CSE8803 Project: Mortality Prediction in ICU patients

Pradeep Vairamani, Masters student in Computer Science, Georgia Institute of Technology

***Abstract*—Accurate prognosis and prediction of a patient's current disease state is critical in an ICU. The use of vast amounts of digital medical information can help in predicting the best course of action for the diagnosis, prognosis and treatment of patients. The proposed technique investigates the strength of using a combination of latent variable models (latent dirichlet allocation) and structured data to transform the information streams into potentially actionable knowledge.**

***Index Terms*—Big data, Health analytics, Data mining, Machine learning, LDA**

I. Introduction and motivation

Healthcare delivery is in the midst of an acute transformation with the adoption and widespread use of health information technology. The increased adoption of electronic health records (EHR) has triggered a number of data analytics, improved medical care management and other innovations. These promise to improve healthcare delivery, efficiency, quality and safety. Big data technologies like Apache spark enable us to capture, store and perform analysis of a large and extended volume of structured and unstructured data. This information can be used to help physicians in improving and optimizing medical care. ICU acuity metrics are routinely utilized to quantitatively characterize the severity of illness of ICU patient populations, and are applied for mortality prediction and benchmarking ICU performance. [3] [7]. However, most of the work has aimed to consolidate structured data [5] [2] and omit free-text clinical notes and reports. Saria et al. [8] showed that integrating structured information with current natural language processing based systems can significantly reduce prediction errors. A similar study by Lehman et al. [6] used Hierarchical Dirichlet Processes for ICU patient risk stratification by combining the learned topic structure of clinical concepts extracted from the unstructured nursing notes with physiologic data for hospital mortality prediction. This gives us an improved prediction of the outcome.

II. Project Life Cycle

The work by Ghassemi et al [4]. highlighted and established the advantages of integrating free-text clinical notes with structured data. The aim of this project is to repeat, validate, analyze and build on this work. This is done by combining the standard physiological results (structured data) and features extracted from free-text data.

A. Data gathering and pre-processing

This project uses the freely available MIMIC 3 [9] database which includes de-identified health data for diverse set of patients. The baseline features like age, gender, SAPS II (Simplified Acute Physiology Score), OASIS (Oxford Acute Severity of Illness Score), APS III (Acute Physiology Score III) scores and mortality outcomes are extracted/constructed from the MIMIC 3 database. The patient mortality outcomes serve as the ground truth for the machine learning models that are applied on this consolidated data.

In addition to the baseline demographic and severity scores, the free-text notes are extracted from MIMIC III. The free-text notes are then cleaned and processed using Apache Spark to handle the extraneous newline/space/non-alphanumeric characters before storing them along with the baseline features. Vocabularies were then generated by tokenizing the notes and the tokens of length less than 3 were discarded. A term count model was then constructed with the remaining tokens and the 100 most common words were discarded. In addition to this, another model was created to explicitly remove the stop words using the Onix stopwords list. The vocabulary size after these operations reduces to 83177.

B. Cohort composition and feature construction

The MIMIC III ICU dataset consists of 46,520 patients (26,121 male and 20,399 female patients) and 2,078,705 clinical notes. Since the predictive value of mortality is most useful early in the treatment of the patient, we consider only the demographic and severity scores at the time of admission. The clinical notes are restricted to the notes that are taken during the first 12 hours from the admit time. We can now construct the features for the given cohort. There are 3 kinds of features that are used in this project.

1) Extracted features: These are features like age and gender which are directly extracted from the MIMIC 3 database.
2) Constructed features: Severity scores like SAPS II, OASIS and ASP III are constructed using the various tables of the MIMIC III data using sql queries.
3) Features derived from notes: 50 topics are generated using Latent Dirichlet allocation (LDA) which posits that each document is a mixture of a small number of topics and that each word's creation is attributable to one of the document's topics. LDA infers a distribution over topics for each document and this can be used as features for our machine learning algorithms.

