

Datennutzungsvereinbarung

The data reported here have been supplied by the United Network for Organ Sharing as the contractor for the Organ Procurement and Transplantation Network. The interpretation and reporting of these data are the responsibility of the author and in no way should be seen as an official policy of or interpretation by the OPTN or the U.S. Government. The entire analysis is based on OPTN data as of June 20, 2020.

KURZFASSUNG

Die Erklärbarkeit von Prognosemodellen spielt insbesondere bei Entscheidungen, die das Leben eines Menschen gravierend beeinflussen können, eine wesentliche Rolle. Additive Regressionsmodelle bieten die Möglichkeit, gut erklärbare Modelle zu erstellen. Durch eine Vielzahl von Erweiterungen lässt sich die Basisvariante an verschiedene Anwendungsfälle anpassen.

Ziel dieser Thesis ist es verschiedene additive Modelle mit Lasso Regularisierung zu recherchieren, um Variablen und Interaktionen zu einer ausgewählten Variable zu selektieren. Im Theorieteil wurden dazu die mathematischen Unterschiede der Verfahren GGLasso, GLINTERNET, GAMSEL und Pliable untersucht.

In einem praktischen Anwendungsfall wurden die vier Methoden auf Nierentransplantationsdaten aus den USA angewandt. Der Kidney Donor Profile Index (KDPI), der für die Nierenqualität steht, spielt eine zentrale Rolle im Transplantationsprozess, daher wurden insbesondere Interaktionen zwischen Merkmalen des Empfängers und dem KDPI untersucht. Hierzu wurde mittels logistischer Regression der Überlebensstatus nach einem Jahr prognostiziert. Durch die Anwendung der oben genannten Methoden, wurden relevante Variablen und Interaktionen identifiziert. Das Modell konnte die Anzahl an Variablen deutlich reduzieren und gleichzeitig die Prognosegüte gegenüber eines Modells ohne Variablenselektion verbessern. Die Erkenntnisse, insbesondere über den Einfluss von Variablen und Interaktionen, wurden in einem interaktiven Dashboard dargestellt.

Um die Methoden systematisch zu vergleichen, wurde zusätzlich eine Simulationsstudie durchgeführt. Hierbei sollten die Verfahren einflussreiche Variablen und Interaktionen unter einer Vielzahl von Störvariablen identifizieren. Mittels dieser Studie und den Ergebnissen aus dem praktischen Anwendungsfall, konnten wesentliche Unterschiede und Probleme in den Verfahren festgestellt werden.

Schlagworte – Nierentransplantation; Additive Modelle; Pliable; GAMSEL; GLINTERNET; GGLasso.

ABSTRACT

The explainability of prediction models plays an important role, especially in cases where decisions have severe consequences on a person's life. Additive regression models provide an opportunity to create explainable models. The basis version can be adjusted to various use cases by numerous expansions.

The objective of this thesis is to research several additive models with lasso regularization to select variables and interactions with a chosen variable. For that in the theoretical part the mathematical differences between the methods GGLasso, GLINTERNET, GAMSEL and Pliable have been examined.

In a practical use case the four methods have been applied on kidney transplant data from the USA. The Kidney Donor Profile Index (KDPI), which represents the kidney quality, plays a major role in the transplantation process, therefore especially the interactions between attributes of the recipient and the KDPI have been examined. For this the survival outcome of the recipient has been predicted by the use of logistic regression models. Relevant variables and interactions were identified by applying the aforementioned methods. The final model had significant less variables and at the same time increased the predictive accuracy compared to a model without variable selection. An interactive dashboard has been created to show the findings, especially the effects of the selected variables and interactions.

To compare the methods systematically a simulation study has been conducted additionally. The methods had to identify important variables and interactions among numerous noise variables. By means of this study and the results of the practical use case significant differences and problems of the methods have been discovered.

Keywords – kidney transplant; additive models; Pliable; GAMSEL; GLINTERNET; GGLasso.