





CRE: An R package for interpretable discovery and inference of heterogeneous treatment effects

Riccardo Cadei ^{1,2*}, Naeem Khoshnevis ^{3*}, Kwonsang Lee ¹, Daniela Maria Garcia ¹, and Falco J. Bargagli Stoffi ¹

¹ Department of Biostatistics, Harvard School of Public Health ² Department of Computer and Communication Science, EPFL ³ Research Computing, Harvard University  Corresponding author * These authors contributed equally.

DOI: [10.21105/joss.05587](https://doi.org/10.21105/joss.05587)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Susan Holmes](#)  

Reviewers:

- [@salleuska](#)
- [@carlyls](#)

Submitted: 17 March 2023
Published: 07 December 2023

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

In health and social sciences, it is critically important to identify interpretable subgroups of the study population where a treatment has notable heterogeneity in the causal effects with respect to the average treatment effect (ATE). Several approaches have already been proposed for heterogeneous treatment effect (HTE) discovery, either estimating first the conditional average treatment effect (CATE) and identifying heterogeneous subgroups in a second stage (Bargagli-Stoffi et al., 2020, 2022; Foster et al., 2011; Hahn et al., 2020), either estimating directly these subgroups in a direct data-driven procedure (Nagpal et al., 2020; Wang & Rudin, 2022). Many of these methodologies are decision tree-based methodologies. Tree-based approaches are based on efficient and easily implementable recursive mathematical programming (e.g., HTE maximization), they can be easily tweaked and adapted to different scenarios depending on the research question of interest, and they guarantee a high degree of interpretability—i.e., the degree to which a human can understand the cause of a decision (Lakkaraju et al., 2016). Despite these appealing features, single-tree heterogeneity discovery is characterized by two main limitations: instability in the identification of the subgroups and reduced exploration of the potential heterogeneity. To accommodate these shortcomings, Bargagli-Stoffi et al. (2023) proposed Causal Rule Ensemble, a new method for interpretable HTE characterization in terms of decision rules, via an extensive exploration of heterogeneity patterns by an ensemble-of-trees approach. CRE is an R package providing a flexible implementation of Causal Rule Ensemble. The package allows for multiple variants of Causal Rule Ensemble algorithm, also including different internal individual average treatment effect (IATE) estimators—i.e., AIPW (Robins et al., 1994), Causal Forest (Athey et al., 2019), Causal BART (Hill, 2011), S-Learner (Hill, 2011), T-Learner (Hansotia & Rukstales, 2002), X-Learner (Künzel et al., 2019).

Statement of Need

Several methodologies for HTE estimation have already been proposed (together with the release of the corresponding packages), but the interpretable discovery of the subgroups and the key factors driving the HTE is still an open challenge. To the best of our knowledge, `causalTree`, based on Causal Honest Tree (Athey & Imbens, 2016), is the unique R package proposing a methodology for interpretable HTE discovery and estimation via decision rules. Still, despite its appealing features, it is also characterized by the limitations of single tree-based methods. Firstly, single-tree-based subgroup identification is sensitive to variations in the training sample (high model variance)—e.g., if the data are slightly altered, a completely different set of discovered subgroups might be found (Breiman, 1996; Hastie et al., 2009; Kuhn et al., 2013). Secondly, it may fail to explore a vast number of potential subgroups

(limited subgroup exploration)—e.g., the subgroups discovered are just the ones that can be represented by a single tree (Kuhn et al., 2013; Spanbauer & Sparapani, 2021). To illustrate, consider a scenario in which two distinct factors independently contribute to the heterogeneity in treatment effects. In such cases, a single tree algorithm may detect only one of these factors, failing to identify the second. In instances where both factors are identified, they are detected sub-optimally as an interaction between the two variables rather than as distinct drivers of the treatment heterogeneity. To account for these shortcomings, we propose CRE (Khoshnevis et al., 2023), an R package providing a flexible implementation of the Causal Rule Ensemble algorithm. CRE provides (i) an interpretable representation of the HTE in observational studies, (ii) via an extensive exploration of complex heterogeneity patterns using decision rules, while (iii) guaranteeing high stability in the discovery.