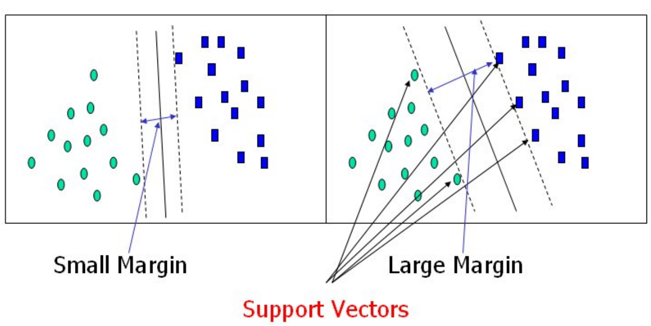
Topic models are a suite of algorithms that uncover the hidden structure in document collections. By discovering patterns of word usage from documents that exhibit similar patterns, words are grouped into thematically coherent structure, called topics. LDA provides a generative model that describes how the documents in a dataset can created . In our scenario, the documents are the clinical notes. LDA looks at each document as a collection of topics and by learning the topic distribution, we can construct the feature vectors. A linear kernel SVM is then trained to create classification boundaries. It is worth noting that the topic learning was done in a completely unsupervised manner i.e. no prior medical knowledge was used. The summary of the top words for each topic is shown in the Appendix.

C. Modeling pipeline

This paper considers two prediction regimes: baseline prediction and combined (structured + unstructured) prediction. The SVM is trained for each of these models and is evaluated against the test set. For the combined model, the structured and unstructured features are joined based on the admission id and averaged per patient. The combined features are randomly split into 2 parts: 70% of the features were used as a training set and the other 30% was eventually used to test the machine learning models. The training set was used to train the support vector machine (SVM) using Stochastic Gradient Descent. SGD incrementally minimizes the primal SVM objective.

$$E(\mathbf{w}) = \frac{\lambda}{2}\|\mathbf{w}\|^2 + \frac{1}{n}\sum_{i=1}^{n} \ell_i(\langle \mathbf{w}, \mathbf{x} \rangle)$$

SVM model represents each point in the feature vector as a point in space such that the vectors of separate categories can be clearly divided by a hyperplane. L2 regularization was used to avoid over-fitting. The figure below [1] shows how SVM performs classification by constructing an N-dimensional hyperplane that optimally separates the data into two categories.



D. Technical approaches

SQL was used for extracting data from the MIMIC database. The extracted data was cleaned using Apache Spark which is a open source distributed computing framework. Spark was also the ideal choice for the analytics infrastructure due to the iterative nature of the stochastic gradient descent method which was used to optimize the LDA topic model. Since the data is stored reliably in-memory, subsequent iterations share data through memory. This provides a huge performance advantage when compared frameworks like MapReduce which are more suitable for batch processing. The trained model was then stored in SVMLite format and python scripts were used to compare the performance of various machine learning algorithms on the given feature vectors.

III. Results

Area under the receiver operating characteristic curve (AUROC) is the primary metric used to evaluate the model described in the previous section. AUROC is equal to the probability that a binary classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one.

Table A. AUROC for various models

Model description	AUROC
Baseline with only age, sex and saps2	0.7256
Baseline with age, sex, saps2 and oasis	0.7296
Baseline with age, sex, saps2 and apsiii	0.7462
All baseline features	0.7432
Baseline + LDA	0.7620

The above table shows that the models that include the latent topic features were better at predicting mortality compared to the models that used only the structured features. In addition to this, we can infer that APS III and SAPS II are better features to predict mortality compared to OASIS.

Table B. Comparision of Machine learning algorithms

ML algorithm	Accuracy	Precision	Recall	F1-score
Logistic Reg	0.6805	0.6790	0.5181	0.5877
SVM	0.6001	0.7907	0.1226	0.2124
Decision Tree	0.6928	0.6522	0.6448	0.6485

The above table compares the performance of various machine learning algorithms on the SVMLight file generated using baseline features. It shows that SVM is a good choice if precision is the most important metric.

