Predicting Risks of Readmission and Mortality for Releasing ICU Patients - A Comparison of various Machine Learning Techniques

L.C. Herstix

Abstract

The amount of data recorded in hospitals is increasing, especially in intensive care units (ICUs). This can be challenging for caregivers since they are not able to evaluate the data manually. However, the collected data also offers tremendous opportunities. Combined with machine learning techniques, many insights can be retrieved that could help caregivers find treatments for their patients. Therefore, this thesis compares the performance of four machine learning algorithms for predicting readmission and mortality risks. In total, eight predictive models are investigated. The four algorithms used are the logistic regression, random forest, multilayer perceptron (MLP) and support vector machines (SVM). The underlying dataset used is the Medical Information Mart for Intensive Care III (MIMIC-III) dataset. A daily model was used for data aggregation and the high imbalance in the dataset was handled by weighing the classes. The outcome of the predictions showed high discrepancies depending on the machine learning technique used. The best performance was achieved by SVM (mortality prediction: AUC of 0.994, readmission prediction AUC of 0.984) followed by random forest (mortality prediction: AUC of 0.931, readmission prediction AUC of 0.940). Therefore, both SVM and random forest produced extremely good results with medical significance (AUC over 0.80) for both predictive cases. Multilayer perceptron was not able to produce any reasonable results since the minority class was not recognized. Logistic regression performed poorly for the prediction of mortality (AUC of 0.62) and only slightly better than random for the prediction of readmission. Further, the differences in the AUC values between the two predictive cases, mortality and readmission, are only approximately 0.01 for both SVM and random forest. This indicates that the approach for data preparation and aggregation followed in this thesis is comprehensive enough to work for diverse predictive cases and that it is equally possible to predict readmission and mortality after discharge.