Deep Learning Forecasting Model for Kidney Function Post-Transplant Exploiting Missing Values

The 44th International Symposium On Forecasting, Dijon, France

Zahrasadat Sajjadifar[1], Axel Geysels[1], Alexandre Arnould[1], Maarten Naesens[2], Oscar Mauricio Agudelo[1], and Bart De Moor[1]

[1]KU Leuven, Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Leuven, Belgium
[2]KU Leuven, Department of Microbiology, Immunology and Transplantation, Nephrology and Kidney Transplantation Research Group, Leuven, Belgium

July 2024

Abstract

While regular and complete time series with evenly spaced timestamps enhance the potential for analysis and modeling, in real-world scenarios irregular time series with missing values are common which pose challenges in effective data handling and modeling. For solving the missing data problem in time series analysis, there are two main approaches either to impute the data before starting the prediction tasks or modify and apply changes directly to the predictive algorithms in a way that they become aware of irregularity in time series and be able to handle the missing values. In our research, we focus on clinical biomarkers time series, particularly the estimated glomerular filtration rate (eGFR) monitored after kidney transplantation for each recipient. The level of eGFR can be interpreted as an alert for performing a transplant biopsy if the values are not in a normal range. However, this normal range is patient-specific and there is no global threshold to determine the abnormal range. The clinically relevant task is to predict future values in a specific time interval from past

observations. The forecasted future eGFR values for each graft can be considered as a patient-specific expected range and the newly measured eGFR values can be benchmarked against the predicted ones to evaluate variances from the expected range and this is going to prompt the patient-specific alarm in clinical practice. Given the frequent tracking of eGFR after transplant, an irregular time series emerges. The missingness in such time series is not random but it is systematic since the missing rate increases over time resulting in larger gaps due to the decrease in the follow-up rate of patients. The contribution of this study is implementing a one-stage sequence-to-sequence forecasting model modified to handle missing values using the gated recurrent unit with decay rate (GRU-D) to exploit the systematic missingness in the predictive model and compare the results with two-stage imputation and forecasting models based on the conventional autoregressive integrated moving average (ARIMA) and sequence-to-sequence recurrent neural network based on the gated recurrent unit (GRU).

Acknowledgements

- KU Leuven: Research Fund (projects iBOF/23/064, C3/20/117, C3I-21-00316), Industrial Research Fund and several Leuven Research and Development bilateral industrial projects;
- Flemish Government Agencies:
  - FWO: SBO project S005319N, TBM Project T001919N; FWO PhD Grant 11K5623N;
  - EWI: the Flanders AI Research Program
  - VLAIO: CSBO (HBC.2021.0076) Baekeland PhD (HBC.20192204)
- EU: This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 885682). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or ERC. Neither the European Union nor the granting authority can be held responsible for them.