MWSUG 2017 – Paper BF03

Untangle Customers' Incrementality Using Uplift Modeling with a Case Study on Direct Marketing

Yi Cao, Alliance Data, Columbus, OH;
Chao Xu, Alliance Data, Columbus, OH;
Hairong Gu, Alliance Data, Columbus, OH

ABSTRACT

Uplifting model is very appealing in direct marketing campaign, mainly because instead of predicting absolute propensity, it predicts the uplift in propensities of response or (sales) driven by promotion. When it is not guaranteed that the incremental campaign ROI will be maximized if customers with the highest propensities of response or sales are targeted, and marketing budget is limited, uplifting model is a better choice than a propensity model as it helps markers to select customers who would shop because of a promotion, instead of those who shop with or without the promotion. Despite of its obvious merit in practice, uplifting model is challenging to develop in practice. In this paper, we build two uplift models-uplift response model and uplift sales model for different business objectives with potential challenges and remedies are discussed. The methods are validated in a real-life marketing campaign from one of our clients.

INTRODUCTION

The primary goal of most direct marketing is to drive change in customer behavior such as extra purchases. A broad consensus of campaign success measurement is ROI which requires calculation of incremental number of purchases or sales through careful use of control group.

However, most marketing activities today, even when measured on the basis of incremental impact, is utilizing conventional propensity model which seeks to characterize customers with highest likelihood of buying. Propensity model itself is not necessarily going to tell marketers which customers are most likely to contribute to incrementality. Different statistical model is needed, targeting the customers whose propensities of response or sales are dramatically driven by incentives.

Uplift modeling, also known as incremental modeling, is a predictive modeling technique that directly models the incremental impact of marketing activities on an individual's behavior. It uses a randomized control to build a predictive model that predicts the incremental response to the marketing action, thus it contributes directly to the success metric of a campaign.

However, uplift model is hard to construct practically and technically. From the practical perspective, firstly, to be able to model incrementality, a well-balanced experimental design with a test group and a control group need to be implemented. The control group needs to have sufficient sample size which, more often than not, might not be realistic because most campaigns tend to hold only a small portion (10% or 5%) as the control group. Secondly, the marketing incentive needs to be consistent each time to be able to model uplift. Otherwise, the predicted uplift might due to difference in incentives rather than customers' characteristics. Thirdly, seasonality would also be a confounding factor in the sense that campaigns executed on holiday seasons usually harvest high amount of sales or incrementality. Those reality complexities make the success of uplift model in sales really difficult. From a technical standpoint, there is no approach approved to be a norm in this domain. A most common approach is "two-model" approach, i.e. direct subtraction of models for the treated and untreated populations (Manahan (2005), Hansotia & Rukstales (2001)). However, some authors also pointed out that it tend to fail rather badly (Nicholas J. Radcliffe & Patrick D. Surry (2011)). Machine learning algorithms might be a remedy, such as K-Nearest Neighbors, Naïve Bayes (Larsen (2010)) and tree-based approach (Rzepakowski & Jaroszewicz (2010)). It is worth to notice that most of those explorations concentrate on predicting uplift in response rate, with little attempt on uplift in sales (Delali (2015)).

In this paper, we elaborate on the modeling practices behind uplift models for both response rate and sales and validate those modeling methods in a case study where an uplift response model and a sales model are developed that successfully identified customers who will generate additional transactions and sales. Meanwhile, we explore common challenges that keep us from building a well-performed uplifting sales model and give a few solutions to those obstacles. Even with vacuum of an uplifting sales model, the alternative is found to rank order uplift sales well. Conclusion is drawn at the end of the paper.

METHODOLOGY

EXPERIMENTAL DESIGN

Special concerns need to be taken in the period of experimental design to collect qualified data for the purpose of modeling uplift, either in response rate or sales. Essentially, the pursuit of modeling uplift is to select those customers who potentially render the largest increase to the campaign ROI. However, because the same customers cannot be in both a test group (with an offer) and a control group (without an offer) simultaneously in a campaign, different customers need to be selected for the two groups.

