User Behavior Data on Taobao/Tmall IJCAI16 Contest

天池小喵萌2018-04-02240539

新建Notebook

内容

Notebook

评论

描述

This dataset is provided by Ant Financial Services  
IJCAI16 Contest: https://tianchi.aliyun.com/competition/entrance/231532/information

数据列表

* 数据名称上传日期大小下载
* IJCAI16\_data.zip2018-04-02258.09MB

文档

**Background**

As mobile devices become ubiquitous in our daily life, location based services (LBS) have become increasingly more important. People are getting more comfortable sharing their real-time locations with various location-based services, such as navigation, car ride hailing, restaurant/hotel booking, etc. As a result, huge amount of user data has been accumulated, which ignites the excitement from the machine learning/data mining community to join in force revealing the magic in the Pandora’s box of our daily life, where the high dimension time-space complexity needs to be explored. In this contest, we will focus on nearby store recommendation when users enter new areas they rarely visited in the past. The contest has two novelties: First, you are supposed to investigate whether the correlation between online and on-site preference helps in recommending nearby stores. Alibaba Group owns Taobao.com and Tmall.com, the largest online retail platforms in China, serves for more than 10 million merchants and over 300 million customers. Meanwhile, the Ant financial’s Alipay offers restaurant and retail store recommendation and payment services, named Koubei, for a number of customers. A user enjoying services provided by these two groups often has a unified online account. While Taobao and Tmall have run for many years and accumulate vast consumers’ behavior data, the nearby recommendation/payment services provided by Alipay are relatively new thus with less data. Second, a set of budget constraints is imposed on the recommender system, for example, due to service capacity or number of coupons available at the stores. As far as we know, such contest setting is novel to the research community, though it is critically important to the blooming location based business.

**Data Set**

In this contest, we aim to predict users' preference on Dec., 2015 (Table 4) based on his/her online/on-site behavior between July 1 st , 2015 and Nov. 30 th , 2015 (Table 1, 2). Moreover, budget constraints are imposed on each merchant (Table 3), simulating the limited discount/coupons available.

This dataset involves following data accumulated on Tmall.com/Taobao.com and the app Alipay.

Remark 1: Due to both business and noise concerns, we remove data in the great promotion period. That is, Nov. 01-Nov. 20 in Table 1 and Dec.12 in Table 4.   
Remark 2: Data are biased sampled from the daily log, thus its distribution would be different from the distribution of our entire business. Nevertheless, we believe it won’t affect too much on users’ preference prediction.

Table 1: Online user behavior before Dec. 2015. (ijcai2016\_taobao)

|  |  |
| --- | --- |
| **Field** | **Description** |
| User\_id | unique user id |
| Seller\_id | unique online seller id |
| Item\_id | unique item id |
| Category\_id | unique category id |
| Online\_Action\_id | “0” denotes “click” while “1” for “buy” |
| Time\_Stamp | date of the format “yyyymmdd” |

Table 2: Users’ shopping records at brick-and-mortar stores before Dec. 2015. (ijcai2016\_koubei\_train)

|  |  |
| --- | --- |
| **Field** | **Description** |
| User\_id | unique user id |
| Merchant\_id | unique merchant id |
| Location\_id | unique location id |
| Time\_Stamp | date of the format “yyyymmdd” |

Table 3: Merchant information. (ijcai2016\_merchant\_info)

|  |  |
| --- | --- |
| **Field** | **Description** |
| Merchant\_id | unique merchant id |
| Budget | budget constraints imposed on the merchant |
| Location\_id\_list | available location list, e.g. 1:356:89 |

Table 4: Prediction result. (ijcai2016\_koubei\_test)

|  |  |
| --- | --- |
| **Filed** | **Description** |
| User\_id | unique user id |
| Location\_id | unique location id |
| Merchant\_id\_list | you may recommend at most 10 merchants here, separated by “:”, e.g. 1:5:69 |

