Georgia Tech College of Computing
Computational Science and Engineering

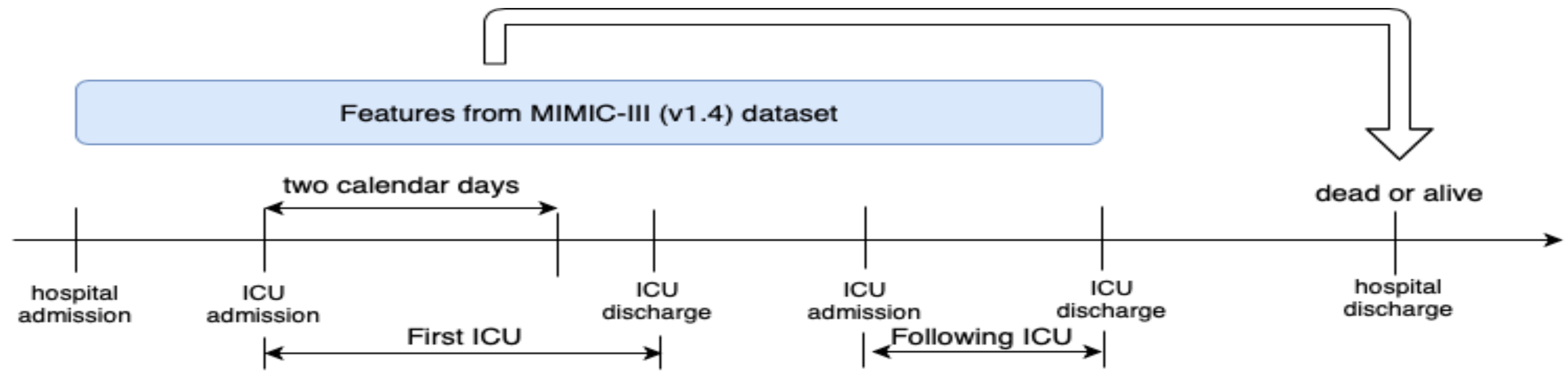
Fall 2019, CSE6250 Project, Team 1

Building Deep Learning Models to Predict Mortality in ICU Patients

Yuanfei Bi

Project Overview

Goal: To identify adult patients death risk in the ICU based on the MIMIC-III (v1.4) data accumulated by the first 48 hours into the first ICU stay.



Risks: MIMIC-III (v1.4) datasets have lots of erroneous entries due to noise, missing values, outliers, duplicate and clerical mistakes etc. And feature selection and extraction is critical for project success.

Payoffs: The project will allow us to apply many aspects of the BD4H course and explore sophisticated deep learning techniques.

Impacts: This project shed light to empower deep learning models into health-related project.

Data Statistics

MIMIC-III (v1.4)

# of patients	46520
# of adult patients [a]	38597
Median age of adult patients	65.8 years
In-hospital mortality of adult patients	11.5%
# of admissions	58976
# of ICU stays	61532
# of ICU stays of adult patients	53423
# of long ICU stays [b] of adult patients	53133
# of the first long ICU stay of adult patients	38418
Avg. length of long ICU stays of adult patients	4.17 days
Avg. length of ICU stays of adult patients	4.14 days
Avg. length of the first long ICU stays of adult patients	4.07 days

[a] Adults: $\geq$ 16 years old.
[b] Long ICU stays: $\geq$ 4 hours.