Algorithm

Causal Rule Ensemble relies on the Treatment Effect linear decomposition assumption, which characterizes the Conditional Average Treatment Effect (CATE) as the sum of $M + 1$ distinct contributions:

$$\tau(x) = \mathbb{E}[\tau_i | X_i = x] = \bar{\tau} + \sum_{m=1}^M \alpha_m \cdot r_m(x)$$

where $\bar{\tau}$ is the ATE, τ_i is the ITE, and for each m in $\{1, \dots, M\}$, r_m is an interpretable decision rule characterizing a specific subset of the covariate space, and α_m is the corresponding Additive Average Treatment Effect (AATE). The CRE procedure is divided into two steps, discovery and inference, and each observation is used for only one of the two steps (honest splitting). The splitting is at random and the percentage allocated to each step is controlled. During the discovery step, CRE retrieves the M decision rules characterizing the heterogeneity in the treatment effect. A set of candidate decision rules is extracted by an ensemble of trees modelling some IATE estimates (Dorie et al., 2020; Polley et al., 2019; Tibshirani et al., 2023) (*fit-the-fit* approach), and among these, only a simple and robust subset of rules is selected for the linear decomposition by the Stability Selection algorithm via LASSO (Friedman et al., 2021; Hofner et al., 2015; Meinshausen & Bühlmann, 2010). During the inference step, CRE estimates the ATE and AATEs, by the normal equations to model some IATE estimates and confidence intervals are provided by bootstrapping. A brief schematic summary of the described procedure (Bargagli-Stoffi et al., 2023) is reported below.

Algorithm Causal Rule Ensemble (CRE)

Inputs: covariates matrix \mathbf{X} , (binary) treatment vector \mathbf{z} , and observed response vector \mathbf{y} .

Outputs: (i) a set of interpretable decision rules $\hat{\mathcal{R}} = \{\hat{r}_m\}_{m=1}^M$,
(ii) ATE $\hat{\tau}$ and AATEs $\hat{\alpha}$ estimates and confidence intervals,

Procedure:

$(\mathbf{X}^{dis}, \mathbf{z}^{dis}, \mathbf{y}^{dis}), (\mathbf{X}^{inf}, \mathbf{z}^{inf}, \mathbf{y}^{inf}) \leftarrow \text{HonestSplitting}(\mathbf{X}, \mathbf{z}, \mathbf{y})$

i. Discovery

$\hat{\tau}^{dis} \leftarrow \text{EstimateIATE}(\mathbf{X}^{dis}, \mathbf{z}^{dis}, \mathbf{y}^{dis})$

▷ e.g. AIPW, CF, CausalBART, S/T/X-Learner

$\hat{\mathcal{R}}' \leftarrow \text{GenerateRules}(\mathbf{X}^{dis}, \hat{\tau}^{dis})$

▷ i.e., tree-ensemble method

$\hat{\mathcal{R}} \leftarrow \text{RulesSelection}(\hat{\mathcal{R}}', \mathbf{X}^{dis}, \hat{\tau}^{dis})$

▷ Stability Selection + LASSO

ii. Inference

$\hat{\tau}^{inf} \leftarrow \text{EstimateIATE}(\mathbf{X}^{inf}, \mathbf{z}^{inf}, \mathbf{y}^{inf})$

▷ e.g. AIPW, CF, CausalBART, S/T/X-Learner

$\hat{\alpha} \leftarrow \text{EstimateAATE}(\hat{\mathcal{R}}, \mathbf{X}^{inf}, \hat{\tau}^{inf})$

▷ Bootstrapping + Linear smoothing

Usage

CRE is available both on CRAN (Khoshnevis et al., 2023) and GitHub, and can be installed and loaded into the R session using:

```
install.packages("CRE")
library("CRE")
```

`generate_cre_dataset()` is a flexible synthetic dataset generator, which can be used for simulations before applying CRE to real-world observational data sets. It generates an outcome array y (binary or continuous), a treatment array z (binary), a covariate matrix (binary or continuous) and the true (unobserved) individual treatment effect ite (useful for performance evaluation). The input parameters specify the dataset characteristics, including the number of individuals (n), the number of covariates (p), the correlation within the covariates (ρ), the number of decision rules (n_rules) decomposing the CATE, the treatment effect magnitude (`effect_size`), the confounding mechanism (`confounding`), and whether the covariates and outcomes are binary or continuous (`binary_covariates`, `binary_outcome`). For a full description of the data generating process and its variants, see Section 4 in Bargagli-Stoffi et al. (2023).

```
set.seed(2023)
dataset <- generate_cre_dataset(n = 5000,
                                rho = 0,
                                n_rules = 4,
                                p = 10,
                                effect_size = 5,
                                binary_covariates = TRUE,
                                binary_outcome = FALSE,
                                confounding = "no")

y <- dataset$y
z <- dataset$z
X <- dataset$X
ite <- dataset$ite
```

We propose here three examples of how to run the Causal Rule Ensemble algorithm by the CRE package.

