Proceedings of the 2019 Pre-ICIS SIGDSA Symposium

Special Interest Group on Decision Support and Analytics (SIGDSA)

Winter 12-2019

Multi-Valued Treatments Uplift Modeling for Continuous Outcomes

Robin M. Gubela

Stefan Lessmann

Multi-Valued Treatments Uplift Modeling for Continuous Outcomes

Prototype Demonstration

Robin M. Gubela
Humboldt-Universität zu Berlin
robin.gubela@hu-berlin.de

Stefan Lessmann
Humboldt-Universität zu Berlin
stefan.lessmann@hu-berlin.de

Abstract

Uplift modeling is an application of causal machine learning and offers an assortment of analytical tools to identify likely responders to a particular treatment such as a medical prescription, a political maneuver, or an advertising stimulus. Although several targeted campaigns co-occur (e.g., through different marketing channels), recent literature has primarily examined the effectiveness of a single treatment. To address the practically more pertinent question of which treatment among several options to choose, we develop a prototype that identifies the most effective treatment for each unit of observation and further generalizes to both binary and continuous outcomes to support classification and regression problems. Using real-world data from e-mail merchandising and e-couponing campaigns, we verify our prototype's financial advantage compared to previous efforts toward the single treatment case.

Keywords

Uplift model, multi-valued treatments, continuous outcome.

Introduction

According to current statistics, the spend on digital advertisements has exceeded 100 billion US-dollars in 2018 (Ha 2019), which confirms the cross-sectoral interest of practitioners in further advancing advertising campaigns. From a process-related viewpoint, a crucial aspect of capitalizing on investments into an advertising campaign relates to the meticulous planning of whom to allocate a campaign incentive (e.g., Kane et al. 2014). Predictive analytics is typically used to recognize units of observation (hereafter referred to as *units*) that are likely performing the incentivized task (e.g., sign a commercial contract). Although the unit-specific information might be estimable, it lacks a causal association between the treatment and the expected outcome, which restricts the quality of the targeting decision. In contrast to units with a low probability to adopt the requested behavior, the corresponding model selects units with a high probability as targets for a particular activity. However, issuing an incentive to an already likely converter unnecessarily wastes financial resources and might even be perceived as irritating from the recipient's perspective. While response models predict this likelihood by considering the units that have obtained the treatment (i.e., the treatment sample), attrition models regard the non-treated units (i.e., the control sample) (Radcliffe 2007).

In contrast to these models, we argue in favor of adopting an uplift model for campaign-based decision-making as it targets a likely responder with an estimated positive effect due to the promotion of a campaign incentive (e.g., Devriendt et al. 2018). To this end, uplift models require data from treatment and control groups. The treatment incidences need to be random or *conditionally independent* from further features to alleviate biased model predictions (Imbens and Rubin 2015). Uplift models predict individual-level treatment effects (ITE), also known as conditional average treatment effects (CATE) (e.g., Gutierrez and Gérardy 2017; Knaus et al. 2018), as they depend on a unit's definite characteristics (e.g., a customer's browsing behavior in an online marketing context or a citizen's past political participation activities regarding an election campaign). For each unit, an uplift model estimates both sign and strength of a treatment's persuasive impact on its desired behavior. Practitioners target units according to their relative ITE in decreasing order. Grounding on scalable machine learning algorithms, uplift models typically benefit from high scalability which is particularly relevant in settings with substantial data amounts such as in customer relationship management (e.g., Gubela et al. 2019).

Most uplift research focuses on marketing applications where the effectiveness of a single treatment (e.g., a specific e-mail promotion or product catalog) to forecast a binary outcome (e.g., a service subscription or product purchase) is examined (e.g., Devriendt et al. 2018). Therefore, researchers develop novel ITE estimators to increase the likelihood of gaining a dichotomous return (e.g., click-through-rates or purchase occurrences). The Qini coefficient (Radcliffe 2007) measures a model's performance and translates the estimated treatment effects into a key performance indicator (KPI) for domain-specific audiences. The results facilitate an analyst's decision-making on whom to allot the focused treatment.

