The Journal of Open Source Software

survxai: an R package for model agnostic explanations of survival models

Aleksandra Grudziaz[1, 2], **Alicja Gosiewska**[2], **and Przemyslaw Biecek**[1, 2]

1 Faculty of Mathematics, Informatics, and Mechanics, University of Warsaw **2** Faculty of Mathematics and Information Science, Warsaw University of Technology

DOI: 10.21105/joss.00961

Software

- Review
- Repository
- Archive

Submitted: 05 September 2018
Published: 17 September 2018

Introduction

Predictive models are widely used in ~~the~~ supervised machine learning. Three most common classes of such models are: *regression models*, ~~here~~ the target variable is continuous numeric, *classification models* ~~here~~ the target variable is binary or categorical and *survival models* ~~here~~ the target is some censored variable. Common examples of censored variables are time to death (but some cases lived for x years and are still alive), cessation of service by ~~the~~ customer or failure of ~~a~~ machine component.

Modern predictive models are often complex in structure. ~~Think about~~ neural networks (Eleuteri, Tagliaferri, Milano, De Placido, & De Laurentiis, 2003) or random forest (Ishwaran, Kogalur, Blackstone, & Lauer, 2008). Such models may be described by thousands of coefficients. Often such flexibility leads to high performance, but makes these models opaque, hard to understand. ~~It~~ is acceptable in cases ~~in which~~ only the model accuracy is important, but in cases that involve human decisions it may not be enough. To trust model predictions one needs to see which features are important and how model predictions would change if some feature was changed.

The area of model interpretability or explanability ~~gains quickly~~ attention of machine learning experts. Understanding of complex models leads not only to higher trust in model predictions but also ~~to~~ better models. Better means that ~~they~~ are more robust and ~~maintain~~ high accuracy on validation data. See examples in `DALEX` (Biecek, 2018) or `iml` (Molnar, 2018) R packages.

Existing tools for model agnostic explanations are focused on regression models and classification problems as in both cases model predictions may be summarised by a single number. Survival models require different approach as predictions are in a form of survival curves. Demand for such explainers ~~leads~~ to some model specific solutions, like `iSurvive` ~~introduced by~~ (Dempsey et al., 2017) for continuous time hidden Markov model. Yet, ~~we are lacking of~~ model agnostic tools for survival models.

The `survxai` fills ~~out~~ this gap. This R package is designed to deliver local and global explanations for survival models in a model agnostic fashion. In the package documentation we demonstrate examples for Cox models and for ~~Survival Random Forest~~ models. The `survxai` package consists new implementations and visualisations of explainers designed for survival models. Functions are well documented, package is supplemented with unit tests and illustrations. Regardless of the complexity of the model, methods implemented in the `survxai` package maintain a certain level of interpretability, important in medical applications (Collett, 2015), churn analysis (Lu & Park, 2003) and others.

