Unstructured Clinical Notes Mining for Mortality Prediction

Vineet Kumar (19BM6JP46) Rohit Bajpai (19BM6JP56)

Presented at Final End Term Project Evaluation
for
(BM60130) Healthcare Analytics
IIT Kharagpur

June 15, 2020

Problem Statement & Motivation

We want to use clinical notes to assess mortality of a patient

We want to use clinical notes to assess mortality of a patient

patient transferred from hospital ward name 289 11 with change in mental status and changes on ct scan problem pneumocephalus assessment patient lethargic but has increased wakefulness as shift progresses follows commands pupils pearla orientated to self only heart rate sinus rhythm with lots ofectopy systolic b p 90 130 ns over 50 ns foley patent draining clear yellow urine lumbar drain clamped site intact n full strength all extremities action echocardiogram done by fellow npo problem pneumocephalus n assessment early placed on open face mask for humidification wfi02 100 and n 15! abg wnl weak cough rhonchi in upper lobes minimal secretions dry mucosa n continued with purposeful mvt but nonverbal no obeying of commands this afternoon pt presenting with a worsening respiratory picture hr n 120 ns 130 rr 30 40 ns labored with abdominal breathing agitated

Deidentified information   Patient Problems (as assessed)   Patient Measurements
Patient Condition (as observed)

Figure: Sample Color Annotated Clinical Note

Problem Statement & Motivation

We want to use clinical notes to assess mortality of a patient

patient transferred from hospital ward name 289 11 with change in mental status and changes on ct scan problem pneumocephalus assessment patient lethargic but has increased wakefulness as shift progresses follows commands pupils pearla orientated to self only heart rate sinus rhythm with lots ofectopy systolic b p 90 130 ns over 50 ns foley patent draining clear yellow urine lumbar drain clamped site intact n full strength all extremities action echocardiogram done by fellow npo problem pneumocephalus n assessment early placed on open face mask for humidification wfi02 100 and n 15! abg wnl weak cough rhonchi in upper lobes minimal secretions dry mucosa n continued with purposeful mvt but nonverbal no obeying of commands this afternoon pt presenting with a worsening respiratory picture hr n 120 ns 130 rr 30 40 ns labored with abdominal breathing agitated

Deidentified information Patient Problems (as assessed) Patient Measurements Patient Condition (as observed)

Figure: Sample Color Annotated Clinical Note

Task

We want to do 1 week, 1 month, 6 month & 1 year long mortality prediction

Dataset[1]

MIMIC–III v1.4 dataset

- Publicly available de-identified health-related data associated with over 40,000 patient
- Recorded in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012.
- Dataset contains information on patient demographics, vital sign measurements, lab test results, clinical procedures, medications, caregiver notes, imaging reports, and mortality labels.

Dataset[1]

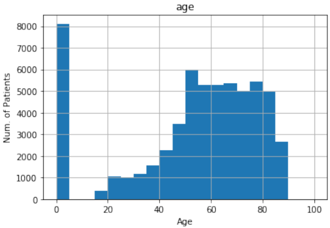
MIMIC–III v1.4 dataset

- Publicly available de-identified health-related data associated with over 40,000 patient
- Recorded in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012.
- Dataset contains information on patient demographics, vital sign measurements, lab test results, clinical procedures, medications, caregiver notes, imaging reports, and mortality labels.

