

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/349762341>

Uplift Modeling: from Causal Inference to Personalization

Conference Paper · April 2021

DOI: 10.1145/3442442.3453701

CITATIONS

0

READS

535

3 authors, including:



[Irene Teinemaa](#)

University of Tartu

23 PUBLICATIONS 426 CITATIONS

[SEE PROFILE](#)



[Dmitri Goldenberg](#)

Tel Aviv University

18 PUBLICATIONS 64 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Alarm-based predictive process monitoring [View project](#)

Uplift Modeling: from Causal Inference to Personalization

Irene Teinemaa
irene.teinemaa@booking.com
Booking.com, Amsterdam

Javier Albert
javier.albert@booking.com
Booking.com, Tel Aviv

Dmitri Goldenberg
dima.goldenberg@booking.com
Booking.com, Tel Aviv

ABSTRACT

Uplift modeling is a collection of machine learning techniques for estimating causal effects of a treatment at the individual or subgroup levels. Over the last years, causality and uplift modeling have become key trends in personalization at online e-commerce platforms, enabling the selection of the best treatment for each user in order to maximize the target business metric. Uplift modeling can be particularly useful for personalized promotional campaigns, where the potential benefit caused by a promotion needs to be weighed against the potential costs. In this tutorial we will cover basic concepts of causality and introduce the audience to state-of-the-art techniques in uplift modeling. We will discuss the advantages and the limitations of different approaches and dive into the unique setup of constrained uplift modeling. Finally, we will present real-life applications and discuss challenges in implementing these models in production.

CCS CONCEPTS

• **Computing methodologies** → *Causal reasoning and diagnostics*; • **Information systems** → *Personalization*; • **Theory of computation** → *Mathematical optimization*.

KEYWORDS

Causality, Uplift Modeling, Causal Inference, Personalization, Heterogeneous Treatment Effects

ACM Reference Format:

Irene Teinemaa, Javier Albert, and Dmitri Goldenberg. 2021. Uplift Modeling: from Causal Inference to Personalization. In *Proceedings of The Web Conference 2021 (WebConf '21)*. ACM, New York, NY, USA, 3 pages.

1 INTRODUCTION

Uplift models [3] are commonly used to estimate the expected causal effect of a treatment on the outcome of individuals, such as subscribing to a service, completing a purchase or responding to a medical treatment. This can be achieved by estimating the *Conditional Average Treatment Effect (CATE)*, defined as the expected increment in a user's outcome probability *caused* by the treatment, given the individual's characteristics. It is particularly useful in the e-commerce setup, where we are interested in estimating the response to website changes for each of the users, in order to personalize their experience.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WebConf '21, April 2021, Ljubljana, Slovenia

© 2021 Association for Computing Machinery.

In the real world we can observe the outcome of a user only under the treatment she actually received and we will not know what would have happened, had she received a different treatment. As a result, we cannot directly calculate the treatment effect for any individual user. This problem, known as the fundamental problem of causal inference [8], poses challenges in CATE estimation, since contrary to classical supervised machine learning, there is no labelled data available.

Various CATE estimation techniques in the literature try to overcome this problem in different ways, falling under two broad categories: *meta-learners* and *tailored methods* [23]. Meta-learning techniques, such as the two-models approach [7], the X-learner [13] and the transformed outcome method [1] allow using classical machine learning techniques for estimating the CATE. Tailored methods, such as uplift trees [20] and various neural network based approaches [9, 16, 22], modify well-known machine learning algorithms to be suitable for CATE estimation.

Over the recent years uplift modeling has become popular in web and e-commerce applications, such as in Facebook, Amazon, Criteo, Uber and Booking.com [4, 5, 17, 25]. In such applications, product improvements and promotions are typically tested via large-scale A/B testing, which allows estimating the overall treatment effect [10]. The data collected during the A/B tests can subsequently be used for uplift modeling, in order to distinguish between the *voluntary buyers*, who would purchase even without receiving the treatment, and the *persuadables*, who would only purchase as a response to the treatment [14].