The Journal of Open Source Software

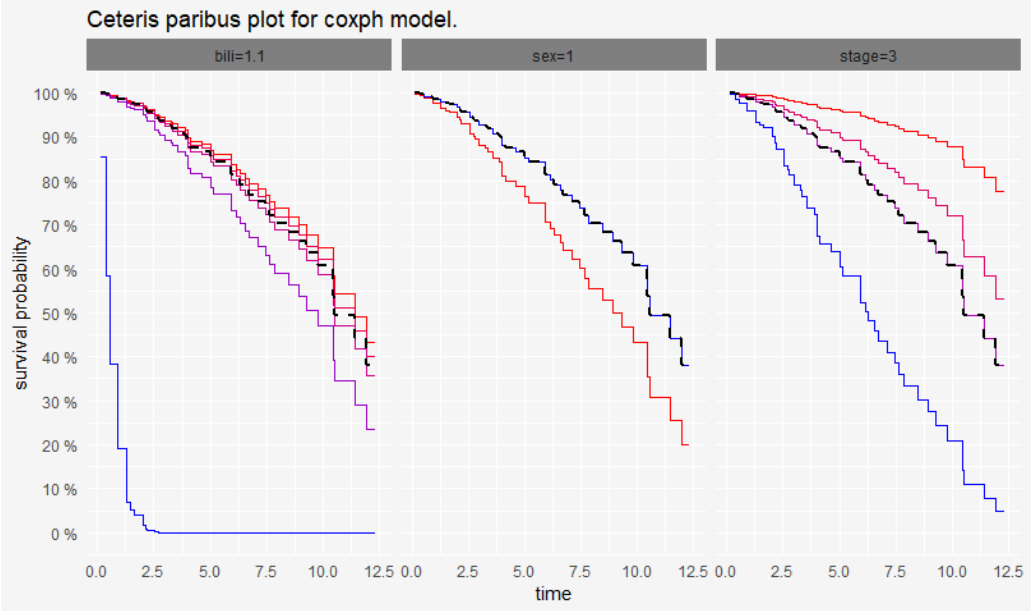

Figure 1: Ceteris Paribus plot for Cox ~~Proportional Hazards~~ model with three variables. ~~Black~~ survival curve corresponds to a observation of interest. ~~Middle~~ panel shows that prediction for sex=0 are worse than for sex=1.

Explanations of survival models

The R package `survxai` is a tool for creating explanations of survival models. ~~It's~~ model agnostic, thus ~~is working with~~ both complex and simple survival models. It also allows ~~to compare~~ two or more models.

Currently, four classes of model explainers are implemented. Two for local explanations (for a single prediction), and two for global explanations (for a whole model).

~~Package~~ `survxai` is available on CRAN and a development version of the package can be found on https://github.com/MI2DataLab/survxai.

Local methods are the explanations of one observation.

- **Ceteris Paribus** plot presents model responses around a single point in the feature space (Biecek, 2018). See ~~an example in Figure 1.~~ Each panel is related to a single variable. ~~Single~~ panel shows how a model prediction (survival curve) would change if only a single variable ~~will be~~ changed. It if ~~usefull~~ for *what-if* reasoning. Each curve in a ~~panle~~ is ~~realted~~ a different value of the selected variable. Ceteris Paribus plot illustrates how ~~may~~ the survival curve change ~~along~~ with the changing values of ~~the~~ variable.

- **break down** plot presents variable contributions in final predictions (Staniak & Biecek, 2018). ~~See an example in Figure 2.~~ The ~~Break Down~~ of prediction for survival model helps to understand which factors drive survival probability for a single observation.

Global methods are model performance and explanations of the conditional model structure.

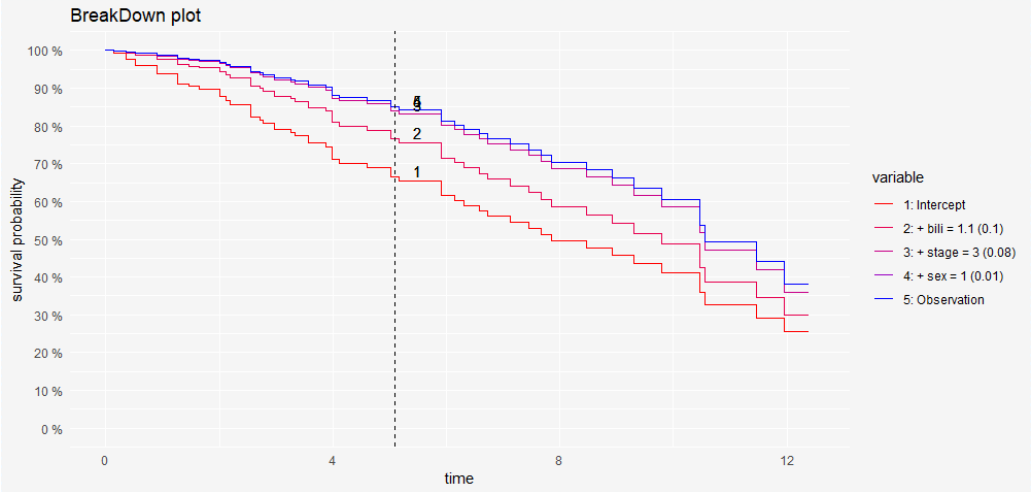

Figure 2: Break ~~Down~~ plot for Cox ~~Proportional Hazards~~ model. Variables bili and stage have highest impact on final prediction.

