8 Experiment Specification

8.1 Algorithms

1. Single Model Methods: Use the same base learners as Two Model Methods. Simply use the treatment indicator W as a normal feature.
2. Two Model Methods: use (1) Linear Regression, (2) Regression Tree, (3) Random Forest, (4) Neural Network (5) Bayesian Additive Regression Tree https://cran.r-project.org/web/packages/BART/index.html as base learner respectively.
3. X-Learner: use the same base learners as Two Model Methods.
4. Y-Learner:
5. Transformed Outcome Methods: use the same base learners as the single model methods. Transform the outcome variable into: $Y^* = \frac{Y_i \cdot (W_i - e(X_i))}{e(X_i)(1-e(X_i))}$, where p is the propensity score $e(X_i) = P(W = 1|X_i)$. After the transformation, use the base-learner as Single/Two model methods.
6. Causal Tree https://github.com/susanathey/causalTree
7. Causal Random Forest https://github.com/grf-labs/grf
8. CFRNet
9. CEVAE
10. Uplift CCIF, Uplift Random Forest, Uplift KNN https://cran.r-project.org/web/packages/uplift/index.html

8.2 Datasets

1. GOTV
2. IHDP
3. Twins
4. Job
5. Criteo https://ailab.criteo.com/criteo-uplift-prediction-dataset/
6. The MineThatData E-Mail Analytics And Data Mining Challenge https://blog.minethatdata.com/2008/03/minethatdata-e-mail-analytics-and-data.html
7. Synthetic datasets: (1) From Debo, (2) From [2], (3) From [12].

8.3 Metrics

with standard deviation (two decimals), and also report relative metric. For example, for the ϵ-ATE metric, relative version would be like $\frac{|\hat{\tau}-\tau|}{\tau}$ etc.

When reporting the metrics, please don't mix up different metrics. For example, if PEHE cannot be calculated from Twins, just put '-' in that space.

References

[1] Kevin J. Anstrom and Anastasios A. Tsiatis. Utilizing propensity scores to estimate causal treatment effects with censored time-lagged data. *Biometrics*, 57(4):1207–1218, dec 2001.

[2] Susan Athey and Guido Imbens. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences of the United States of America*, 113(27):7353–7360, 2016.

[3] Leo Breiman, Jerome Friedman, Charles J. Stone, and R.A. Olshen. *Classification and Regression Trees*. Belmont: Wadsworth, 1984.

[4] Hugh A. Chipman, Edward I. George, and Robert E. McCulloch. BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4(1):266–298, mar 2010.

[5] Diemert Eustache, Betlei Artem, Christophe Renaudin, and Amini Massih-Reza. A large scale benchmark for uplift modeling. In *Proceedings of the AdKDD and TargetAd Workshop*, 2018.

[6] L. L. Doove, E. Dusseldorp, K. Van Deun, and I. Van Mechelen. A comparison of five recursive partitioning methods to find person subgroups involved in meaningful treatment–subgroup interactions. *Advances in Data Analysis and Classification*, 8(4):403–425, dec 2013.

[7] Leo Guelman, Montserrat Guillén, and Ana M. Pérez-Marín. Random forests for uplift modeling: An insurance customer retention case. In *Lecture Notes in Business Information Processing*, pages 123–133. Springer Berlin Heidelberg, 2012.

[8] Leo Guelman, Montserrat Guillén, and Ana M. Pérez-Marín. Uplift random forests. *Cybernetics and Systems*, 46(3-4):230–248, apr 2015.

[9] Pierre Gutierrez and Jean-Yves Gérardy. Causal inference and uplift modelling: A review of the literature. In *Proceedings of The 3rd International Conference on Predictive Applications and APIs*, 2017.

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

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

[12] Jenny Häggström. Data-driven confounder selection via markov and bayesian networks. *Biometrics*, 74(2):389–398, nov 2017.

[13] Kosuke Imai and Marc Ratkovic. Estimating treatment effect heterogeneity in randomized program evaluation. *Ann. Appl. Stat.*, 7(1):443–470, mar 2013.

