Readmission Predict

Wanxue Dong Xuan Wang Yizhuo Li | Prof. Ding Ying | The University of Texas at Austin

Problem

Unplanned readmission to the ICU results in high risk and high cost for the patients.

Hypothesis

- RNN with LSTM can predict the readmission probability with text data.
- XGboost and LightGBM can do the same job with non-text data.
- If two models can be used together, the result will be better.

Project Overview

- Evaluate our models on the MIMIC III public dataset.
- Focus on 30 days readmission prediction.
- Tree based models like Xgboost and LightGBM will deal with numerical and categorical features.
- NLP methods will be used to process text data like doctors' notes and demographic information of the patients.
- Plan to used the pretrained ClinicalBERT model
- We will compare the results from the two models and other results from related works

Features / Metrics

Features for the Tree	Features for NLP	Dependent variable
• Albumin • Glasgow eye • Blood platelets • Age • Temperature etc.	• Early Notes • Discharge summary	• PR curve • ROC curve • Accuracy • Recall

Procedure - Boost

Steps	Details
Data Preparation	Data cleaning, transformation and integration
Feature Selection	Leverage regularization and display top N features
Modeling	Use XGBoost and decision trees to fit the model
Prediction	Predict the readmission rate
Evaluation	Cross-validation (during each iteration); AUROC; Confusion Matrix

Procedure - NLP

Step 1	Step 2	Step 3	Step 4
MIMIC	BERT	natural learning processing NLP language	
preprocess the text data in MIMIC dataset	ClinicalBERT model Or Fine tune the BERT	Train the Model	Evaluate and tune the Model

MIMIC-III Data

- Overview : MIMIC-III integrates de-identified, comprehensive clinical data of patients admitted from 2001 to 2012.

- Possible patient types : Diabetes, Heart attack, Septic shock, Lupus, Stroke etc.
- Measurement / Model predictors : Heart Rate, Temp F, O2 Saturation, NIBP - Mean, Fluids, Creatinine, Alarms On etc.

Related works

	3-days readmission (CIR=15.8%)	7-days readmission (CIR=19.7%)	15-days readmission (CIR=23.5%)	30-days readmission (CIR=27.3%)
Mean AUROC (SD)	0.802 (0.011)	0.809 (0.008)	0.803 (0.007)	0.795 (0.004)
Mean recall (SD)	0.837 (0.016)	0.821 (0.015)	0.799 (0.011)	0.778 (0.013)
Mean precision (SD)	0.421 (0.01)	0.494 (0.012)	0.547 (0.013)	0.587 (0.012)
Mean f1 (SD)	0.363 (0.024)	0.439 (0.022)	0.482 (0.015)	0.502 (0.009)

Table 1. Mean and standard deviation (SD) of the metrics measuring model performance for different outcomes, computed over a stratified 10-fold cross-validation. CIR = class imbalance ratio (proportion of positive cases among the population).

- We would evaluate and compare our work with Nguyen et al.'s (2019) prior work which proposed a model based on a tree boosting method to predict ICU readmission at 3, 7, 15 and 30 days. The results (Table 1) comes from using the tree boosting method on patient data at discharge on the freely available MIMIC-III database. (Nguyen et al., 2019)
- XGBoost, a scalable end-to-end tree boosting system, is widely and effectively used in machine learning predicting problems (Chen and Guestrin, 2016). XGBoost would be beneficial to extract information from numerical and categorical features.
- Reddy and Delen utilizes deep learning methods to predict rehospitalization within 30 days. The authors compare LSTM with other traditional classification methods such as penalized logistic regression and artificial neural networks. (Reddy and Delen, 2018) We would take advantage of LSTM model's ability of processing text to help understand doctors' diagnosis notes.

Expected Outcomes

- Compare results of our method with other existing methods on readmission prediction by applying them on MIMIC-III dataset.
- Evaluate the limitations and specify the directions for further research.

Works Cited

- Nguyen, D. P., Paris, N., & Parrot, A. (2019). Accurate and reproducible prediction of ICU readmissions. *medRxiv*.
- Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).
- Reddy, B. K., & Delen, D. (2018). Predicting hospital readmission for lupus patients: An RNN-LSTM-based deep-learning methodology. Computers in biology and medicine, 101, 199-209.