- **variable response** plot is designed to better understand the relation between a variable and a model output. ~~See an example in Figure 3.~~ ~~Variable~~ response plot illustrates how the mean survival curve change along with the changing values of the variable. It is inspired by partial dependence plots (Greenwell, 2017).

- **model performance** curves present prediction error for the chosen survival model depending on time. ~~See an example in Figure~~ 4. For computing prediction error we use the expected Brier Score (Mogensen, Ishwaran, & Gerds, 2012). At a given time point t, the Brier score for a single observation is the squared difference between observed survival status and a model-based prediction of surviving time t.

Conclusions and future work

Explainers implemented in the `survxai` package allow ~~exploring~~ one or more models in a feature-by-feature fashion. This approach will miss interactions between variables that may be handled by the models. The main problem with ~~integrations~~ is that number of interactions grows rapidly with the number of features what makes it hard to present in a readable form.

Acknowledgments

The work was supported by NCN Opus grant 2016/21/B/ST6/02176.

References

Biecek, P. (2018). DALEX: explainers for complex predictive models. *ArXiv e-prints*.

Biecek, P. (2018). *CeterisParibus: Ceteris paribus plots (what-if plots) for a single observation*. Retrieved from https://CRAN.R-project.org/package=ceterisParibus

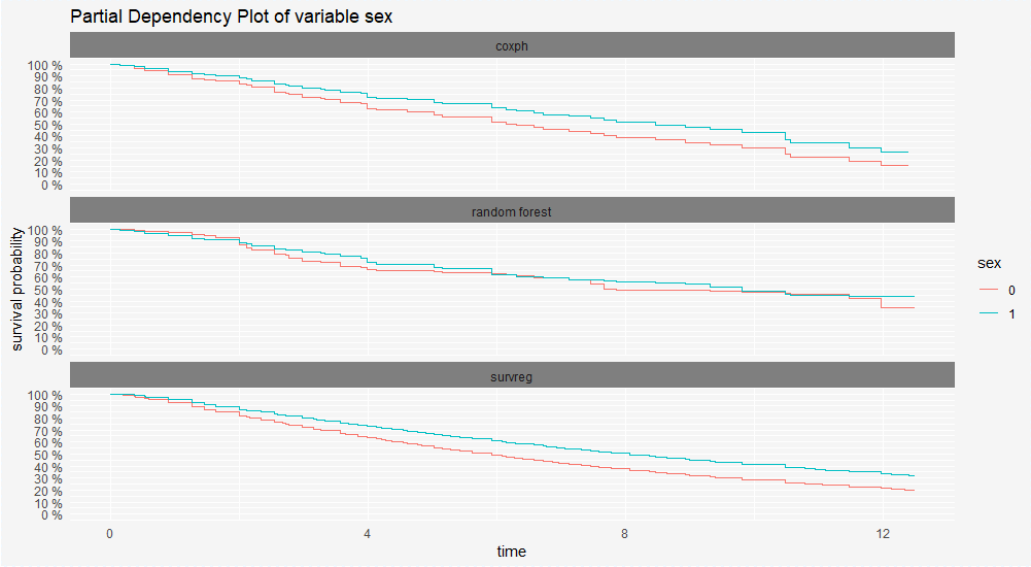

Figure 3: Variable response plot for three models and variable sex. In survival random forest the sex variable affects model predictions in a different way than in other models.