Example 1. Running Causal Rule Ensemble with default parameters. For a detailed description of the default method and hyper parameters refer to <https://nsaph-software.github.io/CRE/articles/CRE.html>.

```
results <- cre(y, z, X)
```

Example 2. Running Causal Rule Ensemble with customized IATE estimator. If `ite` argument is provided, the IATE estimation both in the discovery and inference step is skipped. This argument is useful either to consider new IATE estimators which are not internally implemented in this package, either to compute this estimation *una tantum* (saving the results) and speed up the execution time during the hyper-parameter tuning.

```
# personalized IATE estimation (S-Learner with Linear Regression)
model <- lm(y ~., data = data.frame(y = y, X = X, z = z))
iate_pred <- predict(model, newdata = data.frame(X = X, z = z))

results <- cre(y, z, X, ite = iate_pred)
```

Example 3. Running Causal Rule Ensemble with customized parameters (no need to explicit all the arguments). The method parameters (`method_params`) entail the IATE estimator method, the outcome and propensity score learners, if needed), and the the ratio of data to use for each step. The hyper parameters (`hyper_params`) entail all the other parameters through which Causal Rule Ensemble algorithm can be fine tuned. For a detailed description of all the method and hyper parameters refer to <https://nsaph-software.github.io/CRE/articles/CRE.html>.

```
method_params <- list(ratio_dis = 0.5,
                      ite_method = "aipw",
                      learner_ps = "SL.xgboost",
                      learner_y = "SL.xgboost")
```

```
hyper_params <- list(intervention_vars = NULL,
                     offset = NULL,
                     ntrees = 20,
                     node_size = 20,
                     max_rules = 100,
                     max_depth = 3,
                     t_decay = 0.025,
                     t_ext = 0.025,
                     t_corr = 1,
                     t_pvalue = 0.05,
                     stability_selection = "vanilla",
                     cutoff = 0.9,
                     pfer = 0.1,
                     B = 50,
                     subsample = 0.05)
```

```
results <- cre(y, z, X, method_params, hyper_params)
```

The results are collected in a S3 object containing: the number of decision rules extracted at each step (M), the list of the rules finally selected (rules), a data.frame with the CATE decomposition estimates with corresponding uncertainty quantification (CATE) and the list of selected parameters (method_params and hyper_params).

summarize() and print() display a summary of these results. predict() estimates the Individual Treatment Effect on a new Covariate matrix by the linear decomposition just learnt. plot() visualizes the CATE decomposition estimates in a range bar plot.

```
print(results)
y_pred <- predict(results, X)
plot(results)
```

Figure 1 reports the visualization of the results for Example 3, which perfectly discover the correct CATE decomposition.

Causal Rule Ensemble: Conditional Average Treatment Effect Linear Decomposition

ATE = -0.213 [$-1.063, 0.637$]

