Mortality and Death Time Prediction Models using MIMIC-III

Suiki LAU

CSE 6250
OMSCS
Georgia Institute of Technology

Overview

Background
Objective
Data Source
Implementation
Methodology
Experimental Results
Discussion
Conclusion

Background
Objective
Data Source
Implementation

Literature Review

- Over the past few decades, several ICU scoring systems have been developed using rule-based method / data mining approach, e.g. APACHE, SAPS, SOFA.
- Recently, better mortality prediction models trained on larger input features have been developed using machine learning approach. e.g.
    - Applying Latent Dirichlet Allocation to free-text hospital notes
    - Using Recurrent Neural Network to build benchmark models
    - Applying ensemble methods to improve mortality prediction

- However, most mortality models in the literature were designed for at least 24 hours or 48 hours after ICU admission.
- Could we predict in-hospital mortality in the early stage of ICU, say 6 hours since ICU admission?

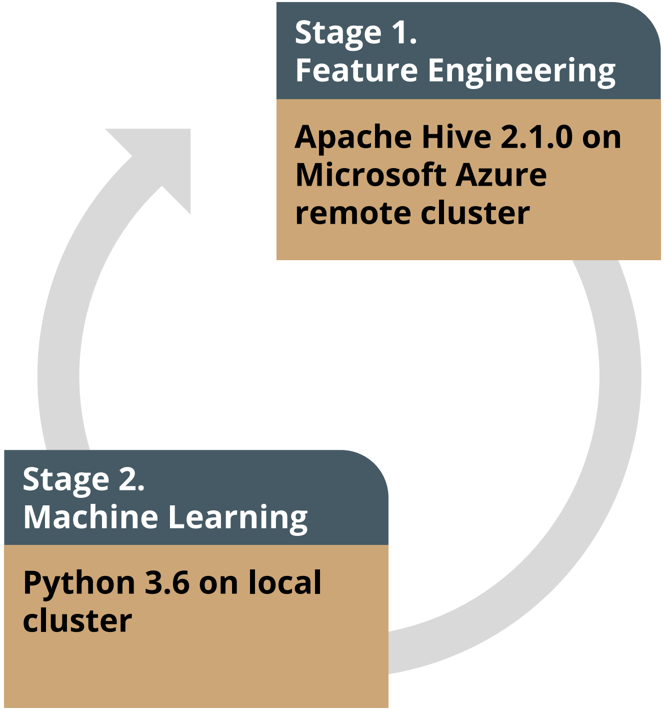
Objective

- The goal is to build a two-phase model framework to predict
  (1) in-hospital mortality and
  (2) death hours since ICU admission
  during the early stage of ICU stay, i.e. first 6-hour since ICU admission.
- The model would be useful to promptly identify high-risk patients who might be dead within hours or days since ICU admission, so that resources can be efficiently allocated during the early stage of ICU stay.

MIMIC-III

- A freely-accessible database of ICU data
- Providing de-identified electronic healthcare records (EHR) of over 60,000 ICU stays for ~40,000 patients who stayed in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012
- Comprising static or temporal information about patients' demographic characteristics, such as gender, age, ethnicity, and various in-hospital measurements, lab tests, procedures and medication of patients during their ICU stay => input features
- Comprising mortality, death time information => predicted labels
- More information at https://mimic.physionet.org/

Implementation



- The project has been implemented in 2 stages.
- MIMIC-III requires around 50GB of space. We chose big data tool like Apache Hive to perform data preprocessing and feature engineering since the dataset is quite large. The process was deployed on Microsoft Azure remote cluster (2 head nodes and 1 worker node, each with 200GB space, 14GB RAM, and 4 processors).
- Data output in Stage 1 was used as feature input for the model training in Stage 2.
- We then build our machine learning pipeline using Python 3.6 on a local cluster (500GB space, 16GB RAM, 4 processors, 4GB GPU).
- The development process was iterative.

Methodology

Exploratory Data Analysis

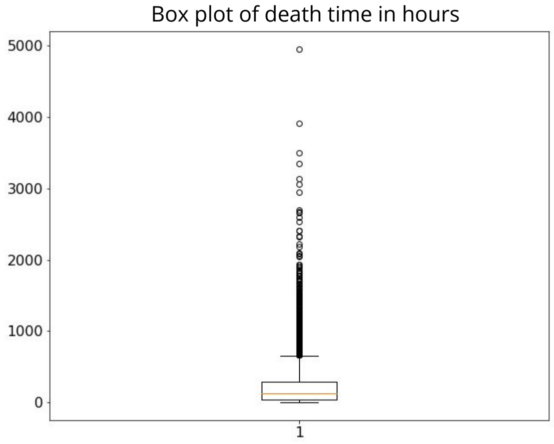
- The original dataset consists of 61,532 distinct ICU stays of 46,520 unique patients.
- To form our study population, we filtered out patients with
    - ICU stays < one hour
    - age less < 16 or age > 89
- The final study population consists of 49,632 ICU stays of 36,343 patients.

