IPRPD
International Journal of Business & Management Studies
ISSN 2694-1430 (Print), 2694-1449 (Online)
Volume 05; Issue no 07: July, 2024
DOI: 10.56734/ijbms.v5n7a1

CAUSAL MACHINE LEARNING IN MARKETING

Martin Huber, Ph.D.[1]

[1]*Professor of Applied Econometrics – Policy Evaluation, University of Fribourg*

Abstract

This study reviews three primary purposes of causal machine learning (CML) in marketing, merging impact evaluation of marketing interventions with machine learning algorithms for learning statistical patterns from data. Firstly, CML enables more credible impact evaluation by considering important control variables that simultaneously influence the intervention and business outcomes (such as sales) in a data-driven manner. Secondly, it facilitates the data-driven detection of customer segments for which a marketing intervention is particularly effective or ineffective, a process known as effect moderation or heterogeneity analysis. Thirdly, closely related to the second point, it allows for optimal customer segmentation into groups that should and should not be targeted by the intervention to maximize overall effectiveness. The discussion is grounded in recent empirical applications, all of which aim to enhance decision support in marketing by leveraging data-driven evaluation and optimization of interventions across different customer groups.

Keywords

Causal Machine Learning, Marketing, Algorithms, Impact Evaluation

1 Introduction

In recent decades, firms have seen a surge in customer data availability, spawning a growing literature on machine learning (ML) in business and marketing for predicting customer behavior like purchase decisions, price sensitivity, and churn, see e.g. Xia, Chatterjee, and May (2019), Hu, Dang, and Chintagunta (2019)), Donnelly, Ruiz, Blei, and Athey (2021), Arevalillo (2021), and Gordini and Veglio (2017). Predictive ML typically lacks insights into the causal effects of specific interventions, like a marketing campaign or a pricing policy, on business outcomes. For instance, when multiple predictor variables (like customer education or income) are similarly capable of capturing specific patterns in the outcome (like purchasing behavior), ML may prioritize some predictors over others in a way that does not reflect the effects of those predictors on the outcome. To address this limitation of predictive ML, a growing literature has developed causal machine learning (CML) algorithms that are tailored to impact evaluation, see for instance Athey and Imbens (2016), Belloni, Chernozhukov, Fernández-Val, and Hansen (2017), Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins (2018), and Athey, Tibshirani, and Wager (2019). Unlike predictive ML, these algorithms are capable of properly assessing the causal effects of marketing interventions, provided certain statistical conditions hold. This capability is crucial for decision support, particularly in determining the viability of implementing a marketing intervention and identifying the optimal customer segments for its implementation.

This study reviews three primary purposes of applying CML in marketing based on recent empirical applications. Firstly, CML enables more credible impact evaluation of marketing interventions by controlling in a data-driven way for important confounders, variables that simultaneously influence the intervention and the outcome. Not properly controlling for confounders entails a biased impact evaluation in which the effect of the marketing intervention on the outcome is (at least to some extent) mixed up with the effects of those confounders on the outcome. Secondly, CML can be applied for effect heterogeneity (or moderation) analysis to detect customer segments for which a marketing intervention is particularly effective or ineffective, see for instance Chernozhukov, Demirer, Duflo, and Fernandez-Val (2018). Thirdly, and closely related to the previous idea, CML permits finding an optimal customer segmentation (based on observed characteristics like customer's income) into groups that should and should not be targeted by the intervention, in order to maximize its overall effectiveness, as discussed in Athey and Wager (2021).

