Uplift Modeling for Cost Effective Coupon Marketing in C-to-C E-commerce

Akihiro Shimizu*†, Riku Togashi*‡, Antony Lam*, Nam Van Huynh†
*Mercari, Inc.
†Japan Advanced Institute of Science and Technology
‡Waseda University
{akihiro-shimizu, riktor, antonylam}@mercari.com,
huynh@jaist.ac.jp

***Abstract*—E-commerce companies often provide marketing incentives such as price discount coupons to motivate new customers to make their first purchase. However, many customers make purchases only when coupons are distributed to them; they stop making purchases after using the coupons. Thus, for cost-effective marketing, it is desirable for companies to distribute marketing coupons to new customers that have the highest potential to make future purchases without continued coupon incentives. However, it is difficult for e-commerce companies to identify the new customers to be targeted within 30 hours of registration. In this study, we address this problem using uplift modeling for cost-effective marketing. Uplift modeling can be used to identify the time when there is a causal relationship between coupon distribution and future non-coupon purchases. The ability to identify these causal relationships can allow a company to distribute coupons to the most promising customers and improve its business. Several studies have explained the benefits of uplift modeling in real-world e-commerce businesses. In this study, we demonstrate the results of uplift modeling for coupon distribution in a real-world Customer-to-Customer (C-to-C) e-commerce platform. We show that uplift modeling decreases marketing costs by 39.0% with only a negligible reduction in the number of acquired customers who make non-coupon purchases.**

***Index Terms*—uplift modeling, causal inference, machine learning, ad coupon targeting**

I. Introduction

To make marketing cost-effective, appropriate customer segments must be targeted. For example, most e-commerce companies distribute price discount coupons to certain new customers for motivating first purchases. Specifically, companies may aim to distribute coupons to new customers who will make non-coupon purchases in the future because this will benefit the business. On the other hand, distributing coupons to customers who will only make a single purchase with their coupon will not be cost-effective. To make marketing more cost-effective, a company needs to identify the customers that are most likely to become loyal customers in the future and ideally, only distribute coupons to these customers.

To achieve these goals, a common practice in industry involves adding distribution conditions based on the experience of marketers and distributing coupons to the appropriate customer segments. For example, a company can have multiple coupons with different conditions on distribution:

- **Coupon A**, Condition 1;
  - Distribute the coupon to customers who looked at products more than five times.
- **Coupon B**, Condition 1 + Condition 2 + Condition 3;
  - Distribute the coupon to customers who looked at products more than five times.
  - Launched the e-commerce application of the company more than ten times.
  - Gave one or more comment.

*All conditions are checked within 30 hours of registration.

In general, considering Coupons A and B, targeting tends to become more accurate when more distribution conditions are added. However, when considering two or more conditions, companies must pay higher costs for tasks that include verifying all possible sets of conditions and conducting many A/B tests to compare the sets of distribution conditions.

Moreover, to make marketing cost-effective, it is important to understand the causal relationship between coupon distribution and future non-coupon purchases. In other words, companies should distribute coupons to customers only if the coupons have a high probability of converting these customers into regular customers in the future and will make purchases even without coupons.

When considering the impact of coupons on purchases, we can consider the four customer segments shown in Fig. 1 to identify the customer segments that should be targeted. The best segment to be targeted is Segment A. Segment A will make purchases only if it receives coupons. In contrast, distributing coupons to Segments B and C may not be as cost effective. There is no need to distribute coupons to Segment B because such customers tend to make no purchases irrespective of the coupons. Further, Segment C may not need coupons to incentivize purchases. Moreover, businesses should not distribute coupons to Segment D because interestingly, such customers do not make purchases if coupons are distributed to them but they make purchases without coupons. An example of Segment D type behavior includes mistakenly distributing coupons to males for the purchase of products for women. The customers could be annoyed by receiving too many irrelevant coupons and ultimately cancel their membership.

Therefore, businesses should distribute coupons to customer

2375-0197/19/$31.00 ©2019 IEEE
DOI 10.1109/ICTAI.2019.00259

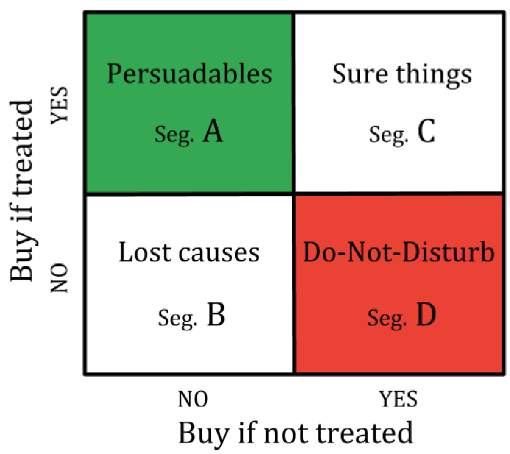

Fig. 1. Four customer segments that may receive coupons. These segments are considered along two axes showing the distribution of coupons and purchasing behavior.

segments such as Segment A because it would be the most cost effective. In this paper, we show that uplift modeling [1]–[3] can be used to effectively determine the most promising customer segments (for example, Segment A). We demonstrate through real-world tests that uplift modeling is a causal inference method that facilitates effective identification of causal relationships between purchases made with and without coupons.