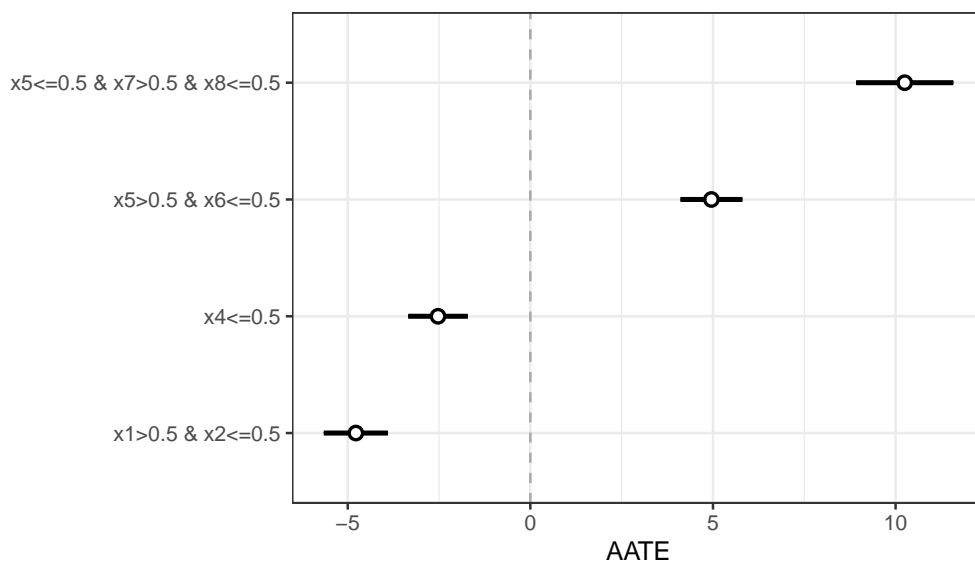


Figure 1: Visualization of Causal Rule Ensemble HTE linear decomposition for Example 3. For each decision rule discovered, the corresponding AATE estimate with 95% confidence interval is reported in a range bar plot. The decision rules are ordered from the most vulnerable (high AATE) to the least, and the ATE is reported on top of the plot.

The observed average execution time of the method varying the number of individuals and observed covariates on R 4.2.1 running on macOS 12.6 on a MacBook Pro 16GB Apple 8-cores M1 processor is reported in Figure 2.

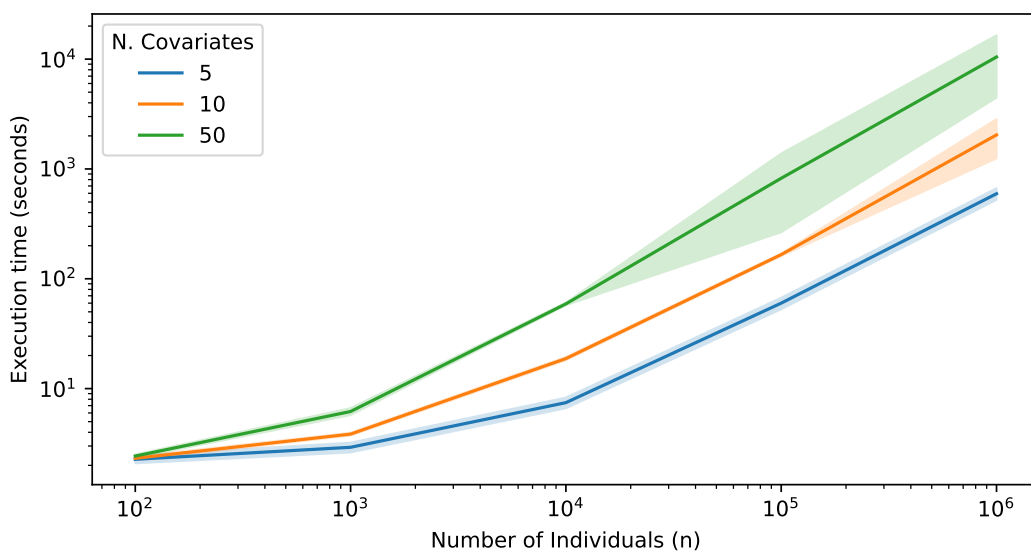


Figure 2: Log-Log line plot reporting the average execution time of `cre()` and standard deviation over 10 seeds per experiment, varying the number of individuals and observed covariates.

Online documentation for the package can be found at <https://nsaph-software.github.io/CRE/>.

Acknowledgements

This work was partially funded by the following grants: NIH: R01ES026217, R01MD012769, R01ES028033, 1R01ES030616, 1R01AG066793, 1R01MD016054-01A1, R01AG066793-02S1 and R01ES028033-03S1; Sloan Foundation: G-2020-13946.