2 Controlling for covariates by causal machine learning

Evaluating the impact of interventions has a long tradition in marketing research, often through experiments with randomly assigned interventions, as seen in Bawa and Shoemaker (1989) (evaluating coupons), Fong, Fang, and Luo (2015) (evaluating discounts), and Gordon, Moakler, and Zettelmeyer (2022) (investigating online ads). However, in many contexts, experimental data is unavailable, leading to reliance on observational data where interventions are not random. In such cases, interventions are typically susceptible to selection bias, meaning that their receipt may depend on customer characteristics that also affect the outcome. Considering, for example, the effect of a loyalty card on purchasing behavior, it might be observed that customers with a loyalty purchase on average more from a store than those without. However, this association may not reflect the true impact of the loyalty card. In fact, past buyers may be more inclined to acquire a loyalty card, such that past purchases affect possession of the card and may also drive current purchases, thus contaminating impact evaluation of the loyalty card on current purchases.

This example is illustrated in Figure 1, which presents a causal graph, see e.g. Pearl (2000), which represents impacts between variables by arrows. Such selection bias issues (e.g., due to differences in past purchases of customers with and without loyalty cards) can be addressed if the analyst can observe those characteristics that reflect all factors that simultaneously affect the choice of the intervention and the outcome. In this case, these characteristics, also known as covariates, satisfy an assumption known as selection-on-observables, see Imbens (2004). This permits evaluating the intervention by comparing the outcomes of customer groups with an without intervention who are similar in terms of covariates, such as age or income. Controlling for covariates in this way avoids confusion between an intervention's impact and any effects of differences in covariates across the groups.

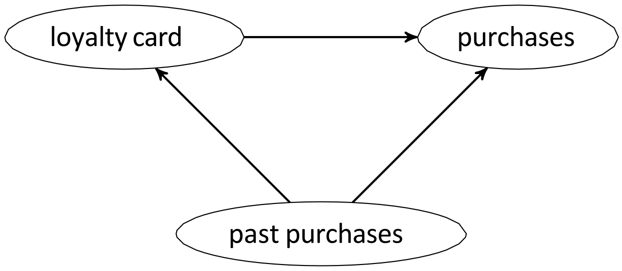

Figure 1: Treatment selection bias when evaluating a loyalty card

Relying on the selection-on-observables assumption involves determining which covariates to control for to ensure comparability between groups receiving and not receiving the intervention prior to its evaluation, which requires thorough domain knowledge. However, in many evaluations, particularly in big data contexts with a large number of variables, it is unclear which covariates are most crucial. Provided that the covariates are informative enough for the selection-on-observables assumption to hold, an unknown subset of covariates might be sufficient to satisfy this assumption. It then appears attractive to have a method that learns from the data which subset should be controlled for. Huber, Meier, and Wallimann (2022), for instance, assess the effect of train discounts in Switzerland on demand shifts, measured as rescheduling train trips to a different time than initially planned. They observe many demand-related factors that may affect both the discount intervention and the demand shift outcome, including train class, month, weekday, day time, public holidays, distance, and also trip-specific points of departure and arrival. Each combination of departure and arrival could be considered as separate covariate, entailing a number of potential control variables that is large compared to their database of roughly 6000 customers. This motivates the use of causal machine learning (CML) to control for covariates importantly affecting the intervention or the outcome in a data-driven manner.

CML involves initially learning statistical models for the intervention and the outcome as functions of the covariates through predictive ML. These models are then incorporated into a so-called doubly robust (DR) statistical function for impact evaluation, as outlined in Robins, Rotnitzky, and Zhao (1994). One popular combination of ML and DR is termed Double Machine Learning (DML), see Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins (2018), and can under certain conditions perform similarly well to cases where the analyst already knows the important covariates to be controlled for a priori. These conditions include not having an excessively large number of crucial covariates that significantly affect the intervention or outcome relative to the number of observations in the data (otherwise, it can be difficult for ML to identify all important covariates).

