A Meta-Learner Framework to Estimate Individualized Treatment Effects for Survival Outcomes

Na Bo   Yue Wei   Lang Zeng   Chaeryon Kang   Ying Ding

1. Guidance

Please down load all of the code and data files named as "Code_and_Data_HTEsurv" are available in github. All simulation files are stored at "Simulations" folder. All real data analysis files are stored at "RealDataAnalysis" folder. We will describe each folder below.

2. Simulation

2.1 File description

All simulation files are stored at "Simulations" folder. Within this folder:

- "intermediate_result_100simulations" folder contains the intermediate results used for reproducing tables and figures simulations. You will need to load these intermediate results if you run "reproduce_simulation_tables_plots.Rmd". Please refer to section "Reproduce simulation tables and figure" below for running this R markdown file.

- "prepared_simulation_data_example" folder contains one training data (train1.RData) and test (test.RData) data prepared by us for simulations when data is generated from Weibull models under balanced design and scenario 1. "Tlearner_result.csv" and "Xlearner_result.csv" are simulation results of D-T and D-X by submitting Python script "DNNSurv_Tlearner_1run.py" and "DNNSurv_Xlearner_1run.py" in "Simulations/source_files" folder, using our prepared training and test data. Please refer to the R markdown file "Example_code_simulation_1run.Rmd" described in section "Example of one run of simulation when data is generated from Weibull models under balanced design and scenario 1" below for running D-T and D-X in Python.

- "source_files" folder contains all functions needed to run R-T, R-X, B-T, B-X, D-T, D-X, weibull true model, logistic true model and data generation process. Here are more detailed descriptions:

  1. baft_TXlearners.R includes functions for running T- and X-learners for bayesian accelerated failure models (BAFT).

  2. baft_bart_np_SurvivalProb.R is sourced to predict survival probabilities in baft_TXlearners.R.

  3. rsf_TXlearners.R includes functions for running T- and X-learners for random survival forests (RSF).

  4. DNNSurv_Tlearner_1run.py includes the function for running T-learner for Cox-based deep neural network models for survival outcomes (DNNSurv) for one simulation.

  5. DNNSurv_Tlearner_100runs.py includes the function for running T-learner for Cox-based deep neural network models for survival outcomes (DNNSurv) for 100 simulations.

6. DNNSurv_Xlearner_1run.py includes the function for running X-learner for Cox-based deep neural network models for survival outcomes (DNNSurv) for one simulation.

7. DNNSurv_Xlearner_100runs.py includes the function for running X-learner for Cox-based deep neural network models for survival outcomes (DNNSurv) for 100 simulations.

8. fun_util.py is sourced in DNNSurv_Tlearner.py and DNNSurv_Xlearner.py for calculating the baseline hazards.

9. simulation_design_survival_StandardLogistic.R includes the function for data generation when true potential survival times are generated from log-logistic models.

10. simulation_design_survival_Weibull.R includes the function for data generation when true potential survival times are generated from Weibull models.

11. StandardLogistic_TrueModel.R includes the function for running the log-logistic model which is considered as the true model when true potential survival times are generated from the log-logistic model.

12. weibull_true.R includes the function for running the Weibull model which is considered as the true model when true potential survival times are generated from the Weibull model.

2.2 Example of one run of simulation when data is generated from Weibull models under balanced design and scenario 1

File "Example_code_simulation_1run.Rmd" gives an example of running one simulation when data is generated from Weibull models under balanced design and scenario 1. You can open and knit or run it by following the instructions in it. It also contains instructions of running D-T and D-X in Python. Other simulation settings can be run in the same way.

To knit this Rmd file, please run code install.packages("bookdown") to install "bookdown" package. Before you knit this file, please follow the instruction in line 24 to add a code chunk of setting up the directory.

"Example_code_simulation_1run.html" is the output after knitting this file. Some browsers may block the file. Please try a different browser if it cannot be opened.

2.3 Reproduce simulation tables and figures

File "reproduce_simulation_tables_plots" contains codes for running all simulations in the manuscript. It also contains instructions of running D-T and D-X in Python. Please follow the instructions in it and use server to run 100 times simulations. This file also gives codes of reproducing simulation tables and figures in the manuscript. You can open and knit or run it by following the instructions in it.

To knit this Rmd file, please run code install.packages("bookdown") to install "bookdown" package. Before you knit this file, please follow the instruction in line 23 to add a code chunk of setting up the directory.

"reproduce_simulation_tables_plots.html" is the output after knitting this file. Some browsers may block the file. Please try a different browser if it cannot be opened.

3. Real data analysis

3.1 File description

All real data analysis files are stored at "RealDataAnalysis" folder. Within this folder:

- "intermediate_result" folder contains the intermediate results used for reproducing tables and figures. You will need to load these intermediate results if you run "RealDataResults_manuscript.Rmd". Please refer to section "Reproduce tables and figures in real data analysis" below for running this R markdown file.

- "data" folder contains partially shared AREDS and AREDS2 data. We do not share full data, but we share outcomes, treatment information and three SNPs for generating tables.

- "source_files" folder contains all functions needed to run R-T, R-X, B-T, B-X, D-T and D-X. Here are more detailed descriptions:

1. aft_bart_np_survival.R includes functions for running T- and X-learners for bayesian accelerated failure models (BAFT).

2. aft_bart_np_SurvivalProb.R is sourced to predict survival probabilities in aft_bart_np_survival.R.

3. DNNSurv_HTE_survival.R includes the functions for running T- and X-learner for Cox-based deep neural network models for survival outcomes (DNNSurv). This requires to source TensorFlow from Python.

4. fun_util.R is sourced in DNNSurv_HTE_survival.R for calculating the baseline hazards.

5. rsf_HTE_survival.R includes the functions for running T- and X-learners for random survival forests (RSF).

3.2 Reproduce tables and figures in real data analysis

"RealDataResults_manuscript.Rmd" contains codes for reproducing plots and tables for real data analysis. Please follow the instructions in it. It also contains instructions about calling Python in R. Since we do not share full data, we provide example codes of running R-T, R-X, B-T, B-X, D-T, D-X and CSF in real data. These codes are commented out.

To knit this Rmd file, please run code install.packages("bookdown") to install "bookdown" package. Before you knit this file, please follow the instruction in line 20 to add a code chunk of setting up the directory.

"RealDataResults_manuscript.html" is the html file after knitting the "RealDataResults_manuscript.Rmd". Some browsers may block the file. Please try a different browser if it cannot be opened.