References

- Athey, S., & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27), 7353–7360. <https://doi.org/10.1073/pnas.1510489113>
- Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, 47(2), 1148–1178. <https://doi.org/10.1214/18-AOS1709>
- Bargagli-Stoffi, F. J., Cadei, R., Lee, K., & Dominici, F. (2023). Causal rule ensemble: Interpretable discovery and inference of heterogeneous treatment effects. *arXiv*. <https://doi.org/10.48550/arXiv.2009.09036>
- Bargagli-Stoffi, F. J., De Witte, K., & Gnecco, G. (2022). Heterogeneous causal effects with imperfect compliance: A bayesian machine learning approach. *The Annals of Applied Statistics*, 16(3), 1986–2009. <https://doi.org/10.1214/21-aos1579>
- Bargagli-Stoffi, F. J., Tortù, C., & Forastiere, L. (2020). Heterogeneous treatment and spillover effects under clustered network interference. *arXiv*. <https://doi.org/10.48550/arXiv.2008.00707>
- Breiman, L. (1996). Heuristics of instability and stabilization in model selection. *The Annals of Statistics*, 24(6), 2350–2383. <https://doi.org/10.1214/aos/1032181158>
- Dorie, V., Hill, J., & Dorie, M. V. (2020). Package “bartcause.” URL: <https://Cran.R-Project.Org/Web/Packages/bartCause/bartCause>. Pdf. <https://CRAN.R-project.org/package=bartCause>
- Foster, J. C., Taylor, J. M. G., & Ruberg, S. J. (2011). Subgroup identification from randomized clinical trial data. *Statistics in Medicine*, 30(24), 2867–2880. <https://doi.org/10.1002/sim.4322>
- Friedman, J., Hastie, T., Tibshirani, R., Narasimhan, B., Tay, K., Simon, N., & Qian, J. (2021). Package “glmnet.” CRAN R Repository. <https://CRAN.R-project.org/package=glmnet>
- Hahn, P. R., Murray, J. S., & Carvalho, C. M. (2020). Bayesian Regression Tree Models for Causal Inference: Regularization, Confounding, and Heterogeneous Effects (with Discussion). *Bayesian Analysis*, 15(3), 965–2020. <https://doi.org/10.1214/19-BA1195>
- Hansotia, B., & Rukstales, B. (2002). Incremental value modeling. *Journal of Interactive Marketing*, 16(3), 35–46. <https://doi.org/10.1002/dir.10035>
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (Vol. 2). Springer. <https://doi.org/10.1007/978-0-387-21606-5>
- Hill, J. L. (2011). Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1), 217–240. <https://doi.org/10.1198/jcgs.2010.08162>
- Hofner, B., Hothorn, T., & Hofner, M. B. (2015). Package “stabs”. <https://CRAN.R-project.org/package=stabs>

- Khoshnevis, N., Garcia, D. M., Cadei, R., Lee, K., & Bargagli Stoffi, F. J. (2023). *CRE: Interpretable subgroups identification through ensemble learning of causal rules*. <https://CRAN.R-project.org/package=CRE>
- Kuhn, M., Johnson, K., & others. (2013). *Applied predictive modeling* (Vol. 26). Springer. <https://doi.org/10.1007/978-1-4614-6849-3>
- Künzel, S. R., Sekhon, J. S., Bickel, P. J., & Yu, B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences*, 116(10), 4156–4165. <https://doi.org/10.1073/pnas.1804597116>
- Lakkaraju, H., Bach, S. H., & Leskovec, J. (2016). Interpretable decision sets: A joint framework for description and prediction. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1675–1684. <https://doi.org/10.1145/2939672.2939874>
- Meinshausen, N., & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 72(4), 417–473. <https://doi.org/10.1111/j.1467-9868.2010.00740.x>
- Nagpal, C., Wei, D., Vinzamuri, B., Shekhar, M., Berger, S. E., Das, S., & Varshney, K. R. (2020). Interpretable subgroup discovery in treatment effect estimation with application to opioid prescribing guidelines. *Proceedings of the ACM Conference on Health, Inference, and Learning*, 19–29. <https://doi.org/10.1145/3368555.3384456>
- Polley, E., LeDell, E., Kennedy, C., Lendle, S., & Laan, M. van der. (2019). Package “SuperLearner.” CRAN. <https://CRAN.R-project.org/package=SuperLearner>
- Robins, J. M., Rotnitzky, A., & Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427), 846–866. <https://doi.org/10.2307/2290910>
- Spanbauer, C., & Sparapani, R. (2021). Nonparametric machine learning for precision medicine with longitudinal clinical trials and bayesian additive regression trees with mixed models. *Statistics in Medicine*, 40(11), 2665–2691. <https://doi.org/10.1002/sim.8924>
- Tibshirani, J., Athey, S., Friedberg, R., Hadad, V., Hirshberg, D., Miner, L., Sverdrup, E., Wager, S., Wright, M., & Tibshirani, M. J. (2023). Package “grf”. <https://CRAN.R-project.org/package=grf>
- Wang, T., & Rudin, C. (2022). Causal rule sets for identifying subgroups with enhanced treatment effects. *INFORMS Journal on Computing*. <https://doi.org/10.1287/ijoc.2021.1143>