Huber, Meier, and Wallimann (2022) use an CML algorithm named causal forest, see Wager and Athey (2018) and Athey, Tibshirani, and Wager (2019), to assess the discount rate's effect on the demand shift outcome. The findings indicate that a one percentage point increase in the discount rate results in a 0.16 percentage point increase in the share of rescheduled trips, a statistically significant effect at the 5% level. Additionally, the authors examine a binary definition of the intervention by comparing discounts of 30% or more versus less than 30%, using DML to estimate the average effect of the binary discount. The results suggest that discounts of 30% or more, on average, increase the number of demand shifts by 3.8 percentage points compared to lower discounts. When a more traditional impact evaluation method, namely matching, see Rosenbaum and Rubin (1983), the estimated impact has a substantially higher variance. Therefore, the data-driven selection of important covariates through CML algorithms may enhance the precision of impact evaluation by reducing its variance.

Langen and Huber (2023) apply CML to evaluate the average impact of various types of coupons issued by a retailer on purchases per customer, considering a dataset of roughly 50,000 observations provided by the retailer as part of the AmExpert 2019 Machine Learning Hackathon, see AML (2019). The data include socio-economic customer characteristics like age, marital status, family size, and income, details on coupons received, coupon redemption, and the purchases made by registered customers between January 2012 and July 2013. The retailer ran multiple coupon campaigns in which customers received between 0 to 37 different coupons that offered discounts for individual items or even a range of products. The results indicate that receiving any coupon has a positive average effect during a campaign period, increasing expected daily expenditures by approximately 60 monetary units per customer. Furthermore, the estimated effects are heterogeneous across different coupon categories, with coupons for drugstore items and other food increasing daily spending by around 80 and 78 monetary units, respectively, and coupons for non-food products decreasing spending by approximately 24 monetary units (likely due to cannibalization effects), all statistically significant.

While Huber, Meier, and Wallimann (2022) and Langen and Huber (2023) exemplify the application of CML for assessing the average impact of marketing interventions, it cannot be overstated that the appropriateness of such analyses depends on the satisfaction of the selection-on-observables assumption. To empirically verify this, Gordon, Moakler, and Zettelmeyer (2022) compared the results of 663 experiments conducted between November 2019 and March 2020 on Facebook, evaluating the impact of ads on conversion outcomes such as purchases, page views, and downloads, with corresponding assessments based on CML and also matching using observational (i.e., non-experimental) data. The experiments encompassed approximately 7.9 billion user observations and over 38 billion ad impressions, with the median ad experiment running for 30 days, involving about 7.3 million users across intervention and control groups. The conversion outcomes considered include variables occurring both earlier and later in a hypothetical purchase funnel. For instance, page views occur early in the purchase funnel, adding items to a cart occurs later, and purchase occurs last. The median experimental impacts on the upper, middle, and lower funnel outcomes amount to increases in conversion rates (measured relative to the conversion rate without ad) of 29%, 18%, and 5%, respectively.

To estimate the impacts in the observational data, the applied CML and matching methods utilize a rich set of covariates, including user characteristics (e.g. age, gender, number of friends, number of ad impressions in the last 28 days, personal interests), estimated intervention-specific conversion probabilities, and pre-intervention outcome variables (past conversion activities). If the selection-onobservables assumption holds and CML or matching properly control for all important factors jointly affecting ad assignments and outcomes, the estimated impacts should generally align closely with the experimental effects. However, using CML (or matching), the median impacts on the upper, middle, and lower funnel outcomes are 83% (173%), 58% (176%), and 24% (64%), respectively, indicating significant upward biases. Thus, although CML outperforms the more traditional matching method, it still notably overestimates the impacts when compared to experimental results. This highlights that CML is not a panacea, its accuracy crucially depends on data quality.

3 Effect heterogeneity

The previous section discussed evaluating the average impacts of marketing interventions in a population of interest, such as all customers. However, since customers are not homogeneous, some may respond differently to certain interventions than others. This motivates effect heterogeneity or moderation analysis, which involves investigating how effects differ across values of covariates, like customer characteristics, to determine which customer segments the intervention is particularly effective or ineffective for. These characteristics could be predetermined by the analyst, such as considering age as relevant covariate to examine whether the effectiveness of the intervention differs across both younger and older customers. Alternatively, a data-driven approach can be used to explore which (a priori unknown) set of covariates is most predictive of effect size and thus, the heterogeneity of effects.