Dataset is in form of multiple tables. For our purpose we only selected

- Admissions Table – ICU Stay and other info
- Patients Table – Patient level and demographic information
- Notes Table – clinical free form textual description

Patient Cohort Selection



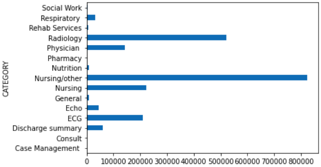
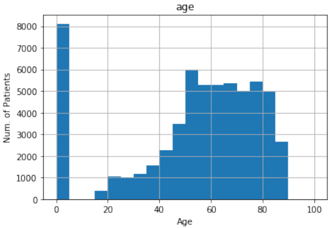


Figure: Age and Frequency Distribution of Type of Notes

Patient Cohort Selection



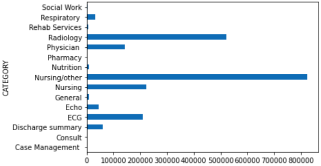


Figure: Age and Frequency Distribution of Type of Notes

Inclusion Exclusion Criteria

We retrieved set of

- Adult patients ($\geq$ 18 years) with at least one associated note from the database.
- We extracted `Nursing Progress Note` from the dataset.

Dataset Preprocessing

- We used **Big query** and its SQL dialect for our cohort selection.
- Mean number of notes per hospital admission is 32 (Median: 15)
- We concatenated all notes of a particular patient with the same hospital admission id.

- We used **Big query** and its SQL dialect for our cohort selection.
- Mean number of notes per hospital admission is 32 (Median: 15)
- We concatenated all notes of a particular patient with the same hospital admission id.

Survival Label

- 'surv_day', which is essentially the number of days the patient has survived after discharge from the hospital

Attribute Label	Frequency (Prevalence)
one_week_mortality	258
thirty_day_mortality	736
one_eighty_mortality	253
one_year_mortality	653

Table: Prevalence in Survival Class

Workflow

- Patients were randomly split into train (70%), validation (15%), and test (15%) sets.
- Since we have a limited number of True labels for any class, (thirty_day_mortality has train prevalence as 11%). It has a class imbalance, to deal we this we did a minority class oversampling.
- To keep things simple and interpretable, we tokenized our notes and then used the `CountVectorizer` on it.

Workflow

- Patients were randomly split into train (70%), validation (15%), and test (15%) sets.
- Since we have a limited number of True labels for any class, (thirty_day_mortality has train prevalence as 11%). It has a class imbalance, to deal we this we did a minority class oversampling.
- To keep things simple and interpretable, we tokenized our notes and then used the `CountVectorizer` on it.

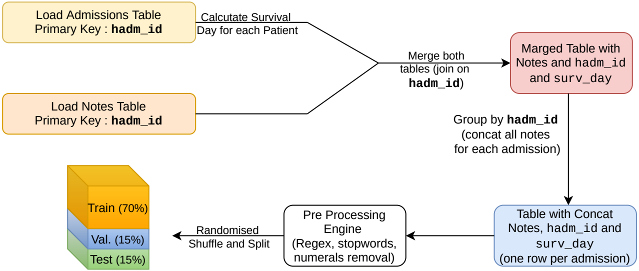
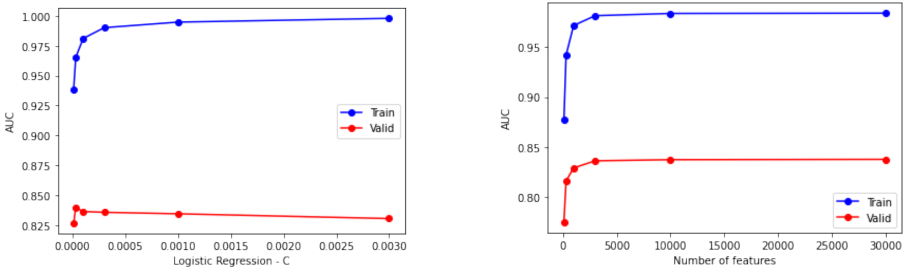


Figure: Workflow of data preparation

Model Description

- We used bag of words approach with logistic regression for initial model building → we can plot the most important features for each **True** and **False** survival class.
- **True** = [`ml, should, time, it, been, cm, cc, dr`]
- **False** = [`po, off, c, i, level, to, s, at`]
- Added new stop words list → removed them from our vocabulary

