

Task 2 : Debiasing

期末報告

組別：第六組

組員：M10815113 賴啓明

M10815103 陳子揚

B 10532037 王韋翔

B 10532030 楊博惟

指導老師：李漢銘 教授

團隊分工

- 楊博惟 – 程式撰寫、資料分析
- 陳子揚 – 程式撰寫、資料分析
- 賴啓明 – 程式撰寫、PPT 製作
- 王韋翔 – 程式撰寫、資訊收集

目錄

Part 1

方法介紹

成績

Part 2

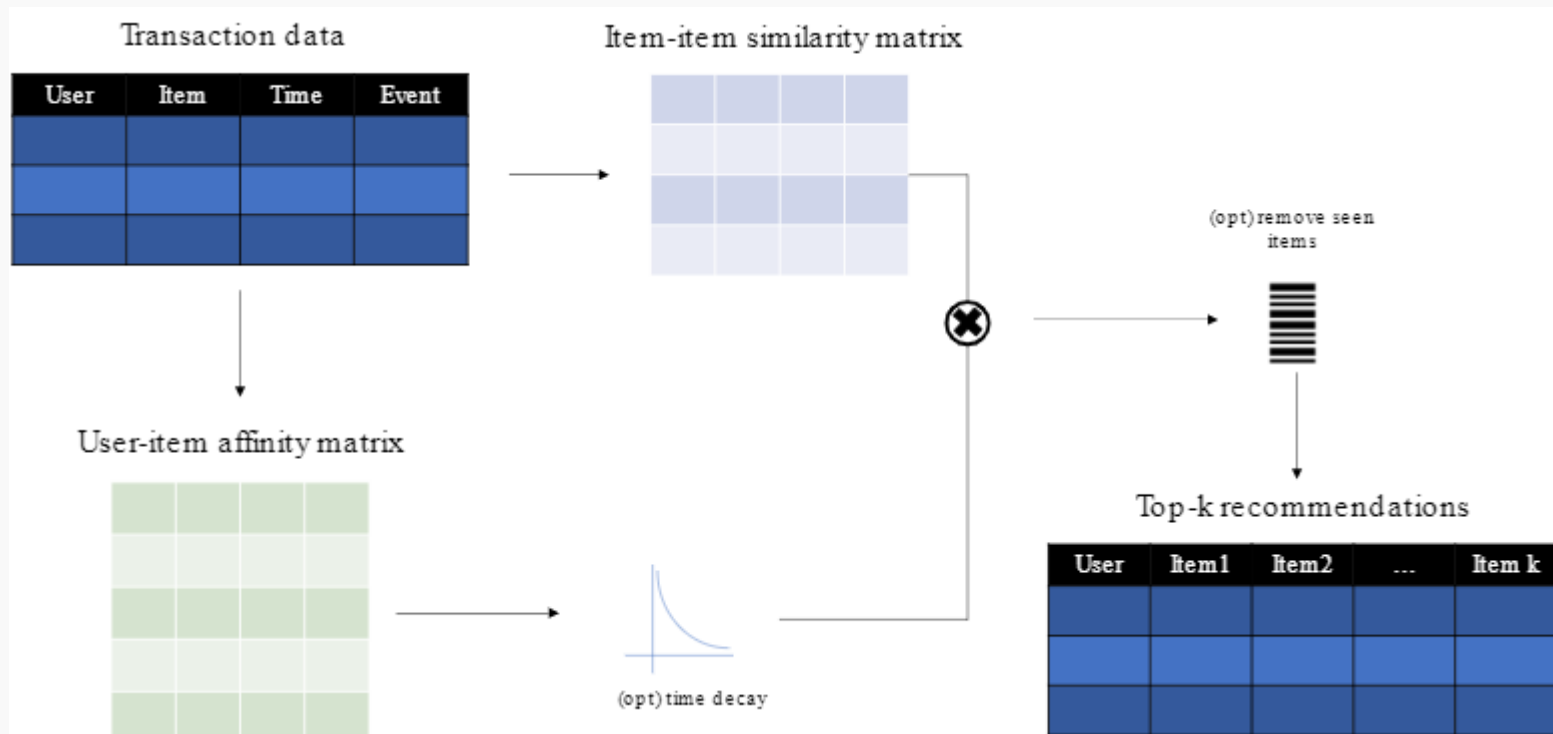
Part 3

結果分析

未來目標

Part 4

方法介紹



成績 (TrackA)

分數



5月7日

5月11日

5月16日

5月22日

5月25日

5月29日

6月03日

— 分數

成績 (TrackA)

98

👤 64

Others-1

0.308

0.662

0.308

0.441

(期中)Date : 2020/05/06 20:20

91↓²

👤 64

Others-1

0.481

1.036

0.481

0.752

0.351

(期末)Date : 2020/06/03 22:30

成績 (TrackB)

62	64	Others-1	7.173	7.388	7.173	7.299
----	----	----------	-------	-------	-------	-------

成績 (TrackB)

Our NDCG@50-full is among the top 10% (score rank < 65)
and the best NDCG@50-half is at **rank 25**.