IV. Discussions

This report confirms the idea put forward by Ghassemi et al that there is rich information in the unstructured clinical notes which can be leveraged to make predictions about the patient's condition. The AUROC in this paper is roughly 0.75. The reason for the lower AUROC in comparison with the aforementioned paper is that this paper tries to make predictions purely based on the information gathered in the first 12 hours. One of the limitations of this paper is that the input to the topic model is not limited to the 500 most informative words per document as per TFIDF. Using this criterion might further increase the predictive power of the model. Another limitation of this paper is that the SOFA (Sequential Organ Failure Assessment) scores are not being used as a feature. Also, the algorithms were evaluated only on the MIMIC-III database which contains data collected from one academic medical center. Ideally, the performance of the model should be evaluated with an ICU database representative of a diverse patient population from different medical centers. A potential extension of this project will be to convert this into a real-time mortality prediction system where the streams of input data could iteratively improve the performance of the predictive model.

V. Conclusion

Predicting the severity of the patients' condition using big data techniques can help healthcare professionals provide adequate care for patients predicted to be at high risk. This can also help in prioritizing ,managing resources and costs. This paper validates the idea that features generated by Latent Dirichlet Allocation models are useful to augment the features constructed using structured data to improve the mortality prediction among ICU patients. In this paper, LDA was used to automatically discover latent structure embedded in the clinical notes. The results of this work could help healthcare providers gain valuable insights while predicting the patients disease state in the ICU.

Acknowledgment

Thanks to Professor Sun for the challenging and the very interesting course. Thanks to all the teaching assistants for the dedication and feedback.

References

[1] https://www.dtreg.com/solution/view/20.
[2] C. J. Anaesth. Limited ability of sofa and mod scores to discriminate outcome: a prospective evaluation in 1,436 patients.
[3] P. J. Beck DH, Smith GB. External validation of the saps ii, apache ii and apache iii prognostic models in south england: a multicentre study.
[4] 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. In *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.
[5] B. Khwannimit. A comparison of three organ dysfunction scores: Mods, sofa and lod for predicting icu mortality in critically ill patients.
[6] W. L. J. L. Lehman, M. Saeed. Risk stratification of icu patients using topic models inferred from unstructured progress notes.
[7] L. Respir. Mortality prediction in intensive care units with the super icu learner algorithm (sicula): a population-based study.
[8] A. K. R. A. A. P. S. Saria, G. McElvain. Combining structured and free-text data for automatic coding of patient outcomes.
[9] M. Saeed. Multiparameter intelligent monitoring in intensive care ii: A public-access intensive care unit database.

