Causal Inference in Online Advertising Proposal

Can Chen and **Jintong Luo** and **Xiaotian Shao** and **Jingyi Zhou**
Department of Data Science
UC San Diego
La Jolla, CA 92093
cac024@ucsd.edu, jil386@ucsd.edu, x5shao@ucsd.edu, jiz238@ucsd.edu

1 Summary

In online advertising, accurately gauging the uplift effect—the behavioral difference, like increased conversion rates, between users who have and haven't seen an ad—is key to crafting effective strategies. This measurement faces challenges due to biases from complexities in ad delivery systems such as selective exposure and system errors, leading to unreliable exposure data. Moreover, despite the advantages of Randomized Controlled Trials (RCTs) in estimating ads' causal effects, they present practical drawbacks like potential revenue loss from ads withheld from control groups and disruptions from existing ad delivery algorithms and operational needs, complicating data collection and analysis.

Our research aims to refine online advertising effectiveness by developing a methodology that addresses biases in uplift modeling. Leveraging large-scale data and controlled trials, our approach seeks to accurately measure ad impacts and optimize ad strategies. This effort will enhance data reliability for advertisers and contribute new insights to the field of online advertising research.

2 Background

In online advertising, accurately assessing campaign effects through causal inference is crucial for refining strategies and optimizing budgets. Measuring the advertising lift effect, or the increment in user conversion, is a key task in digital brand marketing. Typically, this involves comparing conversions between a control group, which does not receive ads, and a treatment group, which does. However, traditional observational studies lack precision due to non-homogeneous groups (Lewis et al., 2011; Varian, 2016). While randomized controlled trials (RCTs) offer a solution by creating homogeneous groups, real-world application faces challenges like exposure bias, affecting the data quality and measurement accuracy of uplift effects (Gordon et al., 2019; Diemert et al., 2018a; Liu and Chamberlain, 2018).

Uplift modeling aims to estimate the uplift effect, also known as the individual treatment effect (ITE) (Devriendt et al., 2018). Existing two-model approaches typically model two separate potential outcomes for the treatment and control groups, and then estimate the upward effect based on the difference between the outputs of the two models(Hansotia and Rukstales, 2002; Hill, 2011; Cai et al., 2011). In addition, single-model approaches to upward modeling can combine the potential outcome models for the treatment and control groups(Hansotia and Rukstales, 2002; Hill, 2011; Cai et al., 2011).

3 Objectives

Our first objective involves systematically evaluating the effectiveness of different foundational meta-learning methods in the context of advertising uplift models. We plan to implement and compare three approaches: the Single Model Approach (S-learner), where a singular model predicts outcomes with and without treatment; the Two Model Approach (T-learner), which uses separate models for treatment and control groups to isolate the treatment impact; and the Multiple Model Approach (X-learner), designed to estimate treatment effects based on user features and make predictions for new users using propensity-score weighted models. These methods will help quantify the uplift effect more accurately by addressing various modeling challenges.

Our second objective is to refine our approach by leveraging state-of-the-art models known for their effectiveness in similar scenarios. After identifying promising models like CIET, GCF, and specialized boosting algorithms, we plan to select the most applicable SOTA model for our dataset and tasks. We will adapt and fine-tune this model to align with our specific requirements in uplift modeling, adjusting feature extractors and prediction components as needed. Finally, we will evaluate the performance of the fine-tuned model against established benchmarks to validate its effectiveness and pinpoint areas for further refinement. Through these efforts, we aim to optimize the precision and adaptability of our uplift models, ensuring they meet the practical demands of modern advertising campaigns.

4 Datasets

The **Criteo Uplift Prediction Dataset**, **ALI-BRANDLIFT Dataset**, and **Hillstrom's Email Marketing Dataset** are pivotal resources for causal inference and uplift modeling. The Criteo dataset, released by Criteo Labs, includes over 25 million data points from a randomized trial aimed at measuring the incremental effect of advertisement exposure on user conversion rates. The ALI-BRANDLIFT Dataset from Alibaba provides insights into how marketing interventions influence brand perception and consumer behavior on e-commerce platforms. Lastly, Hillstrom's Email Marketing Dataset stems from an email marketing experiment and includes data on user demographics, past purchases, and responses to marketing emails.

5 Methods

There are multiple uplift modeling methods proposed, and in this project, we want to explore their effects on the online advertising scenario. We will compare their effects and finally choose the best method, fine-tuning it to make it suitable for various tasks.

5.1 Meta Learning Methods

Meta-learners are common ways to solve uplift modeling by using supervised algorithms to predict treated and control effect. Here we plan to use several popular meta-learners:

- **S-learner**: A single model is trained on the entire population with a treatment indicator. The uplift is calculated as the difference in predicted outcomes with and without treatment. (Lo, 2002).
- **T-learner**: Separate models are trained for the treatment and control groups. The uplift is determined by subtracting the control group predictions from the treatment group predictions (Betlei et al., 2018).
- **X-learner**: This approach involves training additional models to estimate the treatment effect using user features. Separate models for treatment and control can also predict treatment effects for new users, with final scores adjusted by propensity scores (Künzel et al., 2019).

