OLX - Code and the Curious

Rishabh Deshmukh(rishabhd123)

July 30, 2017

1 Training Dataset Quality

- Dataset is quite good, But I didn't get reason for not using English for title and description field in ads_data.csv.
- There are some ads in user_ads.csv which are not there in ads_data.csv.
- There is no use of ads which are not enabled.

2 Data Preprocessing Steps

- Remove all entries from ads_data.csv for which value of enabled is 0.
- Label Encode features event, origin & channel(Refer File Solution1.ipynb).
- Label Encode features user_id, ad_id & category_id(Refer File Solution_Metapath.ipynb).
- Extract month, day & hour from event_time.
- Sort User_data in descending order by= $['event\_month','event\_day','ad\_messages', 'ad\_views','ad\_impressions','images\_count','event']$
- Create augmented user_data as
  $a = augmented\_ads\_data.loc[np.array(user\_data['ad\_id'])]$

3 Recommender System

I have used ensemble of two simple models.

3.1 Meta-Path Based Model (Solution_Metapath.ipynb)

- This model is based on the idea of **Collaborative Filtering**.
- Meta-Path used in this model is user-ad-user-user-ad.
- This will give a matrix of dimension (nu, na) where nu=no. of users, na=no. of ads.

- Refer the code (cell-15 in file Solution-Metapath).
- This model will generate the output file with name sub_mat.csv.

3.2 General Sorting Based Model (Solution1.ipynb)

- This is very simple model based on just sorting and some basic manipulations.
- Sort User_data in descending order by= $['event\_month','event\_day','ad\_messages', 'ad\_views','ad\_impressions','images\_count','event']$.
- Find the top ads for each category(irrespective of the user_id) (Refer cell 13).
- Find the top ads for each user_id & category_id given in the file user_messages_test.csv(Refer cell 14).
- This model will generate the output file with name subm.csv.

3.3 Ensembling (Ensemble.ipynb)

For generating final submission file I took first four entries from subm.csv and last 6 entries from sub_mat.csv and generated the file ads_recommendation.csv.