

Coupon Challenge

Raul Sanchez-Vazquez

25 June 2018

The Problem

The problem:

- ▶ Create a recommender system for coupons.

This would be done in two stages:

- ▶ First, learn a model for classic top-n recommendation.
- ▶ Second, learn a binary classification model where the positive class correspond to actual purchase of the item.

Produce recommendation as:

- ▶ Produce candidate items (RecSys Model)
- ▶ Order by purchase probability (Positive probability score)

Analysis

We first built data profiling documents for:

- ▶ The item catalog.
- ▶ The user catalog.
- ▶ The user session log with purchase information.

We first noticed that the binary classification model would have to deal with class unbalance as shown on the plot below:

PURCHASE_FLG

Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.043198

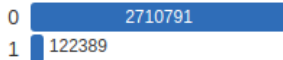


Figure 1

Features

Item features:

- ▶ itemd aged promo
- ▶ itemd nd views
- ▶ itemd nd purchases
- ▶ itemd area
- ▶ itemd ken
- ▶ itemd smalld area
- ▶ itemd kend mostd buy
- ▶ itemd kend mostd buy2
- ▶ itemd larged aread mostd buy
- ▶ itemd larged aread mostd buy2
- ▶ itemd smalld aread mostd buy
- ▶ itemd smalld aread mostd buy2
- ▶ itemd buyersd aged mean
- ▶ itemd buyersd aged median
- ▶ itemd buyersd aged std
- ▶ itemd DISCOUNTd PRICEd percentage
- ▶ itemd price
- ▶ itemd purchased sexd f
- ▶ itemd purchasesd sexd m

Features

User features:

- ▶ userd samed itemd nd views
- ▶ userd samed itemd nd purchases
- ▶ userd lastd itemd larged aread name
- ▶ userd lastd itemd kend name
- ▶ userd lastd itemd smalld aread name
- ▶ userd purchasesd samed itemd area
- ▶ userd purchasesd samed itemd kend name
- ▶ userd purchasesd samed itemd smalld area
- ▶ userd AGE
- ▶ userd SEXd ID

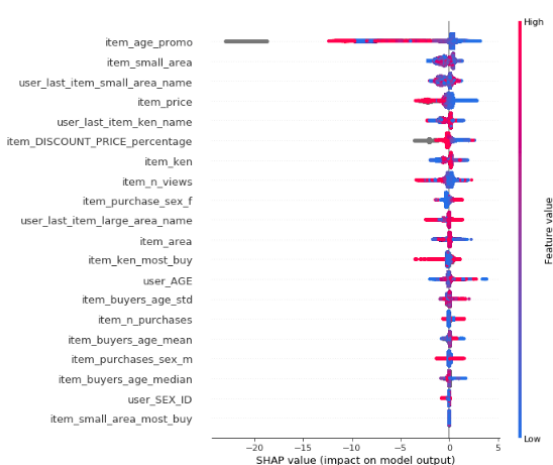
Model Performance

Recsys on Test:

- ▶ P@10: 0.0018949648

Binary Classification on Test:

- ▶ AUC:0.900254



Model Performance

Hybrid recommender:

- ▶ P@10: 0.029458179905312992