The multi-valued treatments setting comprises applications where several incentives co-exist (e.g., Rzepakowski and Jaroszewicz 2012). Analysts make model-dependent targeting decisions and allocate the particular treatment with the highest ITE relative to the other treatments to an individual unit, that is, the treatment with the highest likelihood to alter future behavior. In a broad sense, suppose several types of marketing communications (e.g., offline catalogs vs. online newsletters). More narrowly, we refer to the multi-valued treatments setting if only parameters of otherwise identical incentives alter. For example, these could be campaigns with different contents or varying values of coupon discounts. In contrast to the single treatment setting that studies the effectiveness of one dedicated treatment, analysts assign different treatments to different units in the multi-valued treatments setting.

For the single treatment problem, most estimators consist of decision tree-based structures (e.g., Athey et al. 2019) which have demonstrated high performance in both real-world and simulation experiments (e.g., Guelman et al. 2015; Knaus et al. 2018). Recent literature further lists causal algorithms such as support vector machines (e.g., Zaniewicz and Jaroszewicz 2013), neural networks (e.g., Shalit et al. 2017) and meta-methods that are not limited to a single algorithm (Gubela et al. 2017; Künzel et al. 2019). However, such endeavors focus predominantly on the single treatment case, whereas the case of multi-valued treatments has not yet been sufficiently studied. Only very few methods for multi-valued treatments uplift modeling exist and include decision trees (Rzepakowski and Jaroszewicz 2012; Zhao et al. 2017), k-nearest neighbors (Su et al. 2012) and cluster analysis (Lo and Pachamanova 2015).

Aside from the binary outcome setting, only a single study in the field of uplift modeling has explored continuous outcomes in detail (Gubela et al. 2017). Predicting such an outcome empowers analysts to measure the magnitude of a unit's activity, such as the value of a shopping basket at different time states or financial margins from customer transactions. In many real-world cases, continuous outcome settings are more in line with business KPIs. For instance, marketing analysts are interested in increasing returns on marketing investments (ROMI), and continuous outcomes such as revenue gains and marketing expenditures might serve as suitable approximations. At the same time, the choice which units to treat is more focused as a model considers more granular meta-information of each unit. To our best knowledge, current literature has not yet sufficiently studied continuous outcomes for multi-valued treatments.

We develop a prototype for the multi-valued treatments setting. The prototype calculates treatment-specific ITE for each unit and identifies the most effective treatment based on unit-wise comparisons of treatment effects. It estimates binary and continuous outcomes by conducting classification and regression tasks, respectively. To showcase the prototype's practical relevance, we conduct an empirical analysis with two real-world marketing data sets and compare the prototype's effectiveness against recent efforts in terms of the single treatment case, including recognized methods such as causal forests (Athey et al. 2019). The employed data sets refer to different marketing applications, that is, e-mail merchandising promotions for men and women (Hillstrom 2008), and online shopping campaigns with varying discount values of otherwise identical digital coupons. The e-mail promotion data is publicly available, which facilitates the reproducibility of our prototype's results.

To emphasize the prototype's practical utility, we translate our findings from its predictive performance into business KPIs, such as the spend amount. As a result, we stress that considering the single treatment setting is not optimal if several campaigns co-exist as it ignores a unit's treatment affiliation. Our analyses indicate that allocating the most influential treatment to a particular customer is more appealing from a financial view. Although merchants appreciate the advantage of their campaign-related activities compared to no activity, they lack the insight of each treatment's effectiveness. Such knowledge is crucial, primarily if treatments significantly differ in their persuasive impact and financial value.

References

Athey, S., Tibshirani, J., and Wager, S. 2019. "Generalized Random Forests," *The Annals of Statistics* (47:2), pp. 1148-1178.

Devriendt, F., Moldovan, D., and Verbeke, W. 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), pp. 13-41.