APPENDIX

Top words per topic	
TOPIC 1:	pleural, pulmonary, effusion, bilateral, heart, small, effusions, mild, moderate, evidence
TOPIC 2:	clear, foley, yellow, draining, good, monitor, pulses, lungs, urine, patent
TOPIC 3:	bleeding, units, bleed, unit, prbc, active, transfuse, lower, upper, after
TOPIC 4:	cabg, artery, post, coronary, mediastinal, aortic, cardiac, bypass, valve, pulmonary
TOPIC 5:	daily, home, history, hypertensive, denies, hypertension, lisinopril, dose, admitted, hold
TOPIC 6:	abdomen, small, bowel, free, abdominal, pelvis, fluid, contrast, within, large
TOPIC 7:	sided, neuro, weakness, head, transferred, acute, showed, stroke, have, facial
TOPIC 8:	fall, spine, cervical, spinal, cord, neck, fracture, posterior, lumbar, vertebral
TOPIC 9:	skin, wound, foot, care, area, multiple, impaired, ulcer, applied, noted
TOPIC 10:	review, aspiration, noted, admission, heparin, family, except, likely, resident, care
TOPIC 11:	that, family, wife, when, about, states, called, they, home, first
TOPIC 12:	neuro, head, seizure, pupils, dilantin, repeat, noted, equal, activity, found
TOPIC 13:	head, hemorrhage, contrast, frontal, mass, subdural, intracranial, midline, brain, report
TOPIC 14:	fever, cultures, vancomycin, line, infection, culture, recent, sepsis, picc, daily
TOPIC 15:	fentanyl, line, intubated, lumen, failure, unable, versed, intubation, start, endotracheal
TOPIC 16:	contrast, within, images, evidence, clip, were, final, number, report, admitting
TOPIC 17:	artery, identifier, carotid, were, aneurysm, internal, catheter, into, common, service
TOPIC 18:	lung, upper, lower, lobe, post, pneumothorax, status, seen, single, portable
TOPIC 19:	lower, pericardial, pulmonary, heparin, bilateral, large, extremity, femoral, normal, vein
TOPIC 20:	mental, status, altered, unable, head, more, confused, when, lethargic, found
TOPIC 21:	order, assessment, insulin, sodium, sliding, code, scale, sicu, hours, hour
TOPIC 22:	likely, urine, history, elevated, lactate, acute, pending, consider, also, setting
TOPIC 23:	increased, slightly, please, decreased, than, increase, slight, noted, both, edema
TOPIC 24:	admission, female, very, note, daughter, admitted, over, nursing, room, woman
TOPIC 25:	lung, breathing, ventilation, airway, assessment, intubated, clear, cuff, invasive, sputum
TOPIC 26:	line, placement, catheter, central, reason, number, report, final, admitting, underlying
TOPIC 27:	trach, stent, airway, bronch, tracheostomy, tracheal, neck, upper, after, secretions
TOPIC 28:	lasix, acute, chronic, failure, heart, renal, likely, setting, edema, systolic
TOPIC 29:	cath, groin, site, post, pacer, cardiac, pulses, sheath, iabp, stent
TOPIC 30:	cardiac, found, arrest, have, sinus, episode, after, were, then, bradycardia
TOPIC 31:	denies, abdominal, history, past, reports, signs, code, acute, assessment, well
TOPIC 32:	sats, micu, placed, resp, nursing, note, sent, admit, cont, ward
TOPIC 33:	hypotension, fluid, received, bolus, urine, hypotensive, after, boluses, monitor, output
TOPIC 34:	liver, hepatic, ercp, biliary, portal, gallbladder, normal, vein, pancreatic, abdominal
TOPIC 35:	liver, cirrhosis, esophageal, lactulose, coffee, portal, varices, ground, octreotide, hepatic
TOPIC 36:	intubated, vent, propofol, wean, sedated, resp, thick, care, weaned, suctioned
TOPIC 37:	fracture, trauma, fractures, multiple, report, final, number, lateral, clip, year
TOPIC 38:	been, that, also, have, some, which, does, more, would, several
TOPIC 39:	cancer, mass, metastatic, cell, breast, lung, tumor, biopsy, chemo, diagnosed
TOPIC 40:	renal, insulin, dialysis, esrd, type, glucose, kidney, transplant, acute, chronic
TOPIC 41:	started, after, down, arrival, upon, arrived, received, drip, initially, back
TOPIC 42:	surgical, drainage, drain, post, surgery, repair, dilaudid, small, pacu, incision
TOPIC 43:	reason, underlying, report, admitting, number, eval, final, year, interval, clip
TOPIC 44:	cardiac, ventricular, aortic, valve, mitral, history, normal, systolic, wall, daily
TOPIC 45:	worsening, transferred, high, lung, hypoxia, sats, cough, showed, pulmonary, pneumonia
TOPIC 46:	atrial, afib, coumadin, rate, daily, fibrillation, hold, metoprolol, diltiazem, history
TOPIC 47:	daily, copd, home, tablet, bipap, chronic, albuterol, nebs, history, prednisone
TOPIC 48:	code, radial, assessment, gauge, monitoring, hemodynamic, signs, stress, vital, glycemic
TOPIC 49:	denies, able, oriented, alert, monitor, when, meds, times, nausea, easily
TOPIC 50:	etoh, alcohol, abuse, ciwa, withdrawal, history, valium, ativan, psych, admitted