

Optimal couponing in retail: a bayesian approach

SARAPULOV G. V.

Saint-Petersburg State University

Faculty of mathematics and mechanics

g-eos@yandex.ru

May 16, 2018

Abstract

A problem of optimal couponing in retail is addressed by applying naive bayes classifiers to estimate a probability of item purchase with and without discount and choosing the available coupon with maximum difference between these estimates. It is experimentally shown that proposed methodology is able to significantly increase the conversion rate while maintaining high interpretability and scalability.

I. INTRODUCTION

Growing interest of business in using big data and machine learning leads to new decisions in different known issues. Personalized discount strategy is one of such tasks. Product retail, unlike other fields where recommender systems are actively used (streaming services, e-commerce, search engines etc.), has several traits that make standard approaches unapplicable. Firstly, explicit ratings of items are unavailable due to difficulties of collecting them, so only transactional data is available to inference user preferences. Secondly, it is not enough to recommend an item to a customer to stimulate purchase: customers are already aware of available items, have stable tastes and tend to repeatedly buy those they liked before, and discounts are usually used to increase sales. So a task of recommending an item naturally transforms to a task of offering the most relevant discount coupon.

In this article we propose an approach to couponing based on comparing the probabilities of item purchases with and without discount and choosing the coupon with maximum

difference between these probabilities. This method is highly interpretable (we offer a discount that increases the probability of item purchase most of all) and computationally effective, as it is based on naive assumptions of features conditional independence. Despite its simplicity, described approach is able to generate personalized offers with high conversion rate, as was shown experimentally, thus being particularly useful instrument for marketing tasks.

The article is structured as follows. Section II provides notation and formal definition of the problem. Section III presents a detailed description of optimal coupon choice based on purchase data. Experimental results are summarized in section IV.

II. FORMAL PROBLEM DEFINITION

Here we introduce a notation used further:

- U - a set of customers
- I - a set of items
- $C = \{(i, d)\}$ - a set of available coupons (pairs item-discount)
- $R = \{(t, u, i)\}$ - a set of transactions described as (date, customer-item)
- $D = \{(t, u, i, d, s)\}$ - a set of issued coupons described as (date, customer, item, discount, conversion flag)

Let's also denote $\mathbf{x}_u = \{x_{u1}, \dots, x_{uN}\} \in \mathbf{X}$ as feature vector of customer u , formed on transaction history R for last T days, where $x_{ui} = 1$, if customer u purchased item i , and $x_{ui} = 0$ otherwise. Analogously $\mathbf{y}_u = \{y_{u1}, \dots, y_{uN}\} \in \mathbf{Y}$ denotes the next transaction of a customer u . Each issued coupon from D is represented by vector $\mathbf{z}_u = \{z_{u1}, \dots, z_{uM}\} \in \mathbf{Z}$, where $z_{uc} = 1$, if customer u used coupon c .

Our aim is to provide a customer with personalized discount that would most likely lead to item purchase, while no being excessive: if the customer is likely to buy an item even without discount, then coupon will only lead to margin decrease.

Formally, based on previous transactions R and earlier issued coupons D for each customer $u \in U$ we choose an item i with discount d , i.e. coupon $c = (i, d)$ from a set of

available coupons C , such that the difference between probabilities of this item purchase with coupon $c = (i, d)$ and without discount is maximized:

$$c_u^* = \arg \max_{c \in C, c=(i,d)} P(z_{uc} = 1 | \mathbf{x}_u) - P(y_{ui} = 1 | \mathbf{x}_u) \quad (1)$$

III. PROPOSED APPROACH

To estimate these probabilities, two probabilistic models are needed. First one will be used to predict item purchases in the next transaction depending on content of all previous transactions of customer for period T . Second model will estimate probabilities of coupons conversion based on the same data as first model uses. After we obtain the probabilities of item i purchase and all the corresponding coupons conversion, i.e. coupons that provide some discount d for item i , we will be able to choose an optimal coupon for each customer.

We will use "naive" assumption that Bernoulli random variables x_{ui} are independent given \mathbf{y}_u :

$$P(x_{ui} | x_{uj}, \mathbf{y}_u) = P(x_{ui} | \mathbf{y}_u), i \neq j \quad (2)$$

For each item $i \in I$ we fit a naive bayes classifier $f_i : \mathbf{X} \rightarrow y_i$ to estimate a probability of item i purchase given the content of previous purchases:

$$P(y_i = 1 | \mathbf{x}_u) = \frac{P(y_i = 1) \cdot P(\mathbf{x}_u | y_i = 1)}{P(\mathbf{x}_u)} = \frac{P(y_i = 1) \cdot \prod_{j=1}^N P(x_{uj} | y_i = 1)}{\prod_{j=1}^N P(x_{uj})} \quad (3)$$

Similarly for each coupon $c \in C$ we fit a classifier $f_c : \mathbf{X} \rightarrow z_c$ to estimate the probability of coupon conversion:

$$P(z_c = 1 | \mathbf{x}_u) = \frac{P(z_c = 1) \cdot P(\mathbf{x}_u | z_c = 1)}{P(\mathbf{x}_u)} = \frac{P(z_c = 1) \cdot \prod_{j=1}^N P(x_{uj} | z_c = 1)}{\prod_{j=1}^N P(x_{uj})} \quad (4)$$