[14] Guido Imbens and Donald Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences*. C, 2015.

[15] Maciej Jaskowski and Szymon Jaroszewicz. Uplift modeling for clinical trial data. 2012.

[16] Joseph Kang, Xiaogang Su, Brian Hitsman, Kiang Liu, and Donald Lloyd-Jones. Tree-structured analysis of treatment effects with large observational data. *Journal of Applied Statistics*, 39(3):513–529, mar 2012.

[17] E. L. Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, jun 1958.

[18] Diederik P Kingma and Max Welling. Auto-encoding variational bayes.

[19] Sören R. Künzel, Jasjeet S. Sekhon, Peter J. Bickel, and Bin Yu. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences*, 116(10):4156–4165, feb 2019.

[20] Sören R. Künzel, Bradly C. Stadie, Nikita Vemuri, Varsha Ramakrishnan, Jasjeet S. Sekhon, and Pieter Abbeel. Transfer learning for estimating causal effects using neural networks.

[21] Christos Louizos, Uri Shalit, Joris Mooij, David Sontag, Richard Zemel, and Max Welling. Causal effect inference with deep latent-variable models.

[22] Jared K. Lunceford and Marie Davidian. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statist. Med.*, 23(19):2937–2960, aug 2004.

[23] Mee Young Park and Trevor Hastie. L1-regularization path algorithm for generalizedata linear models. *Journal of the Royal Statistical Society: Series B*, 69(4):659–677, 2007.

[24] Judea Pearl. *Causality*. Cambridge University Press, 2009.

[25] Paul R. Rosenbaum. Model-based direct adjustment. *Journal of the American Statistical Association*, 82(398):387–394, jun 1987.

[26] Paul R. Rosenbaum and Donald B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

[27] Paul R. Rosenbaum and Donald B. Rubin. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, 79(387):516–524, sep 1984.

[28] Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, nov 1987.

[29] Donald B. Rubin. Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine*, 127(8_Part_2):757, oct 1997.

[30] Piotr Rzepakowski and Szymon Jaroszewicz. Decision trees for uplift modeling with single and multiple treatments. *Knowledge and Information Systems*, 32(2):303–327, jul 2011.

[31] Shaun R Seaman and Ian R White. Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research*, 22(3):278–295, jun 2013.

[32] Uri Shalit, Fredrik D. Johansson, and David Sontag. Estimating individual treatment effect: generalization bounds and algorithms.

[33] Michał Sołtys, Szymon Jaroszewicz, and Piotr Rzepakowski. Ensemble methods for uplift modeling. *Data Mining and Knowledge Discovery*, 29(6):1531–1559, sep 2014.

[34] Xiaogang Su, Chih-Ling Tsai, Hansheng Wang, David M. Nkckerson, and Bogong Li. Subgroup analysis via recursive partitioning. *Journal of Machine Learning Research*, 10:141–158, 2009.

[35] Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242, jun 2018.

[36] Ikko Yamane, Florian Yger, Jamal Atif, and Masashi Sugiyama. Uplift modeling from separate labels. 2018.

[37] Łukasz Zaniewicz and Szymon Jaroszewicz. $$l_p$$ l p -support vector machines for uplift modeling. *Knowledge and Information Systems*, 53(1):269–296, mar 2017.

[38] Weijia Zhang, Thuc Duy Le, Lin Liu, and Jiuyong Li. Estimating heterogeneous treatment effect by balancing heterogeneity and fitness. *BMC Bioinformatics*, 19(S19), dec 2018.

[39] Weijia Zhang, Thuc Duy Le, Lin Liu, Zhi-Hua Zhou, and Jiuyong Li. Mining heterogeneous causal effects for personalized cancer treatment. *Bioinformatics*, 33(15):2372–2378, mar 2017.

[40] Yan Zhao, Xiao Fang, and David Simchi-Levi. Uplift modeling with multiple treatments and general response types.