For example, Langen and Huber (2023) examine how providing coupons affects the purchasing behavior of various customer groups defined by predetermined covariates. Specifically, they investigate how the average impact of providing (any) coupons depends on customers' age, income, family size, and pre-campaign expenditures. To achieve this, they follow Semenova and Chernozhukov (2021) and employ linear regression of the previously mentioned DR functions, which represent the expected impact as a function of the covariates, on indicators for membership in specific groups based on age, income, and family size. One finding of Langen and Huber (2023) is that the impact is more pronounced among customers who made either no or rather large purchases in the period prior to the campaign. This is illustrated in Figure 2, which depicts the average impacts across different customer groups defined in terms of past purchases, with dots representing the impact estimates and bands representing the 95% confidence intervals.

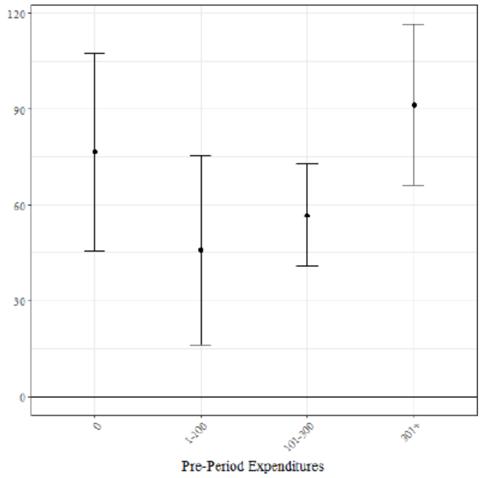

Figure 2: Effect heterogeneity across past purchases

Huber, Meier, and Wallimann (2022) examine heterogeneity in the effect of discounts for train tickets on demand shifts across a predetermined set of covariates that are relevant for characterizing customers and their travel purposes. These include age, gender, travel distance, indicators for leisure trips and commuting (with business trips as the reference category), and traveling during peak hours. Huber, Meier, and Wallimann (2022) find no significant effect heterogeneities across certain covariates like age, gender, or distance. However, they observe heterogeneity in the effect of leisure trips, indicating that, all else being equal, a one-percentage-point increase in the discount rate increases the share of rescheduled trips by 0.29 percentage points more among leisure travelers than among business travelers. This result suggests that leisure travelers may be more flexible in terms of timing compared to business travelers.

Moreover, Huber, Meier, and Wallimann (2022) explore effect heterogeneity in a data-driven manner, employing ML to identify covariates that most strongly predict the effect size of the intervention. To achieve this, the previously mentioned DR function is predicted as a function of all covariates using an ML algorithm like a random forest, see Ho (1995) and Breiman (2001) . Subsequently, covariates are ranked based on their importance for predicting effect heterogeneity, e.g., based on the coefficient estimates in a lasso regression. Across values of more important covariates, the impact of the intervention varies more substantially compared to less important covariates. Huber, Meier, and Wallimann (2022) find that demand-related characteristics such as seat capacity, utilization, departure time, and distance are the most influential predictors of effect size, with customer age also demonstrating some predictive power.

4 Optimal policy learning

Assessing effect heterogeneity is the foundation for optimal policy learning, which aims to optimally assign an intervention across customer segments depending on the expected effects within segments, such that overall effectiveness is maximized, see for instance Kitagawa and Tetenov (2018). This implies that some customer segments with particularly high effects might be targeted by the intervention, while others are not. For instance, Langen and Huber (2023) apply the optimal policy tree algorithm suggested in Athey and Wager (2021) to learn which customer segments should be optimally targeted by coupon campaigns in a way that the overall purchases by customers are maximized. Based on a predefined number of customer segments, this method conducts a data-driven customer segmentation in which only those segments with sufficiently high effects are targeted by coupon campaigns and visually displays the customer segmentation along with the intervention assignment using a decision tree.

