A Matching-based machine learning approach to estimating optimal dynamic treatment regimes with time-to-event outcomes

Lingqi Huang

November 22, 2024

Wang X, Lee H, Haaland B, Kerrigan K, Puri S, Akerley W, Shen J. A matching-based machine learning approach to estimating optimal dynamic treatment regimes with time-to-event outcomes. Stat Methods Med Res. 2024 May;33(5):794-806. doi: 10.1177/09622802241236954. Epub 2024 Mar 19. PMID: 38502008.

History

- Observational data(EHR) has become increasingly important in evidence-based research on **dynamic treatment regimes**, which tailor treatment over time to patients based on their characteristic and evolving clinical history.
- Statisticians and clinicians are interested in identifying an **optimal** dynamic treatment regime that produce the best expected clinical outcomes for each indivisual and thus maximize the treatment benefit over the population.
- Observational data imposes **various challenges** for using statistical tools to estimate optimal dynamic treatment regimes, but the task becomes more sophisticated when the primary interest is **time-to-event**.

Goal

- Apply a matching-based machine learning method to **identify the optimal dynamic treatment regime** with time-to-event outcomes subjects to **right-censoring** using EHR data.
- Illustrate the method with an application to estimate optimal dynamic treatment regimes for patients with **advanced non-small cell lung cancer** using a real-world, nationwide EHR database from **Flatiron Health**.

Problems

- When identifying optimal DTRs from observational data under a causal inference framework, Q- and A-learning depend heavily on the **correct specification** of parametric models.(Handle confounding problem and using IPW)
- Although O-learning does not assume a explicit outcome model, it could be still **vulnerable** to the misspecification of propensity score model.(Extreme weights)
- IPW-based approaches could face difficulties and lead to **biased results** in certain situation.

Matched Learning on DTR for Survival algorithm(MLSurv)

- **Extends** the M-learning methodology(feasible in individualized treatment rule estimation for a single decision stage with continuous/binary outcomes) to estimate optimal two-stage DTRs for time-to-event endpoints using EHR data.
- Utilize matching methods to address both **confoundings** and **censoring**.
- Using **backward induction** framework and implement a **weighted SVM** to estimate DTRs that optimize the time-to-event outcomes.

Improvement

- In contrast to the established IPW-based dynamic treatment regimes methods, the approach provides better protection against **model misspecification and extreme weights** in context of treatment sequences, effectively addressing a prevalent challenge in the longitudinal analysis of EHR.
- The method demonstrates robust performance across a range of senarios.

Settings

Maximum number of stages: K

Binary treatment decisions: $k = 1, \ldots, K, A_k \in \{-1, 1\}$.

Covariates measured at stage k before A_k : $\mathbf{X}_k$

Variable history: $\bar{A}_k = (A_1, \ldots, A_k)$

Indicator variable of entering stage k for person i:

$$\eta_{ik} = \begin{cases} 1 & \text{entered stage k} \\ 0 & \text{otherwise} \end{cases}$$

Survival time during the interval of A_k: $\tilde{T}_k$ ($\tilde{T}_k = 0$ when $k = 0$)

Settings

Accumulated survival time: $\tilde{T} = \sum_{k=1}^{K} \eta_k \tilde{T}_k$

Counterfatual survival time: $\tilde{T}^{\bar{a}_k} = \sum_{k=1}^{K} \eta_k \tilde{T}_k^{\bar{a}_k}$

Counterfactual treatment sequence: $\bar{a}_K = (a_1, \ldots, a_k)$

Censoring Time: C

Observed overall survival: $Y = \min\{\tilde{T}, C\}$

Event indicator: $\delta = I(\tilde{T} \leq C)$

Observed survival time within stage k: Y_k (Not censored: $Y_k = \tilde{T}$)

Clinical History prior to kth treatment decision:

$$H_k = (\eta_{i1}, \mathbf{X}_{i1}, A_{i1}, \tilde{T}_{i1}, \ldots, \eta_{ik}, \mathbf{X}_{ik})$$

When $\eta_{ik} = 0$, A_{ik} and $\mathbf{X}_{ik}$ are unobservable.

Dynamic Treatment Regime(DTR)

DTR: DTR is a set of treatment decision rules:

$$\mathbf{g} = \{g_1(h_1), \ldots, g_K(h_K)\} \in \mathcal{G}$$

Where $\mathcal{G}$ is the set of all possible treatment regimes, and $g_k(h_k)$ represent the decision rule at a specific stage k, is a map from current history to treatment $\{-1, 1\}$.

Value function: denote τ be the end of study, and set restricted survival time

$$T = \min\{\tilde{T}, \tau\}$$

then the value function that assesses the overall benefit of a DTR($\mathbf{g}$) is $V(\mathbf{g}) = \mathbb{E}(T^{\mathbf{g}})$, and the optimal DTR maximize the value function is:

$$\mathbf{g}^{\text{opt}} = \arg\max_{\mathbf{g} \in \mathcal{G}} \mathbb{E}(T^{\mathbf{g}})$$

Assumptions

- **Stable unit treatment value assumption(STUVA)**: Outcome of an individual is **not influenced** by the treatment applied to other individuals.
- **Consistency**: The counterfactual outcome under the observed treatment is the observed outcome
- **Sequential ignorability**: Treatment assignment at a given stage k is independent with future outcome:

$$\{\sum_{j \geq l}^{K} \tilde{T}^{\bar{a}_j} : l = k, \ldots, K\} \perp A_k | H_k, \eta_1, \ldots, \eta_k$$

- **Censoring at random**: At begining of each stage, the future probability of censoring does not depend on future outcome given current history

$$\{\sum_{j \geq l}^{K} \tilde{T}^{\bar{a}_j} : l = k, \ldots, K\} \perp \delta | H_k, \eta_1, \ldots, \eta_k$$

Matching Based Procedure

Denote

$$\mathcal{M}_{ik} = \{j : \eta_{jk} = 1, A_{jk} = -A_{jk}, d(H_{jk}, H_{ik}) \leq \epsilon_{ik}\}$$

as the matched set for a subject i who entered stage k, receiving the alternative treatment, and with similar history as subject i.