To ensure that any observed differences in the campaign outcome are due to the campaign itself instead of intrinsic differences between selected customers in the test and control groups, the demography of the two groups should be kept as homogeneous as possible. Therefore, random sampling or subsampling within segments of population are recommended to select customers into the two groups. It is also recommended to have sufficiently large sample size in each group to improve the accuracy in model fitting and validation.

MODELING

The dataset for modeling can be represented as a matrix with rows representing customers and columns representing observed variables. For each customer i, observed variables can be noted by $\{Y_i, S_i, X_i, T_i\}$, in which Y_i represents whether the offer is accepted by values of 0 and 1; S_i is the amount of sales; X_i is a vector of predictive variables; and T_i indicates whether the customer is in the test ($T_i = 1$) or control group ($T_i = 0$).

A cross validation (CV) procedure is recommended for model building and validation. That is, the total dataset is divided by rows into a few subsets. Each subset is used once as the validation sample with the rest subsets combined as the model fitting sample. The CV procedure iterates through all subsets and aggregates the prediction error as the index for model performance. In practice, to alleviate the computational complexity, the CV procedure can be replaced by a simpler procedure that uses about 70% of data for model fitting and the rest of data for validation.

MODELING UPLIFT OF RESPONSE RATE

If the model of response rate is represented as $P = E(Y|X, T)$, for customer i, her uplift of response rate is the difference of response rates between the scenario when she was in the test group and the scenario when she was in the control group, which is

$$\Delta P_i = P_i|Test - P_i|Control = E(Y_i|X_i, T_i = 1) - E(Y_i|X_i, T_i = 0). \qquad (1)$$

Many modeling approaches can be applied to build the model $P = E(Y|X, T)$. One common approach is logistic regression. The functional form of logistic regression can be written as

$$log\,\frac{P}{1-P} = \alpha + \beta X + \gamma T + \delta XT, \qquad (2)$$

which is similar to a linear regression with the response variable replaced by the logarithmic transformation of the odds ratio $\frac{P}{1-P}$. The interaction term XT is indispensable in the model for the purpose of calculating the uplift in Equation 1 because without this term the predicted uplift ΔP_i for every customer will be exactly the same.

We propose to use the following procedure to build and validate the model for the uplift of response rate:

1. In SAS®, apply the Logistic procedure or the GLM procedure to obtain parameter estimates in Equation 2. Store the parameter estimates.
2. To validate, first, replace all T_i's in the validation dataset with 1, score the dataset to obtain $P_i|Test$, and then do the procedure again with all T_i's replaced by 0 and score to obtain $P_i|Control$. Calculate $\Delta P_i = P_i|Test - P_i|Control$ as the uplift for each customer.
3. Rank the customers by their uplifts into ten deciles. Within each decile, calculate difference between the average response rate of the test group and that of the control group, and use it as the empirical uplift for that decile. If the model performed well, the empirical uplifts should also be in a ranked order with the predicted order.

MODELING UPLIFT OF SALES

In the same fashion, if a customer i's sales is modeled by $S_i = E(S_i|X_i, T_i)$, her uplift of sales can be represented as

$$\Delta S_i = S_i|Test - S_i|Control = E(S_i|X_i, T_i = 1) - E(S_i|X_i, T_i = 0) \qquad (3)$$

Then the problem boils down to the modeling of sales $S = E(S|X, T)$. One intrinsic challenge in building a sales model is illustrated by a hypothetical example in Figure 1. The challenge is that usually a substantial proportion of customers do not have any purchases during campaign promotion period, resulting in large portion of zeros in sales. Due to pike of zeros, the model fitness can be very bad, as shown by the red line.