- We used bag of words approach with logistic regression for initial model building → we can plot the most important features for each **True** and **False** survival class.
- **True** = [`ml, should, time, it, been, cm, cc, dr`]
- **False** = [`po, off, c, i, level, to, s, at`]
- Added new stop words list → removed them from our vocabulary



(a) C Value

(b) Number of Features

Figure: Choice of Hyperparamter for logistic regression

Evaulation Metrics

- We have used Area under the receiver operator curve (**AUC**) to judge the performance of the model. This metric is consistent with other studies published in the domain.
- Apart from AUC, we reported **Accuracy** as well. Accuracy is defined as $\frac{\sum \mathrm{TP} + \sum \mathrm{TN}}{\sum \text{Total Population}}$.
- We also reported **Specificity**, which is defined as

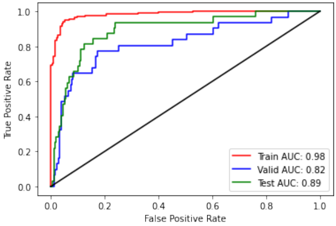
$$\frac{\text{number of true negatives}}{\text{number of true negatives} + \text{number of false positives}}$$

Results

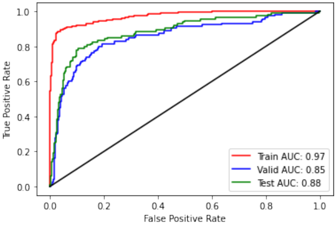
Model →	Logistic Regression			Decision Tree		
Metric →	AUC	Accuracy	Specificity	AUC	Accuracy	Specificity
7 day Mortality	0.891	0.766	0.762	0.743	0.765	0.767
30 day Mortality	0.884	0.855	0.862	0.809	0.810	0.811
6 month Mortality	0.641	0.599	0.597	0.554	0.662	0.671
1 year Mortality	**0.607**	0.569	0.567	0.550	0.542	0.540
Model →	XG Boost			SVM		
Metric →	AUC	Accuracy	Specificity	AUC	Accuracy	Specificity
7 day Mortality	**0.899**	0.834	0.834	0.739	0.679	0.68
30 day Mortality	**0.926**	0.906	0.915	0.665	0.705	0.744
6 month Mortality	**0.678**	0.640	0.640	0.569	0.600	0.609
1 year Mortality	0.604	0.597	0.602	0.500	0.515	0.536

Table: Evaluation Metric for Various Models

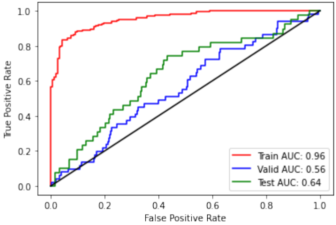
AUC Curve



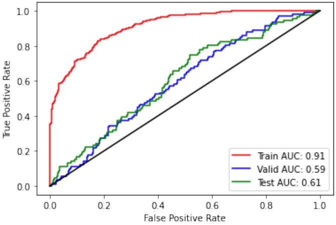
(a) 7 day mortality



(b) 30 day mortality



(c) 180 day mortality



(d) 365 day mortality

Key Inferences

- Comparison with Other State of Art Mortality Predictors from Nursing Notes

Author $\rightarrow$	Rajkomar et.al	Craig et.al	Our Approach
AUC	.94	.70	0.89
Approach	DL & Fusion	CNN	BOW

- Among all the models experimented Logistic Regression and XG Boost performed better compared to others.
- The best AUC was achieved for 30 day mortality prediction and performance drops as we aim for 6 month or 1 year long mortality prediction.

Surival Analysis using Sentiment of Clinical Notes

- No labels $\rightarrow$ Unsupervised syntactic Sentiment Mining

- No labels $\rightarrow$ Unsupervised syntactic Sentiment Mining
- We used Textblob NLP Library. Using `v0.16.0` we extracted two syntactic features from the notes – polarity and subjectivity
- Textblob algorithm tokenizes the text, performs POS tagging, and uses a lexicon to give a sentiment polarity and subjectivity score.