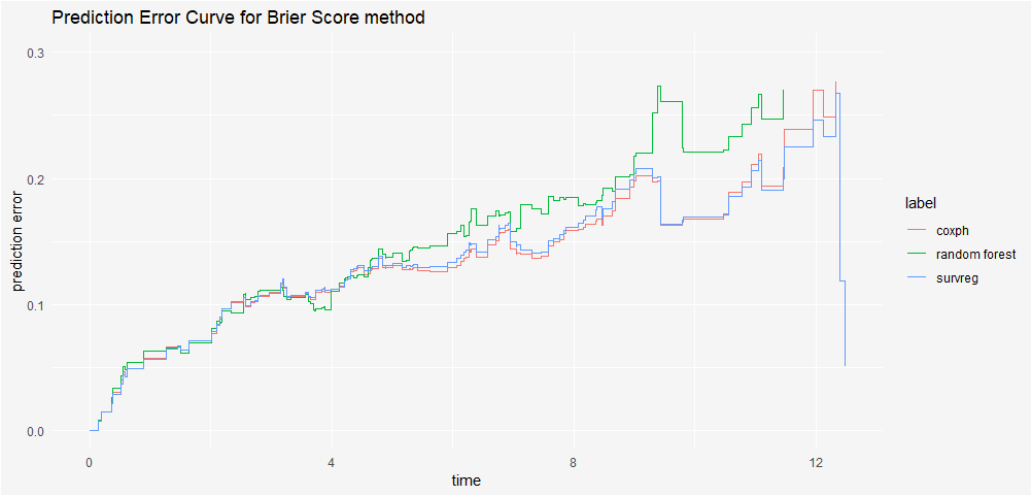

Figure 4: Model performance plot for three models. In random forest model predictions are less accurate after year 4.

Collett, D. (2015). *Modelling survival data in medical research, third edition*. Chapman & hall/crc texts in statistical science. CRC Press. Retrieved from https://books.google.pl/books?id=Okf7CAAAQBAJ

Dempsey, W. H., Moreno, A., Scott, C. K., Dennis, M. L., Gustafson, D. H., Murphy, S. A., & Rehg, J. M. (2017). ISurvive: An interpretable, event-time prediction model for mHealth. In D. Precup & Y. W. Teh (Eds.), *Proceedings of the 34th international conference on machine learning*, Proceedings of machine learning research (Vol. 70, pp. 970–979). International Convention Centre, Sydney, Australia: PMLR. Retrieved from http://proceedings.mlr.press/v70/dempsey17a.html

Eleuteri, A., Tagliaferri, R., Milano, L., De Placido, S., & De Laurentiis, M. (2003). A novel neural network-based survival analysis model. *Neural Networks*, *16*(5), 855–864. doi:https://doi.org/10.1016/S0893-6080(03)00098-4

Greenwell, B. M. (2017). pdp: An R Package for Constructing Partial Dependence Plots. *The R Journal*, *9*(1), 421–436. Retrieved from https://journal.r-project.org/archive/2017/RJ-2017-016/index.html

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests. *Ann. Appl. Statist.*, *2*(3), 841–860. Retrieved from http://arXiv.org/abs/0811.1645v1

Lu, J., & Park, O. (2003). Modeling customer lifetime value using survival analysis—an application in the telecommunications industry. *Data Mining Techniques*, 120–128.

Mogensen, U. B., Ishwaran, H., & Gerds, T. A. (2012). Evaluating random forests for survival analysis using prediction error curves. *Journal of Statistical Software*, *50*(11), 1–23. Retrieved from http://www.jstatsoft.org/v50/i11/

Molnar, C. (2018). Iml: An r package for interpretable machine learning. *Journal of Open Source Software*, *3*(26), 786. doi:10.21105/joss.00786

Staniak, M., & Biecek, P. (2018). Explanations of model predictions with live and breakDown packages. *ArXiv e-prints*.