CSE 6250: Predicting Mortality in ICU Setting (Project Draft).

Vincent La[*], Avi Ananthakrishnan[†]

November 11, 2018

Abstract

FILLER

---

[*]vincent.la@gatech.edu

[†]avinash.ananthakrishnan@gatech.edu

1 Introduction

Accurate knowledge of patient's clinical state is critical in a clinical setting. Patients in an intensive care unit are particularly in a critical state. Thus, if we can have better accurate mortality prediction within an ICU, we can use this to better understand which patients should be prioritized for clinical care and how we can better allocate resources. Within an ICU, there are many measurements being made from from various devices, but we know that there are both problems in terms of false positives and false negatives. While false negatives might seem to be more problematic, being too sensitive and introducing a lot of false positives also has problems as well. Too many false positives results in information fatigue and providers lose trust in signal as it's hard to separate from noise. By improving mortality prediction using data from EMRs and other clinically-relevant data, we can provide better predictions to improve efficiency and quality of care.

2 Background and Literature Review

Siontis et. al (2011) provide an empirical review of methodologies around mortality models. In this paper, the authors use Medline to identify studies published in 2009 that assessed AUC of predicting all cause mortality. What they find is that most studies at the time predicted mortality with modest accuracy, with large variability across populations, which makes sense since for since very high risk populations it may be easier to predict mortality, but for lower risk general populations, the class imbalance is very high with relatively few patients actually dying. A full list of the studies that the authors review and the approaches they use are documented in the paper (linked above). However, we also copy a portion of that table below as an illustrative example:

3 Methodology

What we will do in this project is to try to reproduce and improve models motivated by Unfolding Physiological State: Mortality Model in Intensive Care Units (Ghassemi et. al). In particular, what we find interesting is the use of free-text hospital notes to better predict mortality. This is because in many cases, there are already many mortality models, built both on Business Rules, and machine learning to predict mortality. Perhaps the biggest methodology to improve on existing mortality models is to use unstructured data, to see if there is any additional signal from notes and other text that we don't already get from structured data using EHRs.

   To be more specific, we will use the same dataset as indicated in the Project guidelines, MIMIC III. This data set has 45,000 patients, 60,000 Hospital Admissions, 200,000 Clinical notes, and 6000 mortality events. We will then divide into 80% training, 20% test set. Similar to what is presented in the paper, we will first extract clinical baseline features, including age, sex, and SAPS-II score, from the database for every patient. We will then perform topic modeling, perhaps LDA, to do natural language processing on the notes to extract meaningful features. In terms of evaluating our model, we will predict on multiple iterations of the outcome variable, just as the paper does. For example, to evaluate 24 hour

mortality predictions, in the training set, we will create a 24 hour prediction window and using the data that is transformed into features, we will measure AUC of mortality where outcome variable is mortality within 24 hours. We will then test on held-out test data set to report final test AUC.

4 Experimental Results

FILLER

5 Discussion

6 Conclusion

Filler

References

[1] Angrist, J. D., Cohodes, S. C., Dynarski, S. M., Pathak, P. A., and Walters, C.R. (2013). Stand and Deliver: Effects of Boston's Charter High Schools on College Preparation, Entry, and Choice. *NBER Working Paper,* No. 19275.