Figure 3 illustrates such a policy tree for the optimal allocation of coupons for drugstore items. It suggests issuing coupons to customers with unknown, low, or middle incomes, provided their daily pre-campaign expenditures did not exceed 900 monetary units and/or their family size is no more than 2 members. Customers in the higher-middle income group should receive drugstore coupons if their average daily in-store spending before the campaign did not exceed 100 monetary units. Finally, customers in the high-income group should only receive drugstore coupons if they previously purchased other food products at the store.

While optimal policy trees are a powerful tool for optimal customer segmentation and intervention assignment, one issue worth noting is their potential instability across different datasets, which arises because changes in the sample composition can lead to different subsets of customers being identified as optimal segments. This phenomenon is particularly evident when there are multiple ways of generating subsamples that result in similar average overall effectiveness of the intervention. Even though this instability may seem impractical, the method represents a significant advancement in merging machine learning and impact evaluation to optimally target customers with marketing

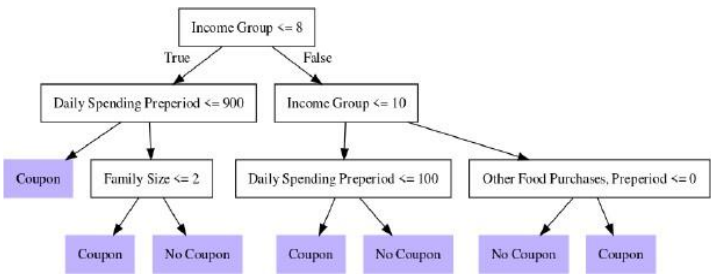

Figure 3: Policy tree for optimal coupon assignment

interventions. This is also because the easily interpretable tree structure makes the decision process regarding intervention assignment accessible in a non-technical manner.

5 Conclusion

Causal machine learning (CML) merges impact evaluation with machine learning, serving three primary purposes in marketing contexts, as outlined in this paper. Firstly, CML facilitates impact evaluation of marketing interventions by controlling for important covariates in a data-driven manner. Secondly, CML enables the detection of effect heterogeneity, allowing for the identification of customer segments that respond differently to marketing interventions. Thirdly, CML permits optimal customer segmentation and targeting by the intervention based on covariates, maximizing the intervention's overall impact. Empirical examples from various studies, including evaluations of coupon campaigns, conversion ads on social media platforms, and the impact of discounted railway tickets on travel behavior, illustrate the practical application of CML in marketing research. Even though the utilization of CML is still limited outside of large tech companies, there is a growing trend towards its adoption, indicating a shift in corporate decision-making practices towards integrating causal methods into business analytics, as highlighted in the study by Hünermund, Kaminski, and Schmitt (2021).

References

AML, H. (2019): “Predicting Coupon Redemption,” url: https://www.kaggle.com/vasudeva009/ predicting-coupon-redemption.

AREVALILLO, J. (2021): “Ensemble learning from model based trees with application to differential price sensitivity assessment,” Information Sciences, 557, 16–33.

ATHEY, S., AND G. IMBENS (2016): “Recursive partitioning for heterogeneous causal effects,” Proceedings of the National Academy of Sciences, 113, 7353–7360.

ATHEY, S., J. TIBSHIRANI, AND S. WAGER (2019): “Generalized random forests,” The Annals of Statistics, 47, 1148–1178.

ATHEY, S., AND S. WAGER (2021): “Policy learning with observational data,” Econometrica, 89(1), 133–161.

BAWA, K., AND R. W. SHOEMAKER (1989): “Analyzing incremental sales from a direct mail coupon promotion,” Journal of marketing, 53(3), 66–78.