Our contributions are as follows:

- Uplift modeling can be used to improve the cost effectiveness of coupon distribution.
- We ultimately show, in a real-world e-commerce business, a cost reduction of 39.0% can be achieved with only a negligible loss to the number of non-coupon purchase user acquisitions.

II. Related Work

Many data analysis methods have been used to improve the cost-effectiveness of digital marketing [4]–[6]. In particular, response models are often used for analyzing the cost-effectiveness of marketing [7]–[9]. For example, we can calculate the probability $P(Buy|NoCoupon)$ using response models. It is quite important to know the probability of a customer making a non-coupon purchase in making business decisions. We can improve prospect targeting using this probability [9], [10].

However, response models offer limited capabilities for analyzing cost-effectiveness. In the case of response models, we cannot confirm the causal relationship between coupon distribution and future non-coupon purchases [11], [12]. To tackle this problem, many studies based on uplift modeling have been conducted to improve the cost-effectiveness of marketing [1]–[3]. Unlike response models, uplift modeling does not predict the customer with the highest purchase probability. However, it predicts a score that determines whether a customer should receive coupons. Uplift modeling predicts a score that is obtained by considering the difference between the probabilities (see 1) of purchases made with and without coupons

$$S = P(Buy|Coupon) - P(Buy|NoCoupon) \tag{1}$$

Many previous studies have focused on improving the cost-effectiveness of marketing using uplift modeling [13]–[15]. However, only a few studies have focused on incorporating uplift modeling into real-world marketing decisions to improve cost-effectiveness. Therefore, in this study, we investigate the effectiveness of uplift modeling used in a real-world setting. Specifically, the business metric we use to measure performance is the total marketing cost for motivating customers to make non-coupon purchases in the future.

III. Method

A. Uplift Modeling

Uplift **U(x)** can be expressed using Pearl's causal inference framework [16], defined as follows:

$$U(x) = E[Y|X = x, T = 1] - E[Y|X = x, T = 0] \tag{2}$$

where E is the expected value of the purchase and $Y \in \{0, 1\}$ represents whether a purchase is made ($Y = 1$) or not ($Y = 0$). Furthermore, $T \in \{0, 1\}$ represents whether treatment was performed. For example, $T = 1$ indicates that a coupon was distributed to customers. X includes the multiple variables related to customer behavior; for example, the number of views (x_1) or the number of times the shopping application was launched (x_2).

The uplift **U(x)** can be calculated by obtaining two conditional probabilities (Eq. 3). These two probabilities are calculated for the cases of when treatment is performed, and when it is not performed.

We would also need data from A/B tests to calculate the uplift. In an A/B test, customers from the same population are divided into two groups. One group is called the test (treatment) group T_R, and the other is called the control (non-treatment) group C. Uplift **U(x)** can be expressed as follows:

$$U(x) = P_{T_R}(Y = 1|X = x) - P_C(Y = 1|X = x) \tag{3}$$

The simplest approach for uplift modeling is the two-model approach. In this approach, two models are required to compute the probabilities in the uplift model (Eq. 3). To estimate these probabilities, algorithms such as RandomForest [17] and Xgboost [18] can be used.

B. Class Variable Transformation

While the two-model approach can be used for uplift modeling, it may be cumbersome to build two models, and the results tend to be difficult to interpret. A solution to this issue is to use the class variable transformation proposed by Jaskowski and Jaroszewicz [19]. The following class variable transformation allows us to calculate uplift with only one model:

$$Z = \begin{cases} 1 \;\; if \;\; T = 1 \;\; and \;\; Y = 1 \\ 1 \;\; if \;\; T = 0 \;\; and \;\; Y = 0 \\ 0 \;\; otherwise. \end{cases} \tag{4}$$

By introducing this transformation, the uplift can be obtained as follows:

$$\begin{aligned} U &= P_{T_R}(Y = 1|X) - P_C(Y = 1|X) \\ &= 2P(Z = 1|X) - 1. \end{aligned} \tag{5}$$

Note that this transformation is based on the assumptions, $P(G = T_R) = P(G = C) = 1$ and that the outcome variable is binary.