Denote

$$\mathcal{MC}_i = \{j : Y_{jk} > Y_{ik}, \delta_j = 1, d(H_i, H_j) \leq \epsilon_i, A_{ik} = A_{jk}\}$$

as a matched psudo-population without right censoring for i that objects in the matched set share similarities in the covariates and treatment history.(We do this because IPCW is recognized to be **vunlerable** to model misspecification)

Algorithm(Two-Stage Study)

Step 1: For patients who entered the second stage denoted as $S_2 = \{i : \eta_{i2} = 1\}$, for subject i who had **censored** restricted survival time, find a mactched set

$$\mathcal{MC}_{i2} = \{j : Y_{j2} > Y_{i2}, \delta_j = 1, d(H_{i2}, H_{j2}) \leq \epsilon_{i2}\}$$

Then compute

$$T_{i2} = \frac{1}{|\mathcal{MC}_{i2}|} \sum_{j \in \mathcal{MC}_{i2}} T_{j2}$$

If the subject i is not censored, T_{i2} would be the observed survival time.

Algorithm(Two-Stage Study)

Step 2: Based on the matched set

$$\mathcal{M}_{i2} = \{j : \eta_{j2} = 1, A_{j2} = -A_{j2}, d(H_{j2}, H_{i2}) \leq \epsilon_{i2}\}$$

for subject i entered the second stage, specify the matching based value function for stage two, which is the expected restricted survival time given that the second stage treatment assignments follow regimen $g_2(h_2)$:

$$V_2(g_2(h_2)) = \frac{1}{|S_2|} \sum_{i \in S_2} \Bigg[I(g_2(h_{i2}) = a_{i2}) * T_{i2} \tag{1}$$

$$+ I(g_2(h_{i2}) \neq a_{i2}) * \frac{1}{|\mathcal{M}_{i2}|} \sum_{j \in \mathcal{M}_{i2}} T_{j2} \Bigg] \tag{2}$$

Algorithm(Two-Stage Study)

From the classification perspective, the maximization problem can be reformulated as a minimization of misclassification errors.
Denote $g_2(h_2) = \text{sign}(f_2(h_2))$ for the second stage decision function f_2.
To obtain $g_2^{opt}(h_2)$, we need to minimize the objective function:

$$\frac{1}{|S_2|}\sum_{i\in S_2} I\left\{ f_2(h_{i2}) * a_{i2} * \text{sign}\left(T_{i2} - \frac{1}{|\mathcal{M}_{i2}|}\sum_{j\in\mathcal{M}_{i2}} T_{j2} \right) \leq 0 \right\} \tag{3}$$

$$* \left| T_{i2} - \frac{1}{|\mathcal{M}_{i2}|}\sum_{j\in\mathcal{M}_{i2}} T_{j2} \right| \tag{4}$$

It has the form of loss function for a weighted classification problem where the weights are constructed using matched pairs.
We use **SVM with hinge loss function** to derive the estimated decision rule for stage two $\hat{g}_2^{\text{opt}}(H_2)$

Algorithm(Two-Stage Study)

Step 3: With the estimated optimal second stage treatment $\hat{a}_2^{\text{opt}}$, construct the pseudo-outcome for stage 1 as:

$$T^{a_1,\hat{a}_2^{\text{opt}}} = T^* = T_1 + \eta_2 * T_2^{\hat{a}_2^{\text{opt}}}$$

Step 4: For subject i who did not enter the second stage and had censored restricted survival time, find the matched set :

$$\mathcal{MC}_{i1} = \{j : Y_{j1} > Y_{i1}, \delta_j = 1, d(H_{i1}, H_{j1}) \leq \epsilon_i, A_{i1} = A_{j1}\}$$

Then $T_i^* = \frac{1}{|\mathcal{MC}_{i1}|} \sum_{j \in \mathcal{MC}_{i1}} T_j^*$.

Note that patients in the matched set can enter the second stage. For subject i did not enter the second stage and had the event observed, $T^* = Y_{i1} = T_{i1}$.

Algorithm(Two-Stage Study)

Step 5: Define the value function for the first stage using the constructed pseudo-outcome T^*:

$$V_1(g_1(h1), g_2^{\text{opt}}(h_2)) = \frac{1}{N}\sum_i \Bigg\{ \left[\left(\frac{1}{|\mathcal{M}_{i1}|} \sum_{j \in \mathcal{M}_{i1}} T_j^* \right) - T_i^* \right] \quad (5)$$

$$* I(g_1(h_{i1}) \neq a_{i1}) + T_i^* \Bigg\} \quad (6)$$

Similar to **step 2**, the weighted classification objective function can be used to estimate the optimal first stage decision rule, $\hat{g}_1^{\text{opt}}(h_1)$

Summary

- The underlying idea of the proposed method is to utilize matching based procedures to obtain the counterfactual restricted survival time, and then apply weighted classification machine learning methods to estimate the optimal DTRs, where the weights depend on counterfactual outcomes.
- MLSurv exhibits robust performance by requiring neither prior knowledge regarding the shape of the decision boundaries, nor the modeling assumptions on treatment propensity and censoring. Notably, it effectively handles cases where extreme weights are likely to occur when applying IPW-based approaches.