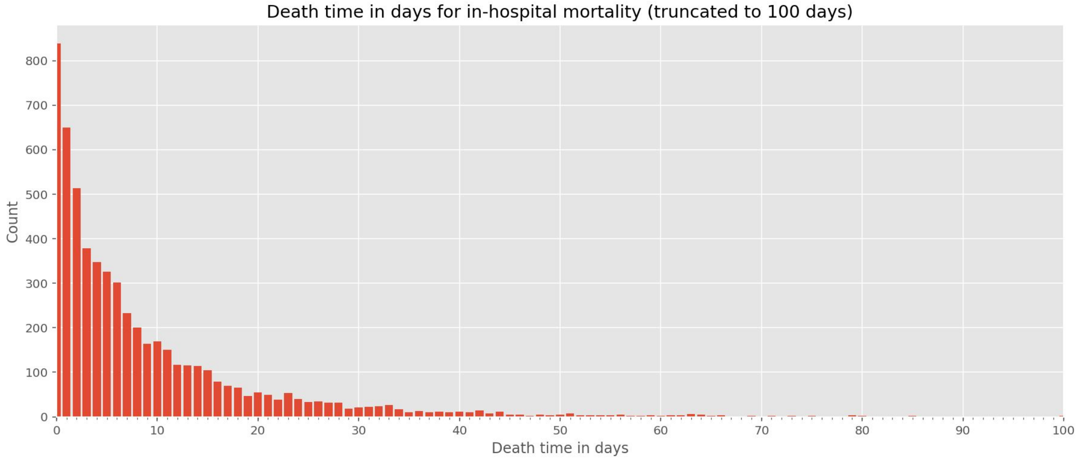
Variables	Statistics
Age	Mean 62.61
Gender	Male 57.79%
Ethnicity	White 71.00%
Admission type	Emergency 82.31%
Number of ICU stays	Mean 1.37
In-hospital mortality ratio	11.62%

Summary statistics of the study population

Exploratory Data Analysis

- Next, we consider only the dead patients with non-negative death time since ICU admission.
- There are 5,718 in-hospital mortality.
- Their average death time since ICU admission is 9.57 days, maximum death time is 206.38 days and minimum death time is 0 day.





Feature Engineering

- For each ICU stay, we have extracted data from the first 6 hours, 12 hours and 24 hours since ICU admission.
- There are altogether 123 extracted features covering 5 static variables and 40 physiological variables.
- The static variables includes
  - admission type
  - number of ICU stays
  - demographic features such as age, gender and ethnicity
- The temporal data of physiological variables includes
  - patients’ vital signs e.g. heart rate and blood pressure
  - Glasgow coma scale
  - blood gases and chemistry values
  - laboratory results
  - urine output
- Most of the temporal variables were aggregated by maximum, minimum, and average during the specified timeframe, except that urine output was aggregated by sum.

Machine Learning

- We propose a two-phase model framework to predict in-hospital mortality and death time in hours.

Phase 1

A binary classifier was trained on the extracted features to predict in-hospital mortality

Phase 2

A multiclass classifier was trained on the same set of extracted features to predict death time in hours since ICU admission for the predicted dead patients in Phase 1

Machine Learning

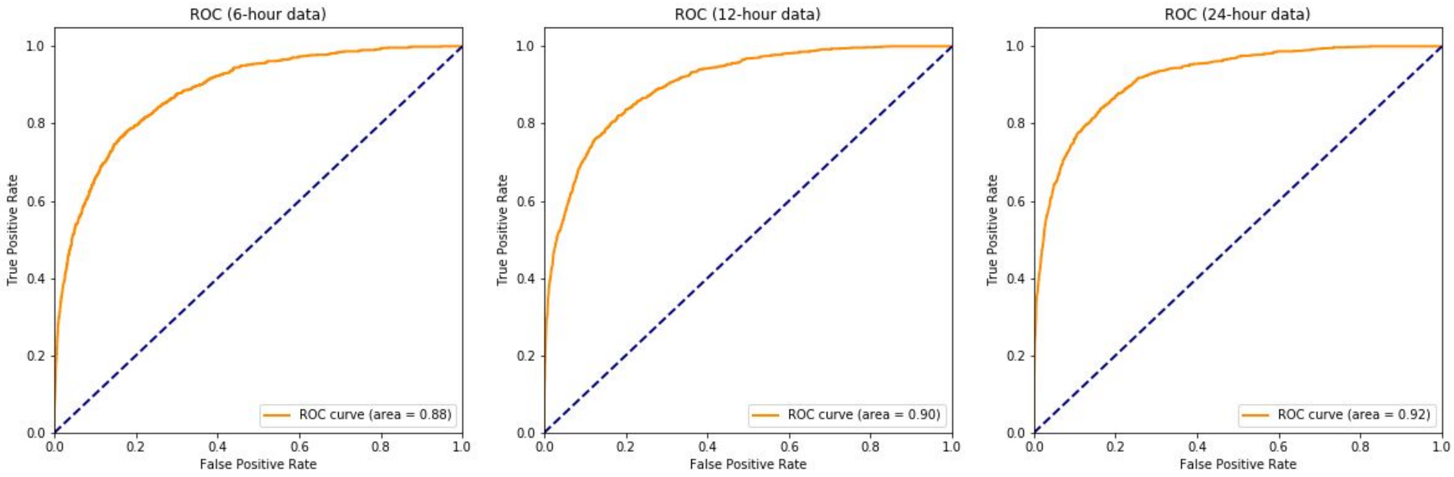
- In Phase 1, the study population consists of 49,632 ICU stays. We split the data into 80% training set and 20% test set.
- We then built a custom machine learning pipeline to select features, transform data, impute missing values and train a **random forest classifier using grid search on 5-fold CV** to predict the in-hospital mortality label.
- We also test the trained model (best parameter set from grid search) on the test set, and compare the model result using 6-hour, 12-hour and 24-hour data.