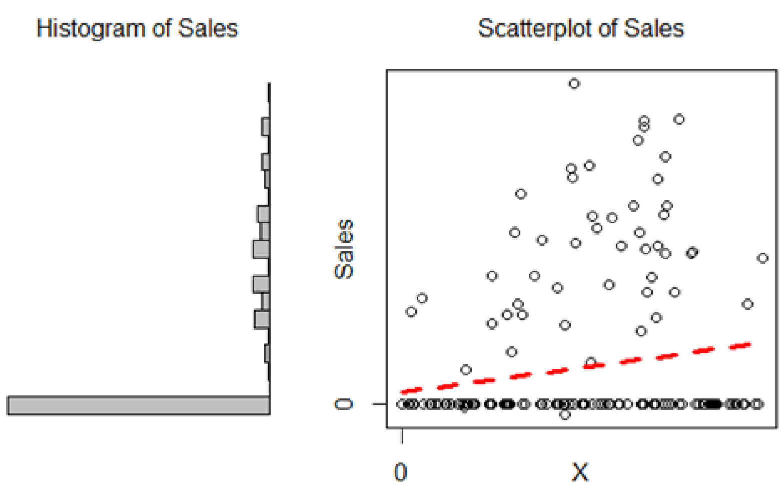

Figure 1: A hypothetical example of the histogram and scatterplot of sales with a large proportion of customers having zeros

In light of that, the traditional modeling approaches such as generalized linear regression cannot take into account excess zeros. Therefore, we consider a machine learning approach, random forest, which is particularly powerful at modeling nonlinear effects in data. A brief introduction of random forest is given below.

The general method of random forests was first proposed by Ho in 1995, and was extended by Leo Breiman and Adele Cutler. Particularly useful for classification, regression, it adopts an ensemble learning approach that is shown to perform superior than a single decision tree model. Given a training set Ɗ, random forest repeatedly selects random samples from the training set, $Ɗ_1, Ɗ_2, \dots, Ɗ_n$, and fits a decision tree model to each sample. Then, the prediction of a new data point X can be gathered by taking the average or the mode of the predictions from each decision tree model.

In the present study, random forest is used to build the model $S = E(S|X, T)$. To further estimate the uplift of sales, first predict the sales of each customer under the treat group by assuming all T_i being 1, and under the control group by assuming all T_i being 0. Then calculate the uplift of each customer ΔS by using Equation 3. To validate, rank the customers by their predicted uplifts into deciles. Within each decile, calculate the difference between the average sales of the test and control group as the empirical uplift of sales. If the model performs well, the empirical uplift of each decile should also follow the ordered ranks.

CASE STUDY

To make it relevant to real-life business scenarios, the data in this study are from an upscale furniture chain store. In these campaigns, $25 off %50 gift cards is issued for loyalty members. The client is not only interested in identifying customers who are more likely to redeem the gift card but also looking for opportunities to maximize campaign ROI given the limited budget. They had been running the same campaign for one and half years thus accrued sufficient amount of data. Table 1 is the description of data set and attributes.

# Loyal customers	427,559
% offered $25 off $50 GC	34%
% not offered anything	66%
Data Attributes	Recency
	Member Tenure
	Credit Limit
	Last year trips
	Last year sales
	Income
	…

Table 1: Data Description

Total available data contain more than 50 attributes ranging from sales, trips, customer demographics to loyalty credit card balance and credit limit information etc. and we also assume that loyalty credit card is their primarily way of making purchases.

RESPNOSE UPLIFT MODEL RESULTS

This section is organized as follows: firstly, results of traditional response model are presented and investigated; then, an uplift response model is proposed for driving shopping traffic.

Aligned with the first goal of the campaign to select customers, a response propensity model using logistic regression was built in SAS®. With the model, a validation dataset was scored and split into 10 deciles based on the predicted propensities. To validate this model, the average response rate in each decile is calculated as shown in Figure 2a. As desired, the empirical uplift in response rate monotonously decreases with the model ranks.

To understand whether customers with high response propensities also contribute to large uplift in response rate, the differences between the average response rate between the test and control groups within deciles from the response model is calculated/ As shown in Figure 2b, customers with the largest response propensities are not those who generate highest incrementaility. Therefore, a shopping propensity model succeeds in selecting customers with high inclination to shop but not with largest uplift in terms of response rate.