A particularly interesting uplift modeling problem arises when the cost of the available treatments varies across individuals. In this setup, it is often the goal to maximize the overall incremental outcome while complying with a constraint on the total cost. The CATE estimations on both the outcome and the cost can be combined to solve a constrained optimization problem [5]. Some recent works address the problem of treatment allocation by estimating the marginal uplift [24], incorporating net value optimization within meta-learners [25], exploring the causal effect with bandits [15], or modeling the optimization task as a min-flow problem [17].

Implementing uplift models in production environments gives rise to several operational challenges. For instance, models that are trained offline might be biased towards historical data and require dynamic calibration [26] according to long-term changes and seasonality trends, which are particularly common in the travel industry [18] and in promotional campaigns in general [12]. This makes bridging between the underlying theory and practical implementations peculiarly relevant to applied data science research.

2 TUTORIAL OUTLINE

The tutorial introduces key concepts on causality as well as recent advances in uplift modeling. The outline of the tutorial is as follows: First, we introduce basic concepts in causal inference under the

Potential Outcomes framework. We continue with an overview of state-of-the-art uplift modeling techniques for evaluating and estimating conditional average treatment effects. Next, we discuss constrained uplift problems, a recent addition to the uplift modeling literature aimed at enabling cost-aware personalized treatment assignment. Lastly, we present real-life applications of uplift modeling and discuss challenges in implementing these models in production. The total duration of the tutorial is three hours.

2.1 Detailed Schedule

- (1) Introduction to Causality (40 min)
 - Potential Outcomes Framework
 - Average Treatment Effects
 - Identifiability of Causal Effects
 - Conditional Average Treatment Effects (CATE)
- (2) Uplift Modeling (60 min)
 - Techniques for CATE Estimation
 - Meta-Learners
 - Tailored Methods
 - Evaluating Uplift Models
- (3) Uplift Modeling with Cost Optimization (40 min)
 - Types of Costs and Return-On-Investment
 - Constrained Optimization
- (4) Applications and Implementation Challenges (40 min)
 - Application Examples
 - Model Robustness
 - Exploration
 - Adaptiveness

3 RELEVANCE TO THE COMMUNITY

In the past, causal inference has been associated mostly with clinical trials and social science applications. However, over the recent years, web applications have become increasingly more interactive, raising the need to estimate causal effects of online interventions to advance from correlation-based models to causal models. Learning the effects of different interventions and adjusting the personalization strategy accordingly has become a key trend in the e-commerce industry.

3.1 Intended Audience

The tutorial is targeted to industry practitioners and empirical researchers who are interested in getting causal insights from observational or interventional data and/or in building personalized e-commerce applications. As prerequisites, basic knowledge of probability, statistics and machine learning is expected. No prior knowledge of causal inference is required.

3.2 Related Tutorials

The tutorial is built upon materials from Booking.com's internal causal inference and machine learning trainings. Parts of the tutorial cover novel materials on recent research papers, talks, other tutorials and practical implementations. Throughout the tutorial we aim to convey our experience from a wide usage of uplift modeling in personalization applications at Booking.com. In our recent tutorial at WSDM 2021 [6], we present key trends in personalization, including an extended chapter on causality and uplift modeling.

The proposed tutorial is intended to deep dive into uplift modeling topic, allowing extensive theoretical review and wide coverage of practical applications.

In prior tutorials by Kiciman and Sharma at WSDM 2019 [11] and CoDS COMAD 2020 [21], the authors cover the importance of causality and pitfalls in conventional machine learning techniques that rely on correlation analysis. They present the concept of counterfactual reasoning and dive into methods for causal inference on large-scale online data. These methods set the ground for uplift modeling applications, providing a controlled training dataset for machine learning solutions.

