Master thesis on the analysis of routine data from outpatient medical care

Abstract

In cooperation with the Institute of Family Medicine at the University of Lübeck, the Department Methods for Big Data at the Karlsruhe Institute of Technology, Scientific Computing Center, is looking for a **master student at the end of a Statistics or Data Science studies to write a master thesis** in the context of a research project on the statistical analyses of routine data from outpatient medical care. The methods used and potentially extended/further developed will be at the intersection of Statistical and Machine Learning. The aim is to combine classical tabular data (such as clinical variables) with unstructured data (such as text or images) to better understand and predict key quantities from outpatient medical data, such as occurrence of certain illnesses or disease events, change of physician or missed appointments.

The data

Timely secondary data analysis is still a rarity in Germany. The Institute for Family Medicine has developed an infrastructure for the daily analysis of anonymized patient data from several primary care practices. The data can be extracted directly from the electronic health records (EHRs) of the participating practices. To date, there is no comparable infrastructure for the analysis of ambulatory care data in Germany. The number of participating practices will be increased in the coming months.

Interpretable modelling of future events in family practice using structured and unstructured features

In addition to structured data, the EHRs from the ambulatory care providers contain a variety of unstructured data, such as clinical notes or clinical measurements collected at irregular intervals (e.g., laboratory tests). These data could play a critical role in predicting future events the in primary care practices. Traditional statistical models are limited in their ability to use such complex data for prediction task. In contrast, artificial neural networks have proven to be very powerful, but at the cost of making it very difficult to understand how the prediction was made (black box). This is especially important in the field of health care. The goal of a master thesis is to develop and apply methods of interpretable artificial intelligence to the prediction of future events in general practice (e.g., change of doctor, missed appointments, occurrence of certain diseases or disease events) using structured and unstructured features extracted from EHRs. The advantages of classical statistical models (interpretability) will be combined with

those of artificial neural networks (taking into account information from unstructured and irregularly collected data).

Key literature

- Rügamer, D., Kolb, C., & Klein, N. (2024). Semi-Structured Distributional Regression. The American Statistician, 78: 88-99.
- Dunstan, J., Villena, F., Hoyos, J. P., Riquelme, V., Royer, M., Ramírez, H., & Peypouquet, J. (2023). Predicting no-show appointments in a pediatric hospital in Chile using machine learning. *Health Care Management Science*, 1-17.
- Huguet, N., Kaufmann, J., O'Malley, J., Angier, H., Hoopes, M., DeVoe, J. E., & Marino, M. (2020). Using electronic health records in longitudinal studies: estimating patient attrition. *Medical care*, *58*(Suppl 6 1), S46.

Key requirements

- Excellent programming skills in R and Python. Experience with tensorflow/Pytorch desirable
- Willingness to work in an interdisciplinary team and interest in both the development and application of advanced statistical learning methods and health care
- Excellent English skills
- Profound experience with Latex
- Interest in contributing to edge-cutting research and potential publications

Contact

If you are interested, please contact Prof. Dr. Nadja Klein or Dr. Christoph Strumann by mail:

Dr. Christoph Strumann
Institute for Family Medicine
UNIVERSITÄTSKLINIKUM
Schleswig-Holstein, Campus Lübeck
Ratzeburger Allee 160 | Haus 50 | 23538
Lübeck
c.strumann@uni-luebeck.de

Prof. Dr. Nadja Klein
Department Methods for Big Data
Scientific Computing Center
Karlsruhe Institute of Technology
Zirkel 2
76131 Karlsruhe
nadja.klein@kit.edu