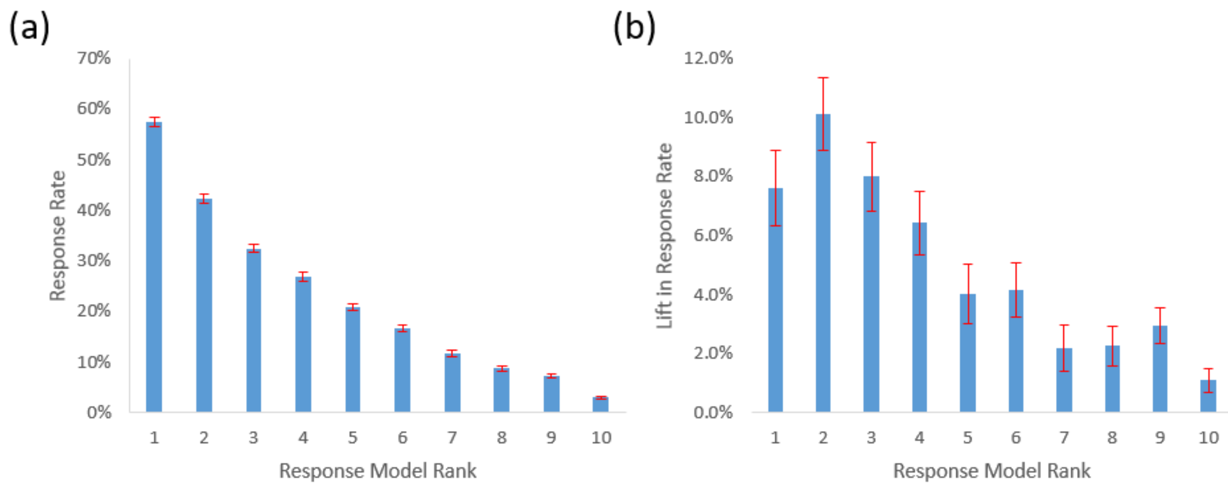

Figure 2. (a) Response Rate and (b) Lift in Response Rate vs. Response Model Rank

To maximize the impact of a marketing campaign, an uplift response model is built to aid us in targeting customers who are most likely to be triggered by promotion. As proposed in Methodology section, after model is built, we rank the customers into 10 deciles according to the predicted uplift sales. Figure 3 shows empirical uplift in response rate, calculated by the difference between the average response rate between test and control groups, within each decile. We can clearly see that the uplift response model successfully identified those customers with the largest uplift in response rate. The rank 1 has as much as 13% lift in response rate and it is indeed the most drivable decile.

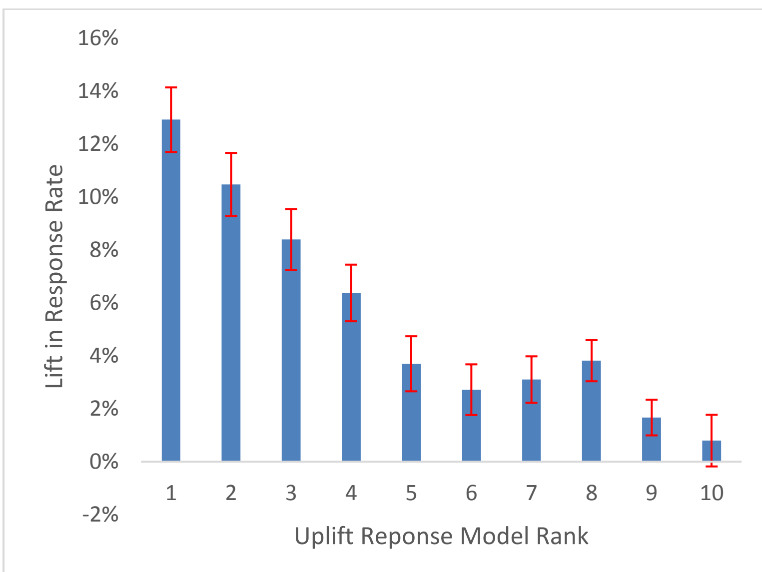

Figure 3. Lift in Response Rate vs. Uplift Response Model Rank

SALES UPLIFT MODEL RESULTS

Another goal of the campaign is to increase sales. To inspect whether the response model renders any information to the uplift in sales, the empirical uplift in sales, calculated by the difference of average sales between the test and the control groups, within the deciles of the response model are plotted in Figure 4.