Another tutorial by Cui et al. [2] at KDD 2020, introduces the strong relationship between causality and machine learning and describe various techniques for treatment effect estimation with advanced learning methods. Our proposed tutorial will cover similar techniques and expand on their applications in uplift modeling and treatment allocation optimization.

Besides these past tutorials, we rely on recent surveys by Devriendt et. al [3], Olaya et. al [19] and Zhang et. al [23] that review state-of-the-art uplift modeling techniques and evaluation methods and compare their relative performance.

3.3 Tutorial Format

The suggested tutorial is a lecture-style tutorial, covering theoretical and practical aspects of uplift modeling. Tutorial participants will be provided with supplemental materials, including tutorial slides, references to relevant literature and open-source libraries for uplift modeling. The tutorial is build upon Booking.com internal training which is conducted online. The international team of presenters will give a special attention to interactive discussions during the session, taking into account the challenges of remote presence.

4 PRESENTERS' BIOGRAPHY

Irene Teinemaa is a Machine Learning Scientist at Booking.com in Amsterdam. She obtained her PhD on the topic of predictive and prescriptive monitoring of business processes from University of Tartu. Her current work is focused on uplift modeling and online experimentation for personalized marketing campaigns. She is one of the trainers of an internal causal inference training and has prior experience with academic teaching. She has published her work in top venues such as TKDD, TIST, Data Mining and Knowledge Discovery, ECML/PKDD and ASONAM.

Javier Albert is a Machine Learning Scientist at Booking.com in Tel Aviv, where he uses uplift modeling to enable the financial viability of large-scale promotional campaigns. His work was recently published at ACM RecSys 2020. Albert has a M.Sc. in Electrical Engineering from Tel Aviv University and two B.Sc. degrees from Technion - Israel Institute of Technology.

Dmitri (Dima) Goldenberg is a Machine Learning Science Team Lead at Booking.com, Tel Aviv, where he leads modeling efforts in cross-product offering and promotions personalization, utilizing online learning and uplift modeling techniques. He obtained his Masters in Industrial Engineering (with honors) in Tel Aviv University, focusing on research of influence maximization in social networks. He was included in the Rector's list for excellence in studies and undergraduate courses teaching. He has presented

and published in numerous top journals, physical and online conferences including ESWA, PyData, RecSys, IEEE TCSS, NetSci and ASONAM.