BELLONI, A., V. CHERNOZHUKOV, I. FERNANDEZ-VAL, AND C. HANSEN (2017): “Program Evaluation and Causal Inference With High-Dimensional Data,” Econometrica, 85(1), 233–298.

BREIMAN, L. (2001): “Random forests,” Machine Learning, 45, 5–32.

CHERNOZHUKOV, V., D. CHETVERIKOV, M. DEMIRER, E. DUFLO, C. HANSEN, W. NEWEY, AND J. ROBINS (2018): “Double/debiased machine learning for treatment and structural parameters,” The Econometrics Journal, 21, C1–C68.

CHERNOZHUKOV, V., M. DEMIRER, E. DUFLO, AND I. FERNANDEZ-VAL (2018): “Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India,” National Bureau of Economic Research Working Paper 24678.

DONNELLY, R., F. RUIZ, D. BLEI, AND S. ATHEY (2021): “Counterfactual inference for consumer choice across many product categories,” Quantitative Marketing and Economics, pp. 1–39.

FONG, N. M., Z. FANG, AND X. LUO (2015): “Geo-conquesting: Competitive locational targeting of mobile promotions,” Journal of Marketing Research, 52(5), 726–735.

GORDINI, N., AND V. VEGLIO (2017): “Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry,” Industrial Marketing Management, 62, 100–107.

GORDON, B., R. MOAKLER, AND F. ZETTELMEYER (2022): “Close Enough? A Large-ScaleExploration of Non-Experimental Approaches to Advertising Measurement,” arXiv preprint arXiv:220107055.

HO, T. K. (1995): “Random decision forests,” in Proceedings of 3rd international conference on document analysis and recognition, vol. 1, pp. 278–282. IEEE.

HU, M., C. DANG, AND P. CHINTAGUNTA (2019): “Search and learning at a daily deals website,” Marketing Science, 38(4), 609–642.

HUBER, M., J. MEIER, AND H. WALLIMANN (2022): “Business analytics meets artificial intelligence: Assessing the demand effects of discounts on Swiss train tickets,” Transportation Research Part B: Methodological, 163, 22–39.

HUNERMUND, P., J. KAMINSKI, AND C. SCHMITT (2021): “Causal Machine Learning and Business Decision Making,” Available at SSRN 3867326.

IMBENS, G. W. (2004): “Nonparametric estimation of average treatment effects under exogeneity: a review,” The Review of Economics and Statistics, 86, 4–29.

KITAGAWA, T., AND A. TETENOV (2018): “Who should be treated? Empirical welfare maximization methods for treatment choice,” Econometrica, 86, 591–616.

LANGEN, H., AND M. HUBER (2023): “How causal machine learning can leverage marketing strategies: Assessing and improving the performance of a coupon campaign,” Plos one, 18(1), e0278937.

PEARL, J. (2000): Causality: Models, Reasoning, and Inference. Cambridge University Press, Cambridge.

ROBINS, J. M., A. ROTNITZKY, AND L. ZHAO (1994): “Estimation of Regression Coefficients When Some Regressors Are not Always Observed,” Journal of the American Statistical Association, 90, 846–866.

ROSENBAUM, P. R., AND D. B. RUBIN (1983): “Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome,” Journal of the Royal Statistical Society: Series B (Methodological), 45(2), 212–218.

SEMENOVA, V., AND V. CHERNOZHUKOV (2021): “Debiased machine learning of conditional average treatment effects and other causal functions,” The Econometrics Journal, 24, 264–289.

WAGER, S., AND S. ATHEY (2018): “Estimation and Inference of Heterogeneous Treatment Effects using Random Forests,” Journal of the American Statistical Association, 113, 1228–1242.

XIA, F., R. CHATTERJEE, AND J. MAY (2019): “Using conditional restricted Boltzmann machines to model complex consumer shopping patterns,” Marketing Science, 38(4), 711–727.