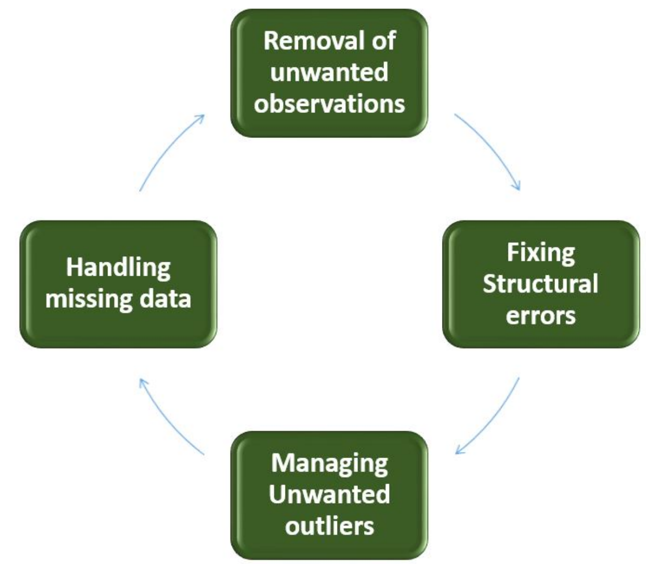
Features

SAPS-II

Features	ItemID	Item Name	Table
glasgow coma scale	723	GCSVerbal	chartevents
	454	GCSMotor	chartevents
	184	GCSEyes	chartevents
	223900	Verbal Response	chartevents
	223901	Motor Response	chartevents
	220739	Eye Opening	chartevents
systolic blood pressure	51	Arterial BP [Systolic]	chartevents
	442	Manual BP [Systolic]	chartevents
	455	NBP [Systolic]	chartevents
	6701	Arterial BP # 2 [Systolic]	chartevents
	220179	Non Invasive Blood Pressure systolic	chartevents
	220050	Arterial Blood Pressure systolic	chartevents
heart reate	211	Heart Rate	chartevents
	220045	Heart Rate	chartevents
body temperature	678	Temperature F	chartevents
	223761	Temperature Fahrenheit	chartevents
	676	Temperature C	chartevents
	223762	Temperature Celsius	chartevents
pao2 / fio2	50821	PO2	labevents
	50816	Oxygen	labevents
	223835	Inspired O2 Fraction (FiO2)	chartevents
	3420	FiO2	chartevents
	3422	FiO2 (Meas)	chartevents
	190	FiO2 set	chartevents

urine output, serum urea nitrogen level, white blood cells count, serum bicarbonate level, sodium level, potassium level, bilirubin level, age, immunodeficiency syndrome, hematologic malignancy, metastatic cancer, admission type.