The unordered values show that customers who have higher propensity to shop do not necessarily purchase more upon the receipt of the offer than others. Thus, to meet the goal of driving incremental sales, an uplift model in sales is needed.

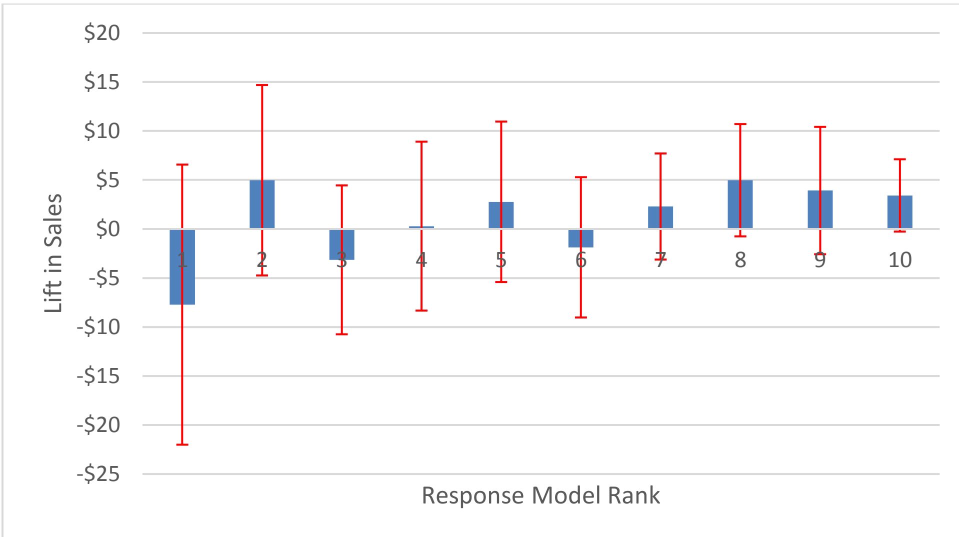

Figure 4. Lift in Sales vs. Response Model Rank

After extensive modeling efforts, the proposed method in the Methodology section for modeling uplift in sales proves unsatisfactory. One reason is due to the large variabilities in sales, making predictions quite inaccurate. Therefore, the estimated uplift, calculated by the deduction of two predictions of sales, becomes even bemired.

Instead of building a sales model, a more direct approach to identifying customers with large uplift in sales is to search for sub-areas in the design space where the difference of the sales between the test and the control groups is large. To facilitate finding those sub-areas, it is useful to first inspect each predictor variable for their relationship to the uplift in sales. After initial data exploration, one such variable found is the sales in the observation window (labeled as obs_sales_avg). Figure 5(a) shows the moving average of sales for the test (black) and the control group (red) along with the variable obs_sales_avg. It appears that the larger the sales in the observation window, the larger the uplift of sales is potentially be. If this observation is reliable, we can use sales in the observation window as the handle to rank uplift in sales. To validate this method, we used the predicted sales to bucket the customers in the validation set into 5 ranks, and then calculated the sales of the test group and the control group in each rank. The difference between the sales of the two groups can be used as the estimate the uplift in that group. As shown in Figure 5(b), the method successfully identified those customers with the largest uplift in sales,

