A Simple yet Efficient Method for a Credit Card Upselling Prediction

Peter Romov

Solution of ECML/PKDD 2016 Discovery Challenge
Task 2 “Upselling prediction”, team “Peter”

Yandex
Data Factory

Approach

User information

	AGE_CAT	LOC_CAT	INC_CAT	GEN	LOC_GEO_X	LOC_GEO_Y
USER_ID						
91498	b	c	b	1	602700	49100

User transactions

	USER_ID	POI_ID	CHANNEL	DATE	TIME_CAT	LOC_CAT	MC_CAT	CARD_CAT	AMT_CAT	GEO_X	GEO_Y
0	91498	28052	p	2014-01-01	a	b	b	d	b	605119.0	58997.8
157469	91498	26943	p	2014-01-08	b	c	e	d	a	602312.0	49289.0
2009139	91498	26943	p	2014-04-02	b	c	e	d	b	602312.0	49289.0
2934556	91498	26943	p	2014-05-14	a	c	e	d	b	602312.0	49289.0
3356486	91498	28052	p	2014-06-03	b	b	b	d	b	605119.0	58997.8
3410068	91498	14978	p	2014-06-05	b	b	e	d	c	603951.0	58409.1
4882432	91498	3419	p	2014-08-11	b	b	b	d	b	587174.0	77853.4
5018046	91498	28433	p	2014-08-16	b	b	a	d	c	590234.0	57492.4
5278427	91498	28052	p	2014-08-29	b	b	b	d	c	605119.0	58997.8
5638258	91498	10913	p	2014-09-13	a	b	a	d	b	591372.0	57098.2

Feature vector:

- Personal (15)
- Cards & Wealth (8)
- Activeness (8)
- Event counters (77)
- Geo (28)

Predict target
(credit card in the future)
using **XGBoost**

Features: User information

Personal features
- Gender
- Age category
- Location category
- Income category

One-hot & label encoded

Card & Wealth features

	C201401	C201402	C201403	C201404	C201405	C201406
USER_ID						
328787	0	1	1	0	0	0

- Number of months labeled as “1”
- Number of label changes “0” -> “1” and vice versa
- Last month labeled “1”

Features: Event counters

	USER_ID	POI_ID	CHANNEL	DATE	TIME_CAT	LOC_CAT	MC_CAT	CARD_CAT	AMT_CAT
0	91498	28052	p	2014-01-01	a	b	b	d	b
157469	91498	26943	p	2014-01-08	b	c	e	d	a
2009139	91498	26943	p	2014-04-02	b	c	e	d	b
2934556	91498	26943	p	2014-05-14	a	c	e	d	b
3356486	91498	28052	p	2014-06-03	b	b	b	d	b
3410068	91498	14978	p	2014-06-05	b	b	e	d	c

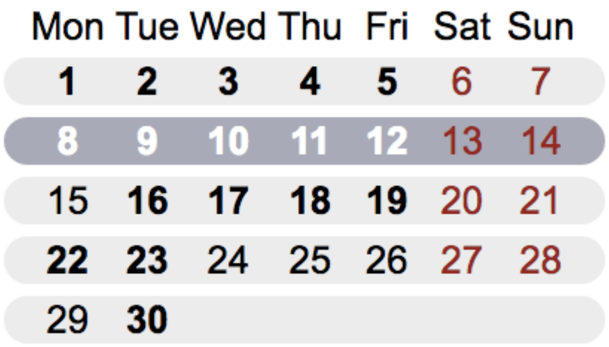
Two features for each category value:

- Number of events
  integer
- Ratio of events with particular value in all user events
  rational in [0, 1]

Categorical features in user transactions:

1. Type of activity (point of sale, webshop, branch)
2. Rounded time period (morning, daytime, evening)
3. Location type (capital, city, village)
4. Marker category of payment (7 anonymous categories)
5. Type of card (credit, debit)
6. Amount of money (low, medium, high)

Features: Activeness

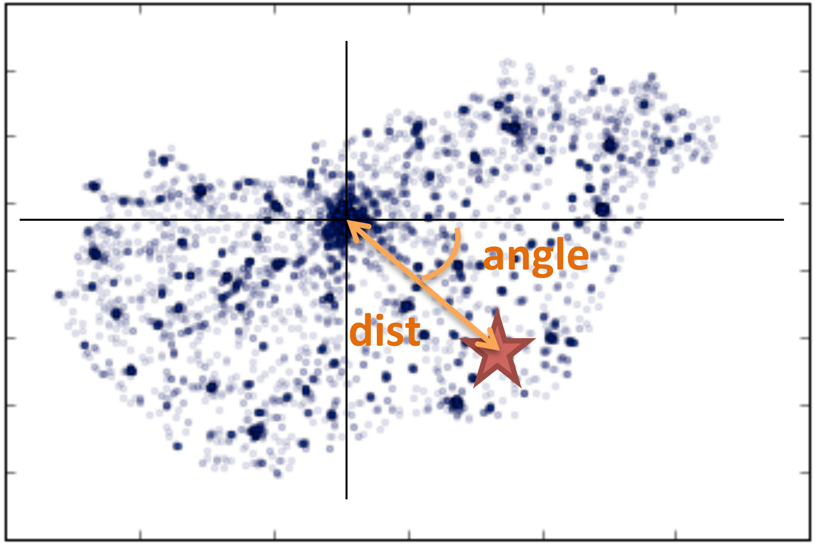


Active = at least one transaction during the period (days and weeks)

Features:

- Number of active / inactive periods
- Ratio of active periods

Features: Geo location



Geo-features of one transaction:

- coordinates
- distance to the center (Budapest)
- horizontal angle of location wrt the center

Aggregation of transaction geo-features:

- min, max
- mean, std
- percentiles (20%, 50%, 80%)

Tuning learning parameters

- Validation: 10x stratified shuffle split on learning (90%) and validation (10%)
- Parameters to tune
  - tree depth
  - learning rate
  - number of trees in ensemble
  - scheme of filling the missing values
  - number of unimportant features to exclude
- Decision
  - Marginal improvement in validation score (about 0.005 with big variance)
  - Biased validation scheme (because of year-to-year changes)
  - Final submission: <u>XGBoost model with default learning parameters</u> (Occam's Razor principle)
    100 trees, max depth = 3, learning rate = 0.1

Feature evaluation

Cross-validation AUC score: **0.7213** (stratified shuffle train/test split)

Feature group	**AUC change** **after removing feature group**	**AUC** **only features from the group**
Personal	-0.0322	0.6615
Cards & Wealth	-0.0137	0.5653
Event counters	-0.0019	0.6738
Activeness	-0.0012	0.6419
Geo location	-0.0004	0.6318

That's it!

Task 2: Upselling prediction

1st, **0.71862**
Team: achm

2nd, **0.71730**
Team: Cosine Vinny

3rd, **0.71589**
Team: Degrees of Freedom

4th, **0.71523**
Team: ISMLL

5th, **0.71479**
Team: TwoBM

6th, **0.71386**
Team: MMNF

7th, **0.71361**
Team: Peter

8th, **0.71341**
Team: Ya

Solution source code:
https://github.com/romovpa/ecmlpkdd2016-otp-bank-upselling