This method predicts the uplift using only one model, however, it assumes that control and test groups contain the same number of customers and that makes it difficult to use in practice. Athey et al. [20] proposed a method that solves this problem:

$$Y_i^* = Y_i(1)\frac{T_i}{\hat{P}(T_i)} - Y_0(1)\frac{1 - T_i}{1 - \hat{P}(X_i)} \tag{6}$$

where $\hat{p}(x)$ is a consistent estimator of the propensity score. They solved the class transformation problem by correcting the objective variable with the propensity score.

IV. Experiments and Results

A. Problem

In this study, we aim to make marketing more cost-effective using uplift modeling; we also aim to motivate customers to make non-coupon purchases in the future after their first purchase using a coupon. Therefore, we predict whether coupons should be distributed to a customer. In addition, such predictions need to be made before the customers' first purchases.

To describe our approach, we first define an objective variable Y. We modify the definition of the objective variable (Sec. III) to be suited to our situation as follows:

$$Y = \begin{cases} 1 \;\; if \;\; T = 1 \;\; and \\ \qquad PurchaseWithoutCoupon = 1 \\ 1 \;\; if \;\; T = 0 \;\; and \;\; FirstPurchase = 1 \\ 0 \;\; otherwise. \end{cases} \tag{7}$$

Since the purpose of coupons is to incentivize non-coupon purchases, it is essential that customers will make non-coupon purchases even after coupon distribution. Moreover, in the case of the treatment group, the objective variable is positive if the customer purchases without a coupon, regardless of whether they used the coupon they received.

After this transformation, the usual class transformation to Z and Y^* is performed; thus the standard uplift modeling that was introduced in Sec. III can be used.

The objective variables Y_{days} were set to 3, 10, and 20 days after coupon distribution, and the date that was the most cost-effective was determined. We created separate models to predict Y_{day} for each time period. Table I shows an example of predictions for each time period.

TABLE I
Predictions of Y_{days}

Customer	Y_{3days}	Y_{10days}	Y_{20days}
A	True	True	True
B	False	True	True
C	False	False	False
...	...	...	...

B. Datasets

The dataset used in this study was based on customer log data from a C-to-C e-commerce application. We utilized this dataset to develop the input features for uplift modeling. Table II shows an example of the input features taken from the dataset.

TABLE II
Example of Dataset Format

Customer	app launch	viewed items	...
A	1	11	...
B	2	5	...
C	2	21	...
...	...	...	...

The dataset contains aggregated customer activity that was monitored during the first 30 hours after each customer registration, and we collected this data for 26 days. For example, we recorded occurrences such as the user opening the app two times and viewing 31 items within 30 hours of signing up.

We also created two groups in this dataset for the experiment. We created a treatment group (coupons were distributed) and a control group (coupons were not distributed). Aside from the initial coupon distribution, each group received no other coupons during this experiment, and each was randomly selected from the same population. All the customers had not bought anything until the day coupons were distributed.

C. Models

We tested several models for uplift calculation and determined the model that was the most effective in this problem. The following models were tested:

- **SGD**, Regularized linear models with stochastic gradient descent (SGD) learning
- **RF**, Random Forest
- **XGB**, XGBoost.

We used the Scikit-learn [21] library in Python to create the SGD, RF, and XGB [18] models. The Pylift library was used for uplift modeling in this experiment. Using this library, we directly calculated the uplift score instead of using the two-model approach.

D. Evaluation

The Qini coefficient and Qini curve [11], [15] are commonly used to evaluate uplift models. The Qini curve formula is defined as a parametric curve with the following equation:

$$g(t) = Y_t^{T_R} - \frac{Y_t^C N_t^{T_R}}{N_t^C} \tag{8}$$

where $g(t)$ is calculated for the first t observations and sorted by uplift value in descending order. The Qini coefficient is defined as follows:

$$f(t) = \frac{g(t)(N_t^{T_R} + N_t^C)}{N_t^{T_R}}. \tag{9}$$

E. Results

TABLE III
EXPERIMENT RESULTS

Model	20 days	10 days	3 days
SGD	0.078	0.101	0.1419
RF	0.05	0.082	0.1723
XGB	0.071	0.101	0.1700