5.2 Direct Modeling Methods

Besides modeling uplift by groups, we can directly predict uplift effects. Causal Inference Based Single-branch Ensemble Trees (CIET) leverage specialized tree structures to optimize the difference in outcomes between treated and control groups, demonstrating effectiveness in fields like online personal loans (Zheng et al., 2023). Similarly, the integration of Graph Neural Networks with causal knowledge has shown remarkable success in marketing, improving the precision in uplift estimation by handling complex data structures (Wang et al., 2023). The Multihead Causal Distilling Weighting method enhances feature selection and interpretability, significantly boosting model performance in applications that require feature control (Wang et al., 2022). Moreover, Generalized Causal Forest (GCF) extends the causal forest framework to better address heterogeneous treatment effects, proving effective in dynamic settings like online marketplaces (Wan et al., 2022). Lastly, boosting algorithms specifically adapted for uplift modeling have also contributed to the field, optimizing the causal impact assessment across various scenarios (Soltys and Jaroszewicz, 2018). These methods collectively push the boundaries of traditional causal inference, offering more robust and scalable solutions for uplift modeling.

5.3 Adaptive Modeling

Based on the state-of-the-art methods identified, we propose an adaptive model capable of being tailored to various tasks and datasets. This involves modifying the feature extractor and prediction head according to the specific requirements of each new task. Subsequently, we will fine-tune the model on these tasks. Our objective is to achieve accuracy levels comparable to those of existing state-of-the-art methods across different tasks. Demonstrating such performance would underscore the robustness and versatility of our model in handling diverse uplift modeling challenges within causal inference.

5.4 Performance Evaluation

Since we can not direct observe the uplift effect in each individual, area under uplift curves (AUUC) an qini coefficients are used to measure the model performance (Diemert et al., 2018b; Radcliffe, 2007).

References

Artem Betlei, Eustache Diemert, and Massih-Reza Amini. 2018. Uplift prediction with dependent feature representation in imbalanced treatment and control conditions. In *Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13–16, 2018, Proceedings, Part V 25*, pages 47–57. Springer.

Tianxi Cai, Tian Lu, Peggy H. Wong, et al. 2011. Analysis of randomized comparative clinical trial data for personalized treatment selections. *Biostatistics*, 12(2):270–282.

Floris Devriendt, Darie Moldovan, and Wouter Verbeke. 2018. 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*, 6(1):13.

Eustache Diemert, Artem Betlei, Christophe Renaudin, and Massih-Reza Amini. 2018a. A large scale benchmark for uplift modeling. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*.

Eustache Diemert, Artem Betlei, Christophe Renaudin, and Massih-Reza Amini. 2018b. A large scale benchmark for uplift modeling. In *KDD*.

Brett R Gordon, Florian Zettelmeyer, Neha Bhargava, and Dan Chapsky. 2019. A comparison of approaches to advertising measurement: Evidence from big field experiments at facebook. *Marketing Science*, 38(2):193–225.

Behram Hansotia and Brad Rukstales. 2002. Incremental value modeling. *Journal of Interactive Marketing*, 16(3):35–46.

Jennifer L. Hill. 2011. Bayesian nonparametric modeling for causal inference. *Journal of Computational & Graphical Statistics*, 20(1):217–240.

Sören R Künzel, Jasjeet S Sekhon, Peter J Bickel, and Bin Yu. 2019. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the national academy of sciences*, 116(10):4156–4165.

Randall A. Lewis, Justin M. Rao, and David H. Reiley. 2011. Here, there, and everywhere: Correlated online behaviors can lead to overestimates of the effects of advertising. In *Proceedings of the 20th International Conference on World Wide Web*.

CH Liu and Benjamin Paul Chamberlain. 2018. Online controlled experiments for personalised e-commerce strategies: Design, challenges, and pitfalls. arXiv preprint arXiv:1803.06258.

Victor SY Lo. 2002. The true lift model: a novel data mining approach to response modeling in database marketing. *ACM SIGKDD Explorations Newsletter*, 4(2):78–86.

Nicholas Radcliffe. 2007. Using control groups to target on predicted lift: Building and assessing uplift model. *Direct Marketing Analytics Journal*, pages 14–21.

Michal Soltys and S. Jaroszewicz. 2018. Boosting algorithms for uplift modeling. *ArXiv*.

Hal R. Varian. 2016. Causal inference in economics and marketing. *Proceedings of the National Academy of Sciences of the United States of America*, 113(27):7310–7315.

Shu Wan, Chen Zheng, Zhonggen Sun, Mengfan Xu, Xiaoqing Yang, Hong-Tu Zhu, and Jiecheng Guo. 2022. Generalized causal forest for heterogeneous treatment effect estimation in online marketplace. *ArXiv*.

Haowen Wang, Xinyan Ye, Zhiyi Zhang, and Yikang Wang. 2022. Multihead causal distilling weighting is all you need for uplift modeling. *2022 International Conference on High Performance Big Data and Intelligent Systems (HDIS)*.

Haowen Wang, Xinyan Ye, Yangze Zhou, Zhiyi Zhang, L. Zhang, and Jing Jiang. 2023. Uplift modeling based on graph neural network combined with causal knowledge. *ArXiv*.

Fanglan Zheng, Menghan Wang, Kun Li, Jiang Tian, and Xiaojia Xiang. 2023. Causal inference based single-branch ensemble trees for uplift modeling. *ArXiv*.