Optimal coupon is then obtained with the following formula:

$$c_u^* = \arg \max_{c \in C, c=(i,d)} f_c(\mathbf{x}_u) - f_i(\mathbf{x}_u) \quad (5)$$

IV. EXPERIMENTAL RESULTS

The described approach was tested on product chain stores. We used transactions of 30-days period before coupon issue to fit our models. The experiment was aimed to increase

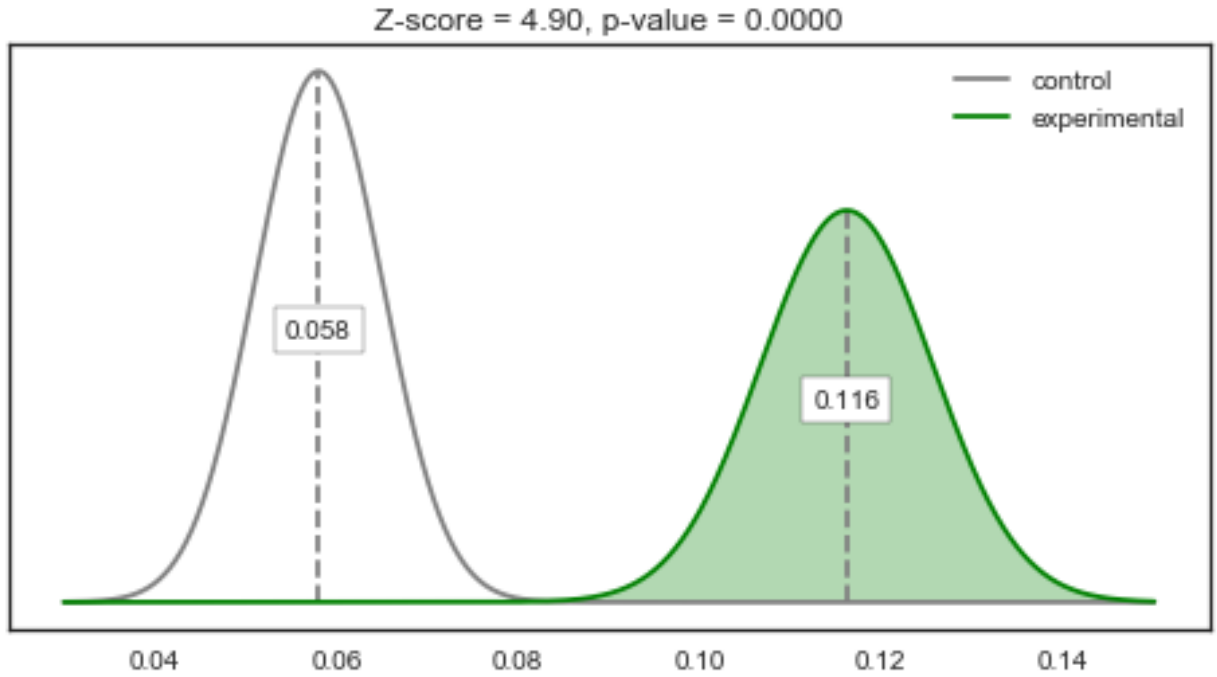


Figure 1: *Conversion rate in experimental groups*

conversion rate, that, in turn, can help increase other marketing indicators (in particular, average order sum and churn rate).

We used a split test to evaluate the conversion rate difference between experimental group, who got coupons provided by proposed method, and control group, who got regular coupons formed by existing recommender system amplified with business logic rules. Picture ?? represents conversion rates in customer groups. Fischer's Z-test shows that conversion rate of experimental group (11.6%) is higher than conversion rate of control group (5.8%) with a confidence level of 99%.

Obtained result is provided by relatively simple probabilistic models and can be improved, though even in current configuration it achieved better result in terms of conversion rate than masive business logic exploited earlier, that took a great amount of resources to be developed and is not hard to maintain. A described approach is flexible to use it for optimization of different metrics. Another advantage is the ability to incrementally update models with new data to adapt to changing internal (pricing policy, mass promotions) and external (seasonality, competition) factors.

REFERENCES

- [1] Yifan Hu , Yehuda Koren , Chris Volinsky, Collaborative Filtering for Implicit Feedback Datasets, Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, p.263-272, December 15-19, 2008
- [2] Ruslan Salakhutdinov, Andriy Mnih, Geoffrey Hinton, Restricted Boltzmann machines for collaborative filtering, Proceedings of the 24th international conference on Machine learning, p.791-798, June 20-24, 2007, Corvalis, Oregon, USA
- [3] Evangelia Christakopoulou , George Karypis, Local Item-Item Models For Top-N Recommendation, Proceedings of the 10th ACM Conference on Recommender Systems, September 15-19, 2016, Boston, Massachusetts, USA
- [4] Christopher M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer-Verlag New York, Inc., Secaucus, NJ, 2006