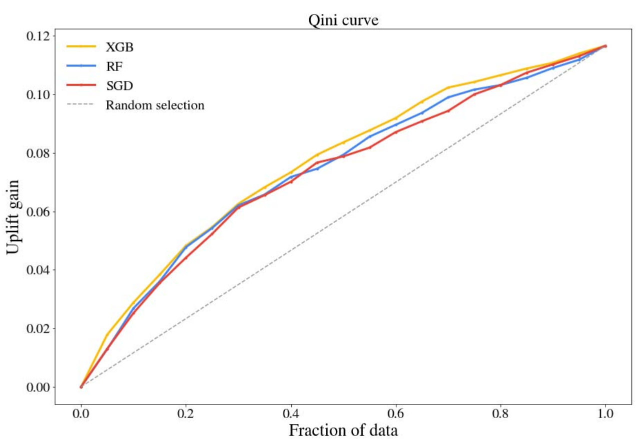

Fig. 2. Qini curve results when forecasting non-coupon purchasing within 3 days. The results for each model are plotted.

We conducted training for the treatment and control groups using 80% of the data (randomly sampled), and the remaining 20% was used for testing. Table III and Fig. 2 show experimental results on the test data. The results show that the coupon was most effective when uplift was maximized after three days. The XGB model was used to gather these results; we confirmed that the XGB model estimates each objective variable sufficiently well, and thus decided to test this model with a real business.

F. Evaluation Using Business Metrics

Our goal was to optimize the impact of the distributed coupon. Therefore, we conducted A/B tests with an e-commerce application to measure the impact of uplift modeling. For this reason, we formed the following groups from the total population of customers:

- **Group A**, customers that never received coupons;
- **Group B**, customers that received the coupons randomly;
- **Group UM**, customers that were chosen to receive the coupons based on uplift modeling.

Then, we subdivided Group UM into two half groups to measure the effect of uplift modeling.

- **Group C**, customers that received coupons based on uplift modeling.
- **Group D**, customers that did not receive any coupons based on uplift modeling.

After we formed these groups, we distributed the coupon to Groups B and C on the same day. Then, we calculated the *non-coupon purchase rate* within 20, 10, and 3 days, and the *coupon usage rate* (Eq. 10) in each group.

TABLE IV
RATE OF PURCHASE WITHOUT ADDITIONAL COUPONS DISTRIBUTED WITHIN n DAYS; COMPARISON BETWEEN GROUPS A, B, AND C

Segments	20 days	10 days	3 days
Never Received Coupons	15.8%	12.1%	6.2%
Random	37.1%	34.5%	30.1%
Uplift Modeling - Received Coupons	49.4%	46.1%	41.0%

We see that the non-coupon purchase rates within n days for Groups A and B are compared in Table IV. There is a difference of 21.3% in the purchase rates within 20 days for each group, thus indicating that coupons have a large influence on non-coupon purchases.

Table IV also shows comparisons between Groups B and C. Our results indicate that customers targeted with uplift modeling make non-coupon purchases more than using random coupon distribution.

TABLE V
PURCHASING RATE WITHOUT ANY COUPONS DISTRIBUTED WITHIN n DAYS FOR GROUPS A AND D

Segments	20 days	10 days	3 days
Never Received Coupons	15.8%	12.1%	6.2%
Uplift Modeling - Never Received Coupons	22.8%	19.2%	12.3%

The non-coupon purchase rates within n days for Groups A and D are compared in Table V. The results indicate that Group D is more likely to make a non-coupon purchase than Group A. In other words, we can predict which customers tend to make non-coupon purchases using uplift modeling.

TABLE VI
COUPON USAGE AND PURCHASE RATE WITHIN 20 DAYS

Segments	Random	Uplift (Group UM)
Usage Rate	28.7%	17.5%
Purchase Rate	37.1%	36.1%

Finally, the non-coupon purchase rate and coupon usage rate (Eq. 10) within 20 days for Groups B and UM (C and D) are compared in Table VI. It can be seen that the purchase rates differ by only 1.0%, which is a difference does not affect the business. On the other hand, the percentage of cost usage (coupon usage) decreased by 11.2%. In brief, using uplift modeling yielded a 39.0% cost reduction (see Eq. 10) with only a negligible reduction in the number of acquired customers that make non-coupon purchases.

$$\begin{aligned} &NPR(\%) = N_P/N \\ &UR(\%) = U/N \\ &CR(\%) = 1 - (UR_B/UR_{UM}) \end{aligned} \tag{10}$$

where:
N = the number of customers in a group
N_P = the number of non-coupon purchase customers in a group
NPR = non-coupon purchase rate
U = the number of usage coupon in a group
UR = usage rate of a group (ex: UR_B = usage rate of Group B)
CR = cost reduction.

