

MLND Capstone proposal: Predicting In-Hospital Mortality of ICU Patients Using Dynamic Time Warping

Ka Hung Wong

Introduction

Advanced and accurate warning of mortality risk could improve the quality and performance in the Intensive Care Unit (ICU) because it can achieve better clinical decision-makings and reduce the wait times for medical interventions (Calvert et al. 2016; Siontis 2011). Acuity scores which are calculated based on patient's features are currently applied in ICU to characterize the severity of a patient's illness. Unfortunately, these models have only modest specificity and sensitivity when they are implemented to predict the patient mortality (Calvert et al. 2016). This motivated PhysioNet to launch challenge in developing predictive model on in-hospital mortality in 2012 ([PhysioNet 2012](#)). The data used for the challenge is consisted of 5 general descriptors and 36 time series of vital signs and laboratory results from the first 48 hours of the first available ICU stay of patients from MIMIC-II dataset. Since it is difficult to have all measurements, it would be better to achieve prediction by a small subset of data. This proposal is inspired by the shortage of full dataset and aims to develop a prediction model based on a small subset of ICU data.

Problem

The problem to be solved in the project belongs to binary classification in supervised learning, in which in-hospital ICU mortality is predicted by the learner trained by labelled dataset.

The labeling of the dataset is defined by the in-hospital death. The label of in-hospital death is marked as positive when the length of stay in days in the hospital is longer than the survival time (the number of days between ICU admission and death if the patient's death was recorded). Otherwise, the in-hospital death is marked as negative.

Datasets and Inputs

The dataset used in the project will be extracted from Medical Information Mart for Intensive Care (MIMIC-III) database (Johnson et al. 2016). MIMIC-III contains information related to patients admitted to critical care units at the Beth Israel Deaconess Medical Center in Boston, Massachusetts. The database contains data associated with 53,423 distinct hospital admissions for adult patients (aged ≥ 16 years) admitted to critical care units between 2001 and 2012.

Database includes 26 tables containing admission information, patients' information such as physiological and laboratory measurements, and more. In the project, MIMIC-III database will be built locally using PostgreSQL. The dataset will be extracted from 7 tables from the database: *admissions*, *chartevents*, *d_items*, *d_labitems*, *icustays*, *labevents*, and *patients*.

Cohort selection is based on three tables: *patients*, *admissions*, and *icustays*.

- *patients*: information about a patient that does not change
- *admissions*: information recorded on hospital admission
- *icustays*: information recorded on intensive care unit admission

The constraints used in selection are:

- Patients are older than 16 years old
- Length of ICU stay is longer than 1 day but less than 10 days
- The first ICU admission during ICU stay
- The unexpected medical events

With the defined cohort, 8 time series of physiological measurements from the first 24 hours of ICU stay of patients will be extracted as the features of the data. The measurements includes heart rate, blood pH, blood pressure, respiration rate, blood oxygen saturation, temperature, and white blood cell count (Calvert et al. 2016).

Approach

Distance-based methods are widely used in classifying time series (Xing, Pei, and Keogh 2010). The idea is to apply distance-dependent classical classification algorithms such as kNN or SVM on time series. In this approach, the distance (or similarly) between two time series are required. One of the popular way to calculate the similarity between two time series is dynamic time warping (DTW) (Kaya and Şule Gündüz-Öğüdücü 2015). In the project, the time series extracted from MIMIC-III will be cleaned and re-sampled to regular time interval. Then, DTW with kNN or SVM will be applied to perform the classification.

Benchmark Model

Simplified acute physiology score (SAPS) II and naive prediction are used as benchmark models. SAPS II is one of the acuity scores widely applied in clinical studies (Gall 1993). It is based on 12 physiological variables in the first 24 hours of ICU admission and admission health status of patients. SAPS II was calibrated to allow conversion of the calculated score into the probability of mortality ([score to prob](#)). The score and the converted probability can be calculated in MIMIC III dataset by executing the SQL script, which is available in

MIMIC Code Repository hosted on GitHub ([SAPSII](#)). A naive prediction of 100% mortality is used as the second benchmark model since the dataset is imbalanced with higher survival ratio.

Metrics

Since it is important to predict the mortality precisely, precision is thus one of the metrics need to be considered. Also, it is bad to have a lot of false negative since we do not want to miss any high risk patients. Therefore, F1 score is a good choice to evaluate the developed model and benchmarks.

Project Design

The workflow can be summarized:

1. Build MIMIC-III database

The MIMIC-III database is developed in PostgreSQL at Windows platform. Psycopg is used as a database adapter to execute SQL script in Python environment.

2. Cohort selection

SQL scripts are required to identify the interested group of patients for analysis.

3. Data extraction

SQL scripts are required to extract the patient's information on physiological and laboratory measurements. The extracted data will be stored in forms of CSV format.

4. Data cleaning

Pandas will be used to load the CSV data and to perform data cleaning. Two processing are required before analyzing the data. First, the time series from database has irregular time interval and re-sampling is required. Second, there is missing data in the data and imputation is needed. Third, the cleaned data will be divided into training, validation, and hidden testing sets.

5. DTW implementation

DTW will be implemented in Python using dynamic programming.

6. Machine learning algorithm implementation

Distance based algorithm such as kNN and SVM will be the candidate for the binary classifiers. Scikit-learn library will be used to perform the algorithms.

7. Metric evolution

The results will compare with the benchmarks (SAPS II and naive predictor). Parameter tuning will be performed to boost the results.

References

- Calvert, Jacob, Qingqing Mao, Jana L. Hoffman, Melissa Jay, Thomas Desautels, Hamid Mohamadlou, Uli Chettipally, and Ritankar Das. 2016. "Using Electronic Health Record Collected Clinical Variables to Predict Medical Intensive Care Unit Mortality". *Annals of Medicine and Surgery* 11 (November). Elsevier BV: 52–57. doi:10.1016/j.amsu.2016.09.002.
- Siontis, George C. M. 2011. "Predicting Death". *Archives of Internal Medicine* 171 (19). American Medical Association (AMA): 1721. doi:10.1001/archinternmed.2011.334.
- Johnson, Alistair E.W., Tom J. Pollard, Lu Shen, Li-wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. 2016. "MIMIC-III a Freely Accessible Critical Care Database". *Scientific Data* 3 (May). Springer Nature: 160035. doi:10.1038/sdata.2016.35.
- Xing, Zhengzheng, Jian Pei, and Eamonn Keogh. 2010. "A Brief Survey on Sequence Classification". *ACM SIGKDD Explorations Newsletter* 12 (1). Association for Computing Machinery (ACM): 40. doi:10.1145/1882471.1882478.
- Kaya, Hüseyin, and Şule Gündüz-Öğüdücü. 2015. "A Distance Based Time Series Classification Framework". *Information Systems* 51 (July). Elsevier BV: 27–42. doi:10.1016/j.is.2015.02.005.
- Gall, J. R. Le. 1993. "A New Simplified Acute Physiology Score (SAPS II) Based on a European/North American Multicenter Study". *JAMA: The Journal of the American Medical Association* 270 (24). American Medical Association (AMA): 2957–63. doi:10.1001/jama.270.24.2957.