Gubela, R. M., Bequé, A., Gebert, F., and Lessmann, S. 2019. "Conversion Uplift in E-Commerce: A Systematic Benchmark of Modeling Strategies," *International Journal of Information Technology & Decision Making* (18:3), pp. 747-791.

Gubela, R. M., Lessmann, S., Haupt, J., Baumann, A., Radmer, T., and Gebert, F. 2017. "Revenue Uplift Modeling," *Proceedings of the 38th International Conference on Information Systems (ICIS 2017)*, Seoul, Korea: AIS.

Guelman, L., Guillén, M., and Pérez-Marín, A. M. 2015. "A Decision Support Framework to Implement Optimal Personalized Marketing Interventions," *Decision Support Systems* (72), pp. 24-32.

Gutierrez, P., and Gérardy, J.-Y. 2017. "Causal Inference and Uplift Modelling: A Review of the Literature," *International Conference on Predictive Applications and APIs*, pp. 1-13.

Ha, A. 2019. "US Digital Advertising Exceeded $100B in 2018 (IAB Report)." Retrieved 19th September, 2019, from https://techcrunch.com/2019/05/07/iab-internet-advertising-report

Hillstrom, K. 2008. "The MineThatData E-Mail Analytics and Data Mining Challenge." From https://blog.minethatdata.com/2008/03/minethatdata-e-mail-analytics-and-data.html

Imbens, G. W., and Rubin, D. B. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences*, New York: Cambridge University Press.

Kane, K., Lo, V. S., and Zheng, J. 2014. "Mining for the Truly Responsive Customers and Prospects Using True-Lift Modeling: Comparison of New and Existing Methods," *Journal of Marketing Analytics* (2:4), pp. 218-238.

Knaus, M. C., Lechner, M., and Strittmatter, A. 2018. "Machine Learning Estimation of Heterogeneous Causal Effects: Empirical Monte Carlo Evidence," *IZA Discussion Paper No. 12039*.

Künzel, S. R., Sekhon, J. S., Bickel, P. J., and Yu, B. 2019. "Metalearners for Estimating Heterogeneous Treatment Effects Using Machine Learning," *Proceedings of the National Academy of Sciences* (116:10), pp. 4156-4165.

Lo, V. S., and Pachamanova, A. D. 2015. "From Predictive Uplift Modeling to Prescriptive Uplift Analytics: A Practical Approach to Treatment Optimization While Accounting for Estimation Risk," *Journal of Marketing Analytics* (3:2), pp. 79-95.

Radcliffe, N. J. 2007. "Using Control Groups to Target on Predicted Lift: Building and Assessing Uplift Models," *Direct Marketing Analytics Journal* (1), pp. 14-21.

Rzepakowski, P., and Jaroszewicz, S. 2012. "Decision Trees for Uplift Modeling with Single and Multiple Treatments," *Knowledge and Information Systems* (32:2), pp. 303-327.

Shalit, U., Johansson, F. D., and Sontag, D. 2017. "Estimating Individual Treatment Effect: Generalization Bounds and Algorithms," in *Proceedings of the 34th International Conference on Machine Learning (ICML 2017)*, Sydney, Australia.

Su, X., Kang, J., Fan, J., Levine, R. A., and Yan, X. 2012. "Facilitating Score and Causal Inference Trees for Large Observational Studies," *Journal of Machine Learning Research* (13), pp. 2955-2994.

Zaniewicz, L., and Jaroszewicz, S. 2013. "Support Vector Machines for Uplift Modeling," in *Proceedings of the 13th IEEE International Conference on Data Mining Workshops*, W. Ding, T. Washio, H. Xiong, G. Karypis, B. Thuraisingham, D. Cook and X. Wu (eds.), Dallas, Texas, USA: IEEE Computer Society, pp. 131-138.

Zhao, Y., Fang, X., and Simchi-Levi, D. 2017. "Uplift Modeling with Multiple Treatments and General Response Types," *Proceedings of the 2017 SIAM International Conference on Data Mining*: SIAM, pp. 588-596.