Introduction tournament title

This applies to season1 tournament title game and season 2, season 3 of the online game (season 2 announced after the game title).
In Lynx, tens of millions of users every day through the brand find their favorite products, the brand is connecting consumers with goods most important link. Is the task of the tournament title four months according to the user's behavior log in Lynx, build brand preferences of users and to predict their future behavior within a month of purchase under the brand goods.

We will open the following data types:

Field	Field Description	Extraction Description
user_id	Users Mark	Sampling & Field Encryption
Time	Time behavior	Accuracy to-day level & hidden Year
action_type	Users of the brand type of behavior	Including clicks, purchase, add to cart, collecting four kinds of behavior (hits: 0 buy: 1 Favorite: 2 Shopping Cart: 3)
brand_id	Brand Digital ID	Sampling & Field Encryption

Users of any commodity behavior will be mapped to a row of data. All of which have been summarized as a commodity ID corresponding brand ID. Users and brands are done a certain degree of data sampling, and have done encrypted digital ID. All times are accurate to-day behavior level (hidden Year).

Evaluation indicators

We want teams predicted, the higher the better prediction accuracy rate brand, but also want to overwrite the user and the brand better, so with the most commonly used precision and recall rate as the list of targets.

Accuracy: Note: N is the number of teams to predict user pBrands ¡ i forecast for the user, he (she) will buy the brand list number hitBrands ¡ brand i predict a user's list of user i buy real intersection of a brand number $\text{precision} = \frac{\sum_i^N \text{hitBrands}_i}{\sum_i^N \text{pBrands}_i}$

Recall: Note: M is the number of users actually generated turnover of bBrands ¡ i true for users to buy the brand number hitBrands ¡ predict the number of users of the brand list and i buy real brands intersection $\text{Recall} = \frac{\sum_i^M \text{hitBrands}_i}{\sum_i^M \text{bBrands}_i}$

Finally, we use the F_1 -Score to fit precision and recall rate, and contest the final race results ranked by F_1 score prevail.

$$F_1 = \frac{2 * P * R}{P + R}$$

Using data on different seasons

Season 1 Season Statistics

1) data acquisition Format:
Participants via the contest's official website after successful registration, you can download the data directly to a local, direct use. (Download the contestants personal center)
2) Data Description:
original documents provided are around about 4M, involving about 1,000 users Lynx, Lynx thousands of brands, a total of more than 100,000 records of behavior. Users four kinds of behavior type (Type) corresponding codes are:
Hits: 0; purchase: 1; Collection: 2; cart: 3
3) Submit Rating Method:

Participants will predict the user into a text file in the following format:
user_id \ t brand_id, brand_id, brand_id \ N
(20 words or less) to upload the results file name limitation, the file must be txt format.

Figure:

a. txt

```
1 1232——>23423,234234,234234
2 1234——>1234,12312,123,1233,123345
3 1235——>344,5362,1324,56,743
4 2344——>3214,123
5 1238——>2134
```

The predicted results files uploaded to the official website of the results of the contest can be submitted to the entrance. (Entrance contestants submit the results of individual centers)

Season 2 Season Statistics

1) Data Access:

Participants can log in Alibaba Tianchi cluster access to data, and the data can not be downloaded, all data analysis, calculation, submit ratings will be completed in the Tianchi cluster. Log Tianchi cluster approach, please leave

Italy subsequent guidance documents and e-mail notifications.

2) Data Description:

the amount of data provided, involving ten million days cat users, 10,000 Lynx brand, behavior record time span of four months.

Tianchi training data provided on the cluster table t_alibaba_bigdata_user_brand_tota_1, the fields are: user_id, brand_id, type, visit_datetime. As shown in Figure

Field	Type	Label	Comment
user_id	string		
brand_id	string		
type	string		
visit_datetime	string		

Users four kinds of behavior type (Type) corresponding codes are:

Hits: 0; purchase: 1; Collection: 2; cart: 3

3) Submit Rating Method:

Participants completed the user brand preference predict the results need to be placed in the designated format data sheet (non-partitioned tables), requiring the results table named: t_tmall_add_user_brand_predict_dh,

contains user_id and brand two. Where brand preference field is predicted brand ID, across multiple brands (",") separated by commas.

For example:

user_id	brand
1000000	15397,22253,21082,20410
10000002	185,5300
10000005	22757,14020,2967,3747
10000008	732
10000011	3112
10000014	13779,8441,22204,8441
10000017	20890,3691,18158,1754
10000020	9822,9822
10000023	12379,4511,16526,10493
10000026	6936,2590