REFERENCES

- [1] Susan Athey and Guido W Imbens. 2015. Machine learning methods for estimating heterogeneous causal effects. *stat* 1050, 5 (2015), 1–26.
- [2] Peng Cui, Zheyang Shen, Sheng Li, Liuyi Yao, Yaliang Li, Zhixuan Chu, and Jing Gao. 2020. Causal Inference Meets Machine Learning. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 3527–3528.
- [3] 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 (2018), 13–41.
- [4] Eustache Diemert, Artem Betlei, Christophe Renaudin, and Massih-Reza Amini. 2018. A large scale benchmark for uplift modeling. In *KDD*.
- [5] Dmitri Goldenberg, Javier Albert, Lucas Bernardi, and Pablo Estevez. 2020. Free Lunch! Retrospective Uplift Modeling for Dynamic Promotions Recommendation within ROI Constraints.. In *Proceedings of the 14th ACM Conference on Recommender Systems*.
- [6] Dmitri Goldenberg, Kostia Kofman, Javier Albert, Sarai Mizrahi, Adam Horowitz, and Irene Teinmaa. 2021. Personalization in Practice: Methods and Applications. In *Proceedings of the 14th International Conference on Web Search and Data Mining*.
- [7] Behram J Hansotia and Bradley Rukstales. 2002. Direct marketing for multichannel retailers: Issues, challenges and solutions. *Journal of Database Marketing & Customer Strategy Management* 9, 3 (2002), 259–266.
- [8] Paul W Holland. 1986. Statistics and causal inference. *Journal of the American statistical Association* 81, 396 (1986), 945–960.
- [9] Fredrik Johansson, Uri Shalit, and David Sontag. 2016. Learning representations for counterfactual inference. In *International conference on machine learning*. 3020–3029.
- [10] Raphael Lopez Kaufman, Jegar Pitchforth, and Lukas Vermeer. 2017. Democratizing online controlled experiments at Booking.com. *arXiv preprint arXiv:1710.08217* (2017).
- [11] Emre Kiciman and Amit Sharma. 2019. Causal Inference and Counterfactual Reasoning (3hr Tutorial). In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. 828–829.
- [12] Hyeon Min Kim and Thomas Kramer. 2006. “Pay 80%” versus “get 20% off”: The effect of novel discount presentation on consumers’ deal perceptions. *Marketing Letters* 17, 4 (2006), 311–321.
- [13] Sören R Künnel, 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 (2019), 4156–4165.
- [14] Yi-Ting Lai, Ke Wang, Daymond Ling, Hua Shi, and Jason Zhang. 2006. Direct marketing when there are voluntary buyers. In *Sixth International Conference on Data Mining (ICDM’06)*. IEEE, 922–927.
- [15] Ying-Chun Lin, Chi-Hsuan Huang, Chu-Cheng Hsieh, Yu-Chen Shu, and Kun-Ta Chuang. 2017. Monetary discount strategies for real-time promotion campaign. In *Proceedings of the 26th International Conference on World Wide Web*. 1123–1132.
- [16] Christos Louizos, Uri Shalit, Joris M Mooij, David Sontag, Richard Zemel, and Max Welling. 2017. Causal effect inference with deep latent-variable models. In *Advances in Neural Information Processing Systems*. 6446–6456.
- [17] Rahul Makhijani, Shreya Chakrabarti, Dale Struble, and Yi Liu. 2019. LORE: a large-scale offer recommendation engine with eligibility and capacity constraints. In *Proceedings of the 13th ACM Conference on Recommender Systems*. 160–168.
- [18] Themis Mavridis, Soraya Hausl, Andrew Mende, and Roberto Pagano. 2020. Beyond algorithms: Ranking at scale at Booking. com. In *Proceedings of the Fourth Workshop on Recommendation in Complex Scenarios. CEUR-WS*.
- [19] Diego Olaya, Kristof Coussement, and Wouter Verbeke. 2020. A survey and benchmarking study of multitreatment uplift modeling. *Data Mining and Knowledge Discovery* 34, 2 (2020), 273–308.
- [20] Piotr Rzepakowski and Szymon Jaroszewicz. 2012. Decision trees for uplift modeling with single and multiple treatments. *Knowledge and Information Systems* 32, 2 (2012), 303–327.
- [21] Amit Sharma and Emre Kiciman. 2020. Causal Inference and Counterfactual Reasoning. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*. 369–370.
- [22] Jinsung Yoon, James Jordon, and Mihaela van der Schaar. 2018. GANITE: Estimation of individualized treatment effects using generative adversarial nets. In *International Conference on Learning Representations*.
- [23] Weijia Zhang, Jiuyong Li, and Lin Liu. 2020. A unified survey on treatment effect heterogeneity modeling and uplift modeling. *arXiv preprint arXiv:2007.12769* (2020).
- [24] Kui Zhao, Junhao Hua, Ling Yan, Qi Zhang, Huan Xu, and Cheng Yang. 2019. A Unified Framework for Marketing Budget Allocation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1820–1830.
- [25] Zhenyu Zhao and Totte Harinen. 2019. Uplift modeling for multiple treatments with cost optimization. In *2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, 422–431.
- [26] Yunhong Zhou, Deeparnab Chakrabarty, and Rajan Lukose. 2008. Budget constrained bidding in keyword auctions and online knapsack problems. In *International Workshop on Internet and Network Economics*. Springer, 566–576.