Mortality Prediction in ICU

Shaikh Shamid

***Abstract*—Given patient's medical record in ICU, we predict the mortality based on structured features. This prediction can greatly improve the efficiency and quality of care by having the early knowledge of a patient's condition. In addition to the structure features, Ghassemi et al. [1] used the latent variable models (viz. Latent Dirichlet Allocation) to turn the clinical notes into meaningful features, and showed the effectiveness of predictive power of these features for patient mortality. This work considered only one prediction categories: baseline prediction based on structured data (such as age, sex, saps score). The goal of this project is to repeat the first part of the study[1] using big data tools (e.g., Hadoop and Spark).**

***Index Terms*—Big data, Health analytics, Data mining, Machine learning, Healthcare and medicine; Topic modeling, LDA, Support vector machines.**

I. Introduction

In ICU, the accurate knowledge of patient's condition is very important to figure out which patient needs great attention because ICUs are busy. An increased amount of electronic medical data available has made it easier for researchers to analyze it and make mortality predictions, as evidenced by the growing number of articles related to this topic. Our motivation in this project is to do the mortality prediction to prioritize the task, and hence improve the effectiveness and quality of care.

In 2009, 118 validated mortality prediction tools were published [2] with modest accuracy (such as APACHE [3], SAPS-II [4], and SOFA [5] with median reported AUCs of 0.77, 0.77, and 0.84, respectively), and large variability of these measures across various diseases and population subgroups. These models are also based on numeric, waveform, or admission baseline data (not actionable). Several research works have also used clinical notes in addition to structured data to formulate the model. Saria et al. [6] combined structured data with concepts from the discharge summaries to achieve a better patient outcome classification F1 score. Similarly, [7] described clinical texts in a subgroup of ICU patients were predictive of mortality, and an RBF SVM achieved a retrospective AUC of 0.855 for in-hospital mortality. Finally, Lehman et al. [8] applied Hierarchical Dirichlet Processes to nursing notes from the first 24 hours for ICU patient and demonstrated that unstructured nursing notes were enriched with clinically meaningful information. This work represents part of the research done by Ghassemi et al. [1] in the following ways:

- No patients were excluded by age (i.e. newborns) or by low volumes of incoming data.
- All code and data was built using Scala and Apache Spark for scalability.
- No time-varying nature of ICU notes event
- studied only in-hospital mortality

II. Problem formulation

The problem we address in this work is to predict patients in-hospital mortality probabilities bases on structured data available in electronic medical records, including age, sex, SAPS score, minimum SAPS score, maximum SAPS score, SOFA score, minimum SOFA score and maximum SOFA score for every patient from icustay_detail table of MIMIC 2v26 database [9]. We attempt to reproduce their Retrospective Models for patient mortality. In order to measure the success of our model, we will be using area under ROC and area under Precision Recall (PR).

III. Approach and implementation

We constructed an analytics architecture that utilizes distributed computing environment in order to process this large quantity of data in an efficient manner. We wrote our application using Scala and the Scala Spark 1.3.1 API. Apache Spark is an open-source distributed processing engine that uses the speed of in-memory computing in order to process big data very quickly. MLlib is Spark's machine learning library, that implements many of the widely-used machine learning algorithms and other utilities in a distributed fashion. Much of our data processing and analysis used MLlib's implementations. We work with a data set from MIMIC 2v26 database [9] constructed from patients admitted to ICUs at Boston's Beth Israel Deaconess Medical Center during the period 2001 to 2007 consisting record of 32,425 patients. We extract age, gender, SAPS score,minimum SAPS score, maximum SAPS score, SOFA score,minimum SOFA score, maximum SOFA score and mortality outcome from the admissions records. Patients were filtered out if they did not have recorded age, initial SAPS I score, minimum SAPS I score, maximum SAPS I score, or gender. We identify hospital mortality outcome based on the hospital_expire_flg. A feature vector is constructed using the patients age, gender, SAPS score,minimum SAPS score, maximum SAPS score, SOFA score,minimum SOFA score and maximum SOFA score to build models for predicting in-hospital mortality. Patients were divided into a training/test split of 70% of patients assigned to the training set, and 30% of patients assigned to the testing set. Thus our cohort composition consists of 70% patients in train set and 30% in test set.

IV. Experiment Design and evaluation

We focused on predicting only one outcome (in-hospital mortality) for this experiment. This admission baseline and derived features model uses the structured features of age, gender, SAPS and the SOFA scores extracted from the data (8 features total). Because there was a large class imbalance for this experiment, we used subsampling on the negative class

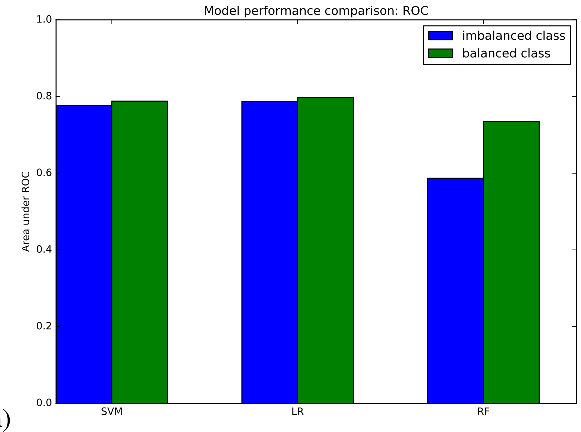
Model	AU-PR	AU-ROC
SVM	0.826	0.788
LR	0.828	0.797
RF	0.821	0.735

