricist)	570.04
Mean (genre_ids, language, source_screen_name)	512.88
Count from future (registered_via, song_id)	495.69
Time to the next listening (msno, source_system_tab)	487.80
Count from future (song_id, source_type)	462.39
Mean (gender, registered_via, song_id)	458.42
Count from future (song_id)	454.10
Time from the previous listening (msno)	412.73
Mean (gender, source_screen_name, source_type)	381.84
Mean (language, msno, source_screen_name)	380.28
Mean (registered_via, source_system_tab)	361.06




















Таблица: Single model's score

Model	Train set	Matrix factorization feature	Score
XGBoost	1	Yes	0.74421
XGBoost	1	No	0.74377
XGBoost	2	Yes	0.74380
XGBoost	2	No	0.74331
Catboost	1	Yes	0.74496
Catboost	1	No	0.74441
Catboost	2	Yes	0.74451
Catboost	2	No	0.74384


















Blending: 0.74688

https://github.com/VasiliyRubtsov/wsdm_music_recommendations/blob/master/pipeline.ipynb

Public leaderboard

#	△priv	Team Name	Kernel	Team Members	Score ?	Entries	Last
1	—	Bing Bai			0.75001	263	1mo
2	▼ 1	Rubtsov Vasily			0.74982	32	1mo
3	▲ 1	Magic Recommenders		 	0.74909	105	1mo
4	▼ 1	ekffar			0.73826	148	1mo
5	▲ 1	Devin			0.73579	58	1mo
6	—	EdZhang			0.72930	95	1mo
7	—	Chih-Ming			0.72566	82	2mo
8	—	rn51			0.72078	49	1mo
9	▼ 1	nanaya tachibana			0.71892	17	1mo
10	▲ 1	nPowered		  	0.71799	180	1mo
11	—	wehaveNOidea		 	0.71680	129	1mo
12	—	yese			0.71652	51	2mo
13	▼ 1	Newsta		  	0.71315	183	1mo

Private leaderboard

#	△pub	Team Name	Kernel	Team Members	Score ?	Entries	Last
1	—	Bing Bai			0.74787	263	1mo
2	▲1	Magic Recommenders		 	0.74693	105	1mo
3	▼1	Rubtsov Vasily			0.74688	32	1mo
4	▲1	Devin			0.73680	58	1mo
5	▼1	ekffar			0.73666	148	1mo
6	—	EdZhang			0.73015	95	1mo
7	—	Chih-Ming			0.72669	82	2mo
8	—	rn51			0.72095	49	1mo
9	▲1	nPowered		  	0.71907	180	1mo
10	▼1	nanaya tachibana			0.71855	17	1mo
11	—	wehaveNOidea		 	0.71840	129	1mo
12	—	yese			0.71704	51	2mo
13	▲1	action			0.71264	78	1mo