- In Phase 2, we filtered out dead patients with negative death time. A total of 5,718 ICU stays of dead patients were split into 80% training set and 20% test set.
- We label each data to one of the three specified classes, then built a custom machine learning pipeline to train a **multiclass random forest classifier using grid search on 5-fold CV** to predict the death time label.
- We also test the trained model (best parameter set from grid search) on the test set, and compare the model result using 6-hour, 12-hour and 24-hour data.

Experimental Results

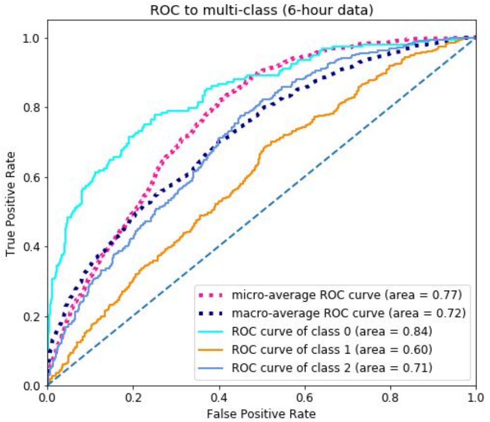
Phase 1 Evaluation

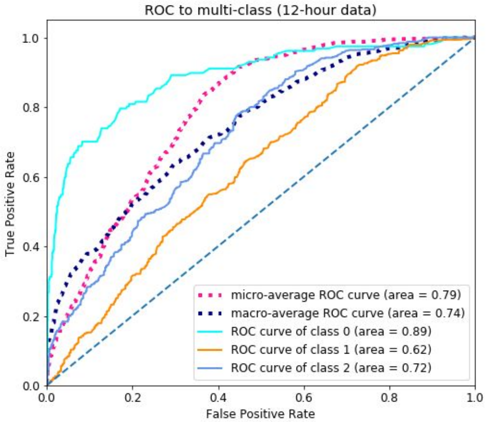
- The predictive performance of the mortality model trained on 6-hour data has competitive performance (AUROC 0.88) with the same model trained on 12-hour and 24-hour data (AUROC 0.90 and 0.92 respectively).

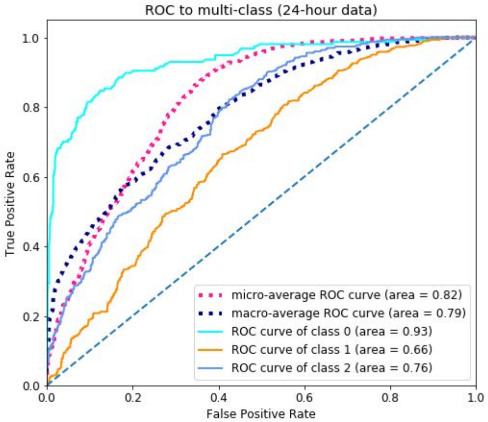


Phase 2 Evaluation

- The death time multiclass classifier trained on 6-hour data also provides an effective base (micro-average AUROC 0.77) to give a rough estimate of death hours since ICU admission. The result is competitive to the models trained on 12-hour or 24-hour data (micro-average AUROC 0.79 and 0.82 respectively).







Discussion
Conclusion

Discussion

- Can we use regressor instead of classifier to more precisely predict death time in Phase 2?
- Can we use LSTM to handle time series data instead of using aggregation of features?

Conclusion

- Although the models trained on the data in the first 24-hour since ICU admission give better performance, the first 6 hours of ICU data provides enough information for mortality prediction and a rough estimate of death hours since ICU admission.
- The proposed framework provides a base to promptly identify high-risk patients who might be dead within hours or days since ICU admission in the early stage of ICU stay, and there are potential avenues for improvement.