TABLE I
MODEL PERFORMANCES

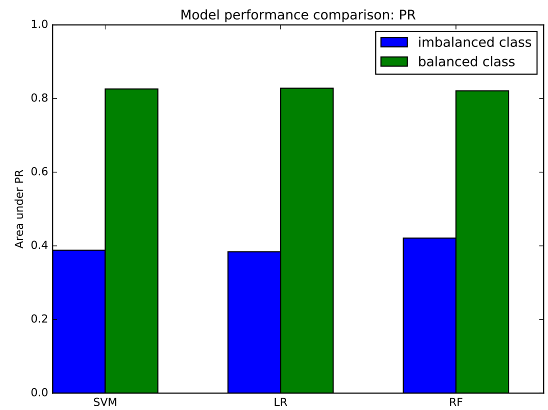
in order to create at least a 70%/30% split of negative and positive class data.

One linear model (Logistic Regression) and two non linear models (SVM and Random Forest) were used to assess the features inherent relationship with the target mortality. For non linear models, we also tested non-parametric, tree based ensemble Random forest (RF) model to figure out the underlying relationship between features and targets. We compared the model performances to find out which type of algorithms find the classification boundaries better for this data set. The model is evaluated by the Area Under the Receiver-Operator Curve (AUC) and Area Under the Precision and Recall (PR), which can be interpreted as the quality of the rank ordering of patients likelihood of having a mortality event. A value of 1.0 would be a perfect rank-order of severity, while a value of 0.5 would be a random rank-ordering.

Figure 1 shows the model performances for both balanced and imbalanced class. In the case of area under PR, indeed the model performances are greatly improved by subsampling on the negative class in order to make the class balanced. The final AU-PR for all three models are consistent with values 0.82, but the final AU-ROC metrics are ranged between 0.73 and 0.80 as shown in table I. And the final result shows that the linear model is sufficient to find a descent classification boundary for this type of static mortality prediction problem where we used only admission baseline and derived features.



(a)



(b)

Fig. 1. Model performance comparison for both balanced and imbalanced class (a) Area under ROC, (b)Area under PR.

V. CONCLUSION

This study was able to reproduce similar results to the part of the original study [1] by studying multiple classification algorithms in big data platforms. As big data technologies make it possible to build faster and more complex tools that are scalable, which can result in enhanced quality of care, particularly in the ICU by doing precise and efficient analysis of patient's data.

This work can be improved by extracting each patient's clinical notes from noteevent table of the same database and use LDA model to construct time-varying features with a fixed window size.Then it would be interesting to show that the combined features achieving higher AUC values than the baseline and derived features.

For the technology stack, we use Spark version 1.3.1 with its Scala API. In particular, Spark MLlib containing common machine learning algorithms. We use Spark ML which is a new high-level API for Spark MLlib: it defines a pipeline containing standard component like Cross-validation or metrics.

ACKNOWLEDGMENT

I would like to thank our Professor, and all the TAs for their time and effort in helping to make this work a success.

VI. SUPPLEMENT MATERIAL

A. Code link

https://github.gatech.edu/sshamid3/project-submission

REFERENCES

[1] M. Ghassemi, T. Naumann, F. Doshi-Velez, N. Brimmer, R. Joshi, A. Rumshisky, and P. Szolovits. Unfolding Physiological State: Mortality Modelling in Intensive Care Units. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, **KDD 14**, pages 75-84, New York, NY, USA, 2014. ACM.

[2] G. Siontis, I. Tzoulaki, and J. Ioannidis. Predicting death: an empirical evaluation of predictive tools for mortality. **Archives of internal medicine**, pages archinternmed-2011, 2011.

[3] W. A. Knaus, D. Wagner, E. e. a. Draper, J. Zimmerman, M. Bergner, P. G. Bastos, C. Sirio, D. Murphy, T. Lotring, and A. Damiano. The apache iii prognostic system. risk prediction of hospital mortality for critically ill hospitalized adults. **CHEST Journal**, 100(6):1619?1636, 1991.

[4] J. Le Gall, S. Lemeshow, and F. Saulnier. A new simplified acute physiology score (saps ii) based on a european/north american multicenter study. **JAMA**, 270(24):2957?2963, 1993.

[5] J.-L. Vincent, R. Moreno, J. Takala, S. Willatts, A. De Mendon?ca, H. Bruining, C. Reinhart, P. Suter, and L. Thijs. The sofa (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. **Intensive care medicine**, 22(7):707?710, 1996.

[6] S. Saria, G. McElvain, A. K. Rajani, A. A. Penn, and D. L. Koller. Combining structured and free-text data for automatic coding of patient outcomes. In **AMIA Annual Symposium Proceedings**, volume 2010, page 712. American Medical Informatics Association, 2010.

[7] M. Ghassemi, T. Naumann, R. Joshi, and A. Rumshisky. Topic models for mortality modeling in intensive care units. In Proceedings of **ICML 2012** (Machine Learning for Clinical Data Analysis Workshop), Poster Presentation, Edinburgh, UK, June 2012.

[8] L.-w. Lehman, M. Saeed, W. Long, J. Lee, and R. Mark. Risk stratification of icu patients using topic models inferred from unstructured progress notes. In **AMIA Annual Symposium Proceedings**, volume 2012, page 505. American Medical Informatics Association, 2012.

[9] M. Saeed et al. Multiparameter Intelligent Monitoring in Intensive Care II: A public-access intensive care unit database. **Critical Care Medicine**, 39(5):952?960, May 2011.