V. CONCLUSION

In this study, we used uplift modeling to achieve a cost reduction of 39.0% with only a negligible reduction (1.0%) in the number of acquired customers who make non-coupon purchases. Cost reduction received considerable attention in this study but other important business metrics such as lifetime value and return on investment can be used in the future to change the objective variable of uplift modeling. After having determined who the coupons should be distributed to, we plan to address the question of when and what needs to be done for marketing using machine learning and causal inference.

CONFLICT OF INTEREST

The project described in the study was concluded at the end of the experimental phase and thus no commercial, or production deployment of the methods presented in the study ever took place.

ACKNOWLEDGMENT

We would like to thank Editage (www.editage.com), Mr. Kimura, Dr. Yamaguchi, and Mr. Ueta for English language editing.

REFERENCES

[1] P. Gutierrez and J.-Y. Gérardy, "Causal inference and uplift modelling: A review of the literature," in *International Conference on Predictive Applications and APIs*, 2017, pp. 1–13.
[2] R. Gubela, A. Beque, F. Gebert, and S. Lessmann, "Conversion uplift in e-commerce: A systematic benchmark of modeling strategies," *International Journal of Information Technology & Decision Making*, 2019.
[3] F. Devriendt, D. Moldovan, and W. Verbeke, "A literature survey and experimental evaluation of the state-of-the-art in uplift modeling: A stepping stone toward the development of prescriptive analytics," *Big Data*, vol. 6, no. 1, pp. 13–41, 2018, pMID: 29570415. [Online]. Available: https://doi.org/10.1089/big.2017.0104
[4] M. Berry and G. Linoff, *Mastering data mining: The art and science of customer relationship management*. John Wiley & Sons, Inc., 1999.
[5] M. J. Berry and G. S. Linoff, *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons, 2004.
[6] E. Y. Huang and C.-j. Tsui, "Assessing customer retention in b2c electronic commerce: an empirical study," *Journal of Marketing Analytics*, vol. 4, no. 4, pp. 172–185, Dec 2016. [Online]. Available: https://doi.org/10.1057/s41270-016-0007-x
[7] P. Baecke and D. Van den Poel, "Improving purchasing behavior predictions by data augmentation with situational variables," *International Journal of Information Technology & Decision Making*, vol. 9, no. 06, pp. 853–872, 2010.
[8] Z. CHEN, "From data mining to behavior mining," *International Journal of Information Technology & Decision Making*, vol. 05, no. 04, pp. 703–711, 2006. [Online]. Available: https://doi.org/10.1142/S0219622006002271
[9] K. Coussement, P. Harrigan, and D. F. Benoit, "Improving direct mail targeting through customer response modeling," *Expert Syst. Appl.*, vol. 42, no. 22, pp. 8403–8412, Dec. 2015. [Online]. Available: https://doi.org/10.1016/j.eswa.2015.06.054
[10] V. S. Y. Lo, "The true lift model: A novel data mining approach to response modeling in database marketing," *SIGKDD Explor. Newsl.*, vol. 4, no. 2, pp. 78–86, Dec. 2002. [Online]. Available: http://doi.acm.org/10.1145/772862.772872
[11] N. Radcliffe, "Using control groups to target on predicted lift: Building and assessing uplift model," *Direct Marketing Analytics Journal*, pp. 14–21, 2007.
[12] P. Rzepakowski and S. Jaroszewicz, "Uplift modeling in direct marketing," *Journal of Telecommunications and Information Technology*, pp. 43–50, 2012.
[13] V. Lo, "From predictive uplift modeling to prescriptive uplift analytics: A practical approach to treatment optimization while accounting for estimation risk," *Journal of Marketing Analytics*, vol. 3, 08 2015.
[14] R. Gubela, S. Lessmann, J. Haupt, A. Baumann, T. Radmer, and F. Gebert, "Revenue uplift modeling," in *International Conference on Information Systems (ICIS)*, 12 2017.
[15] Diemert Eustache, Betlei Artem, C. Renaudin, and A. Massih-Reza, "A large scale benchmark for uplift modeling," in *Proceedings of the AdKDD and TargetAd Workshop, KDD, London,United Kingdom, August, 20, 2018*. ACM, 2018.
[16] J. Pearl, *Causality*. Cambridge university press, 2009.
[17] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
[18] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. ACM, 2016, pp. 785–794.
[19] M. Jaskowski and S. Jaroszewicz, "Uplift modeling for clinical trial data," 2012.
[20] S. Athey and G. W. Imbens, "Machine learning for estimating heterogeneous causal effects," 2015, Research Papers. [Online]. Available: https://EconPapers.repec.org/RePEc:ecl:stabus:3350
[21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.