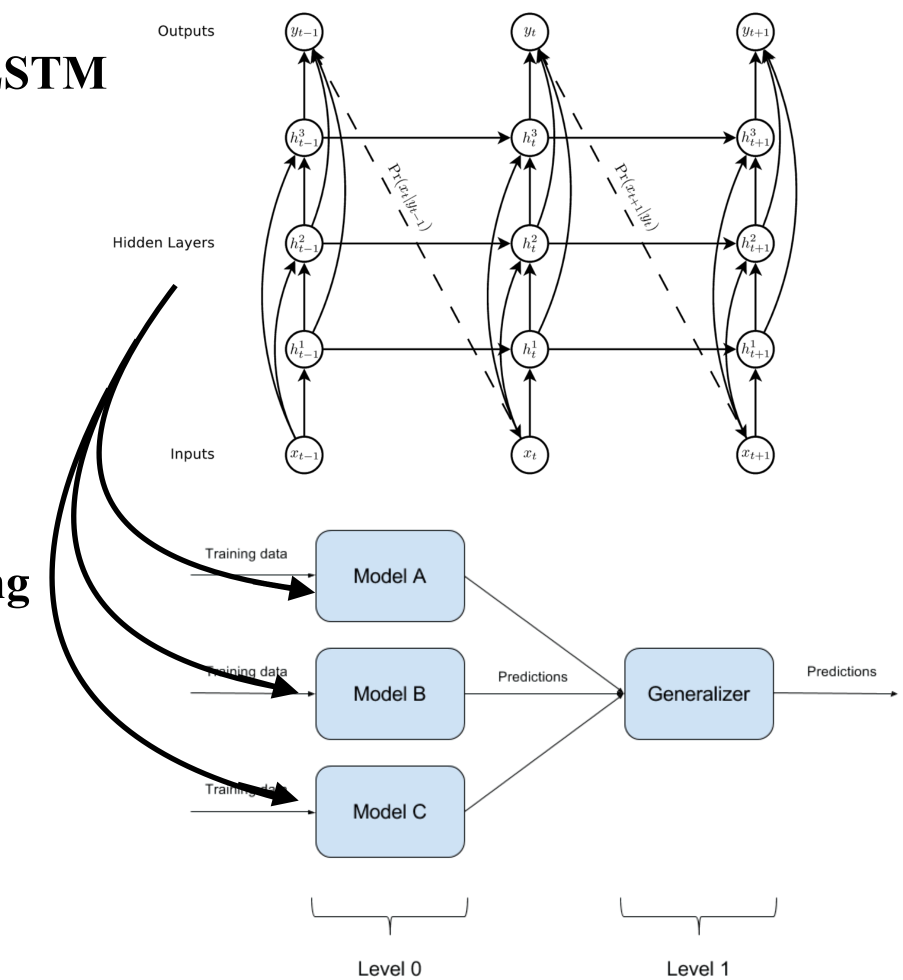
Data Cleaning



Models

RNN-LSTM

Stacking



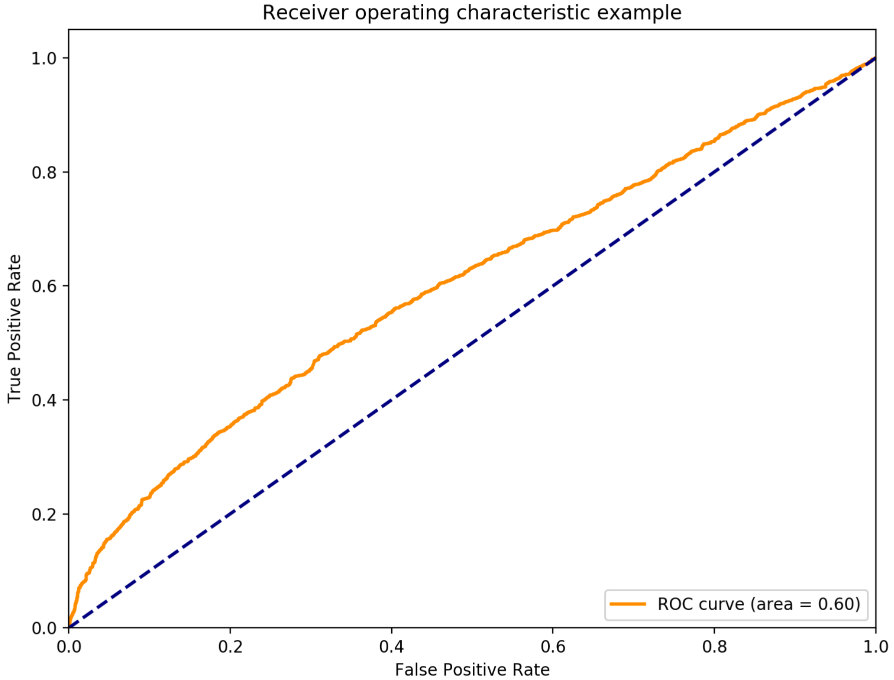
v.s. logistic regression

Experiments Setup

- Data statistics: PostgreSQL
- Feature selection and extraction: Spark
- LSTM ensemble model and logistic regression model: Python/Pytorch

Prediction Results

Model	Precision	Recall	F1	AUC
RNN-LSTM model	0.620	0.711	0.662	0.600
Logistic Regression	0.610	0.650	0.620	0.560



Conclusions & Discussion

- The RNN-LSTM model consistently outperforms the baseline logistic regression model. Particularly, the AUC of the RNN-LSTM model is higher than logistic regression by 4%.
  Able to take advantage of sequential nature of time-series features.

- It is important to perform sequential analysis of clinical data. The RNN-LSTM model implemented in this work has only three layers which lacks capability to efficiently capture both sequential and non-sequential features.

- It is worthwhile to mention that identifying efficient features to predict ICU survival is not trivial. Some other data which are available from the MIMIC-III dataset, such as fluid balance and monitor data, have not been incorporated in this work.

Future Work

- More sophisticated data preprocessing steps and deep learning models will be conducted to capture the characteristics of the massive MIMIC-III datasets.

- More extensive ICU datasets will be employed to evaluate and improve the models.

- Aggregate additional features and quantifying their impact on prediction accuracy.