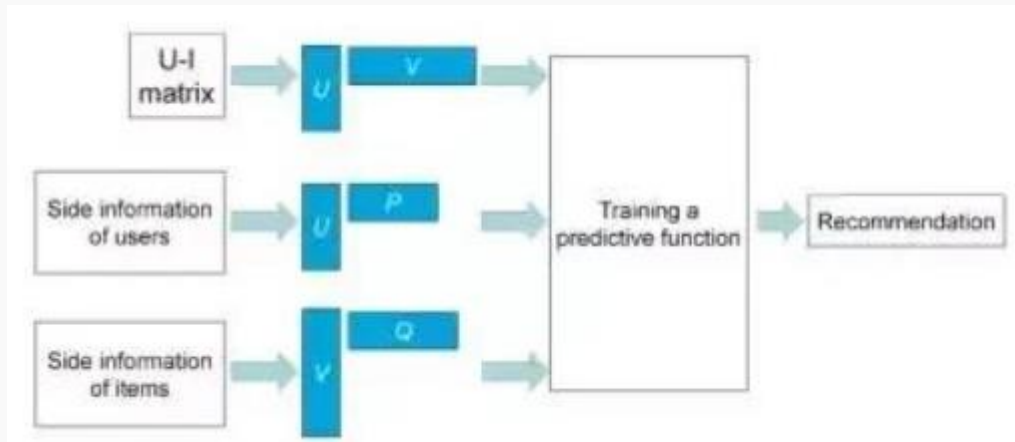
We are very close to the top 20 teams.

new_rank	score_rank	hitrate_full	ndcg_full	hitrate_half	ndcg_half	submit_time	team_name
1	1	7.6795	7.2977	7.9028	7.4516	2020-06-11 23:05:09	aister
2	14	7.6319	7.2494	7.8112	7.3920	2020-06-11 19:32:08	DeepWisdom
3	5	7.6321	7.2742	7.7187	7.3633	2020-06-11 19:17:25	TheAvengers
4	4	7.5423	7.2770	7.6969	7.3239	2020-06-11 18:29:18	hello dog
5	9	7.6246	7.2576	7.7180	7.3204	2020-06-11 20:58:14	ECNU@KDDCUP20
6	10	7.5966	7.2559	7.5372	7.3031	2020-06-11 18:48:57	Challengers
7	2	7.5862	7.2833	7.4903	7.2569	2020-06-06 23:40:31	MOH
8	11	7.5632	7.2521	7.5002	7.2365	2020-06-09 10:16:05	云儿是真名
9	6	7.5748	7.2679	7.5076	7.2333	2020-06-09 23:25:40	DB
10	3	7.6221	7.2774	7.4802	7.2261	2020-06-12 00:16:36	Rush
11	13	7.5941	7.2505	7.4910	7.2239	2020-06-08 13:17:13	千纸鹤L
12	18	7.5548	7.2241	7.4764	7.2045	2020-06-11 10:36:50	无名小队
13	15	7.5248	7.2401	7.4117	7.1943	2020-06-09 10:48:09	第一次打天才
14	8	7.5505	7.2628	7.3802	7.1928	2020-06-08 11:11:44	Who knows
15	7	7.5861	7.2657	7.4533	7.1890	2020-06-11 11:04:51	ML-Terminator
16	40	7.4527	7.1909	7.4226	7.1874	2020-06-06 10:39:37	build_success
17	19	7.5165	7.2160	7.3904	7.1753	2020-06-11 23:31:07	INNOVA-TSN/UCM
18	17	7.5032	7.2263	7.3674	7.1746	2020-06-11 15:55:38	前排大佬让一下
19	12	7.5457	7.2512	7.4189	7.1739	2020-06-11 22:50:57	聪聪呀
20	35	7.4024	7.1955	7.3118	7.1714	2020-06-07 23:30:20	一个人的战斗
21	63	7.3991	7.1720	7.3371	7.1664	2020-06-11 14:31:32	华开如梦
22	22	7.4477	7.2043	7.3516	7.1661	2020-06-06 06:58:04	少年加油
23	26	7.4314	7.1999	7.3094	7.1560	2020-06-08 21:26:09	最爱吃菠萝
24	44	7.4353	7.1864	7.3289	7.1399	2020-06-07 23:09:39	TCBigData
25	62	7.3884	7.1725	7.2994	7.1398	2020-06-11 21:19:33	64
26	20	7.4723	7.2082	7.3159	7.1397	2020-06-10 22:18:30	建议弃赛
27	21	7.4656	7.2044	7.3407	7.1386	2020-06-08 23:53:32	求求大家让我们
28	81	7.5055	7.1547	7.3378	7.1360	2020-06-11 21:16:17	LTL-Rec

方 法 改 進

- 考慮時間和先後順序
- 捨去較遠的點擊紀錄
- 去除只點一兩個物品的使用者

- Collective matrix factorization
- 效果不好，原因: user 及 item 的資料集不全



未 來 目 標

- 訓練不同的model並組合在一起，加強準確性
- Find user latent features