PERSONALIZED TREATMENT SELECTION MODEL FOR SURVIVAL OUTCOMES

Matteo Pedone [1], Raffaele Argiento[2] and Francesco C. Stingo[1]

[1] Department Statistics, Computer Science and Applications, University of Florence, (e-mail: matteo.pedone@unifi.it, francescoclaudio.stingo@unifi.it)

[2] Department of Economics, University of Bergamo, (e-mail: raffaele.argiento@unibg.it)

ABSTRACT: Precision medicine, a patient-centric approach to disease treatment, has attracted considerable interest in recent years. Building on a prior method focused on short-term outcomes, we introduce a model that clusters patients based on similar predictive characteristics and treatment responses, enabling optimal therapeutic strategy selection via predictive inference for new patients, incorporating long-term survival outcomes.

KEYWORDS: Nonparametric Bayes, personalized medicine, predictive probability, product partition models, time-to-failure endpoints.

1 Introduction

The field of oncology has shifted towards personalized treatments that take into account the heterogeneity of cancer pathogenesis. This is driven by the recognition that cancer molecular mechanisms are complex and multifactorial, involving multiple biomarkers and pathways. To address this complexity, the focus has shifted towards developing therapies that are based on multiple biomarkers.

Statistical methods for personalized treatment selection need to consider the uniqueness of each tumor and individual patient characteristics. The common assumption of statistical exchangeability among patients should be relaxed, and patients should only be treated as exchangeable to the extent to which their tumors are molecularly similar. By leveraging individual patient

Matteo Pedone gratefully acknowledges the support of the European Union - Next GenerationEU, UNIFI Young Independent Researchers Call - BayesMeCOS Grant no. B008-P00634.

characteristics, these personalized treatment strategies have the potential to improve treatment efficacy and patient outcomes.

? proposed a hybrid two-step approach for accurate treatment selection that integrates a Bayesian predictive model with prognostic and predictive biomarkers. Patients are grouped based on molecular similarity using heuristic clustering algorithms, and a Bayesian model predicts treatment response probabilities for each competing treatment. This approach improves upon existing methods by relaxing the assumption of full exchangeability among patients and utilizing complementary sources of information.

? developed a fully Bayesian method that builds upon **?**'s approach and improves upon it by jointly performing clustering and prediction using a nonparametric approach. By combining the two tasks into a single model, the need for multi-step procedures is eliminated. The nonparametric approach provides a sound framework for both clustering and prediction, accounting for the uncertainty in all modeling steps. Ultimately, the individualized treatment rule fully accounts for patients' heterogeneity, as confirmed by improved prediction performances compared to **?**'s method (**?**).

? used a categorical outcome to evaluate treatment effectiveness after a post-therapy follow-up period. However, this approach may be limited since it only considers short-term outcomes. To address this limitation, we suggest using time-to-event analysis to base treatment selection on long-term outcomes such as disease progression, relapse, or death. This approach can better capture treatment effectiveness.

2 Survival model

We examine a group of n patients from past clinical studies who were treated with T different treatments. The patients' predictive and prognostic biomarkers were measured, along with the survival times. The treatments are indexed by $a = 1, \ldots, T$, and the total number of treated patients is denoted by $n = \sum_{a=1}^{T} n^a$, where n^a is the number of patients receiving therapy a. The observed survival times of patients are represented as vectors $\boldsymbol{t}^a,\ a = 1, \ldots, T$. However, due to limited study duration, not all patients experience the event of interest, resulting in "right-censored" data. To account for this, binary indicator vectors $\boldsymbol{d}^a,\ a = 1, \ldots, T$ are introduced to identify patients whose event was observed during follow-up and those who were censored. In the case of patients who received treatment a and experienced an observed event or censored time ($d_i^a = 1$), their time to event is represented by t_i^a. On the other hand, if $d_i^a = 0$, meaning that the patient did not experience an event or was not cen-

sored during the study period, then t_i^a represents the length of their follow-up.

We use an accelerated failure time (AFT) model to analyse right-censored survival data, that takes into account the $p-$ and $q-$ dimensional vector of prognostic and predictive features. Prognostic and predictive markers are denoted as $\boldsymbol{z}_i^a$ and $\boldsymbol{x}_i^a$, respectively, measured on the $i-$th patient who received treatment a. It is assumed that patients with similar genetic profiles are likely to have similar responses to a given treatment. We assume that $\Pi_{n^a}^a = S_1^a, \ldots, S_{C_{n^a}^a}^a$ is a given treatment-specific partition of the indices $1, \ldots, n^a$, where $C_{n^a}^a$ is the number of clusters among patients treated with therapy a, and $n_j^a = |S_j^a|$ is the number of patients in cluster j, for $j = 1, \ldots, C_{n^a}^a$.

$$\log(t_i^a) = \mu_j^{a\star} + \boldsymbol{z}_i \boldsymbol{\beta} + \sigma\varepsilon_i,$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is the vector of regression coefficients and $\boldsymbol{\varepsilon}$ is the error vector. Assuming a minimum value Gumbel distribution for the error terms $\varepsilon_1, \ldots, \varepsilon_n \sim \text{Gumbel}(0,1)$, gives rise to the Gumbel AFT model. Moreover, we assume $\beta_k \overset{\text{iid}}{\sim} N(0, \lambda_k^2\tau^2), \lambda_k, \tau \overset{\text{iid}}{\sim} HC(0, 1/p)$ (namely, a horseshoe prior), and $\sigma \sim U(a_\sigma, b_\sigma)$. Moreover, $\mu_j^{a\star}$s are cluster-specific parameters. We assume a product partition model with covariates (PPMx, **?**) for the joint distribution of the clustering and the cluster-specific parameters $(\Pi_{n^a}^a,\ \mu_j^{a\star})$, to induce independence across clusters and conditional independence within clusters. The joint law of $(\Pi_n, \mu_j^{a\star})$ is assigned hierarchically as:

$$\begin{aligned} \mu_j^{a\star} \mid \boldsymbol{\zeta}, \Pi_{n^a}^a &\overset{\text{ind}}{\sim} N(\theta, \boldsymbol{\Sigma}) \\ \Pi_{n^a}^a &\sim PPMx(\boldsymbol{x}). \end{aligned}$$

All the details pertaining to the specification, construction, and posterior inference of PPMx can be found in **?**.

3 Example

We conduct a simulation study to evaluate the properties of the proposed method on finite samples. We consider 200 patients assigned to two competing treatments and use piecewise constant exponential distributions to generate survival outcomes, that is we do not simulate from our model. Our simulation design is inspired by the work of **?**. To evaluate the performance of the clustering procedure, we generate synthetic covariates (5 predictive and 5 prognostic biomarkers) with a known clustering structure (a two-component mixture of normals). The validation set comprised 100 patients. In Table **??** we report the

Adjusted Rand Index (ARI), the Mis-assigned Optimal Treatment (MOT), and Proportion of Treatment Utility (PTU) averaged over 10 replications (standard deviation in parenthesis).

ARI	MOT	PTU
1.00	7.50	0.93
(0.00)	(5.92)	(0.06)

Table 1: Results simulation study.

In terms of clustering, the model demonstrates a remarkable level of effectiveness. The quantity of non-optimal treatment assignments is approximately 8 per 100 patients, with a considerable standard deviation. However, the high PTU value suggests that the misassigned patients may belong to a subgroup with similar treatment benefits across therapies.

4 Conclusion

Overall, our study's preliminary results are promising and suggest that the proposed method has potential for accurately assigning treatments using long-term outcomes.