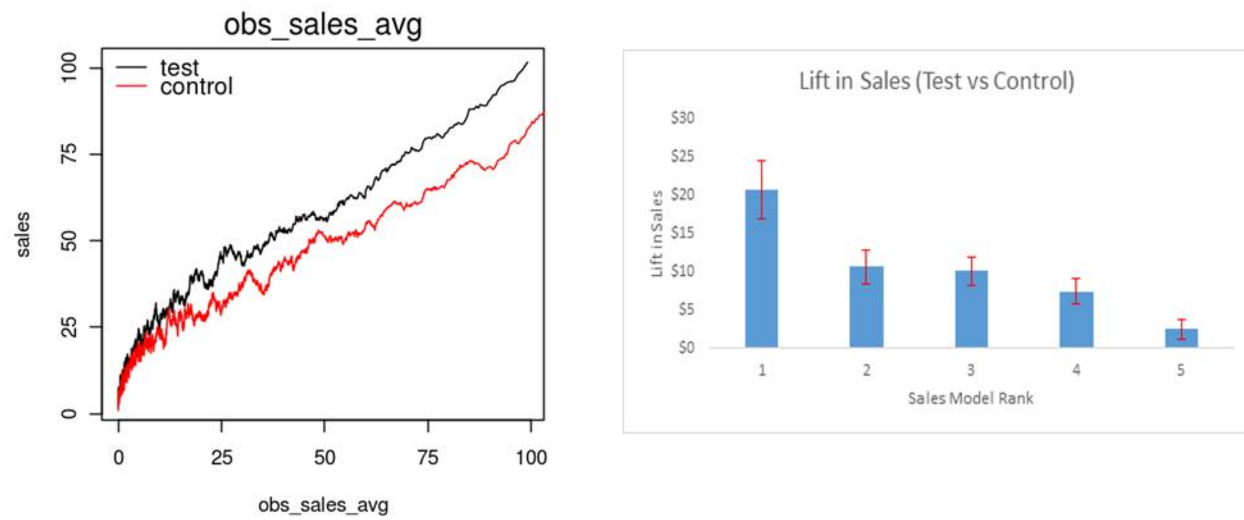

Figure 5. (a. left hand) Moving average of sales for the test and control group along with the predictor variable obs_sales_avg; (b. right hand) Validation of the uplift model by plotting the observed uplift with the predicted ranks of uplift

The underlying idea of this approach is that there are sub-areas in the design space where the uplift of sales are larger compared to other areas. To estimate the uplift in those areas, it is necessary to have a sufficient number of customers in both the test and the control group in that area. Otherwise, the estimates of the uplift in that area may not be accurate. In the current demonstration, only one predictor variable obs_sales_avg is used to construct those sub-areas. To extend from this example, other informative variables can also be included for the construction of sub-areas. However, it becomes more difficult to ensure enough sample size in high-dimensional sub-spaces. Therefore, the method needs further development for high-dimensional cases.

CONCLUSION

We prove through a case study that simply selecting customers with the largest response rate does not necessarily guarantee maximizing incrementality. Empowered by the data from a well-designed, real-world marketing example, the case study successfully developed statistical procedures that identified customers whose response rate or sales have the largest boost due to promotion. The result showed that customers who are predicted to have the largest response rate, and those who have the largest uplift in response rate, and those who have the largest uplift in sales do not overlap. Therefore, marketers need to think in advance business priorities in order to decide which aspect of lift is of most concern. The paper addresses those issues that arise during the process of developing uplifting models and the corresponding solutions. Therefore, the procedure can serve as a reference for the future efforts for those pursuing incrementality of marketing campaigns.

REFERENCES

1. C. Manahan, 2005. A proportional hazards approach to campaign list selection. In SAS® User Group International (SUGI) 30 Proceedings.

2. B. Hansotia and B. Rukstales, 2001. Incremental value modeling. In DMA Research Council Journal, pages 1–11.

3. P. D. Surry and N. J. Radcliffe, 2011. Quality measures for uplift models. Submitted to KDD2011.

4. K. Larsen, 2010. Net lift models.

5. P. Rzepakowski and S. Jaroszewicz, 2010. Decision trees for uplift modeling. IEEE Conference on Data Mining, pages 441–450.

6. Ho, Tin Kam (1995). *Random Decision Forests* (PDF). Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995. pp. 278–282.

7. D. Agbenyegah (2015). Get the highest bangs for your marketing bucks using incremental response models in SAS® Enterprise Miner. MWSUG2015.

ACKNOWLEDGMENTS

We would like to profusely thank Alliance Data Predictive Analytics team to provide us opportunity to present our work at MWSUG2017 conference.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Yi Cao
Enterprise: Alliance Data Card Services
Address: 3100 Easton Square PL, Columbus, Ohio, 43219
Email: yi.cao@alliancedata.com