- No labels → Unsupervised syntactic Sentiment Mining
- We used Textblob NLP Library. Using `v0.16.0` we extracted two syntactic features from the notes – polarity and subjectivity
- Textblob algorithm tokenizes the text, performs POS tagging, and uses a lexicon to give a sentiment polarity and subjectivity score.

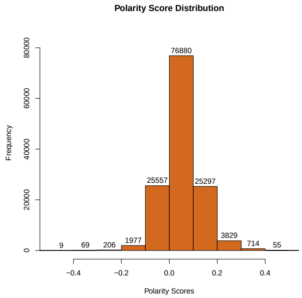


Figure: Polarity Distribution among notes

Survival Analysis

We used polarity, subjectivity with SAPSII[2] score to do the survival analysis of the patients. We partitioned the polarity and subjectivity into 4 quartiles and plotted the Kaplan Meier curve.

We used polarity, subjectivity with SAPSII[2] score to do the survival analysis of the patients. We partitioned the polarity and subjectivity into 4 quartiles and plotted the Kaplan Meier curve.

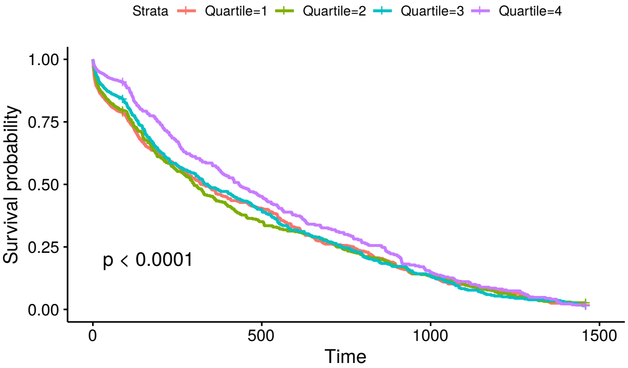


Figure: Survival Curves using Only Subjectivity and SAPSII Score

KM Curve

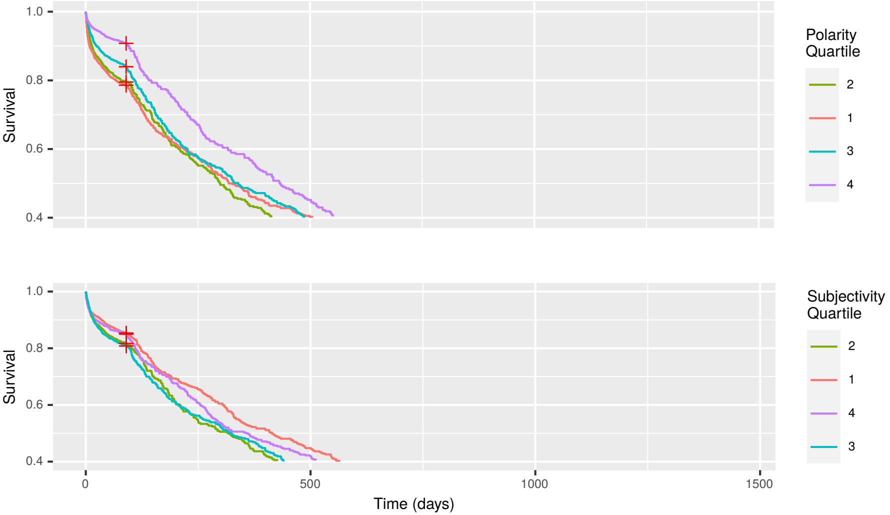


Figure: Kaplan Meier Curve

Concluding Remarks

- Unstructured text in the EHR notes contains meaningful information that can be mined for building intelligent medical decision support systems.
- If we could use the data from multiple sources, we can genuinely validate our claim.

Future Directions

- Incorporate other **structural data** (which is mostly time series such as patient measurements, etc.), **sentiments** in our model; and build a simple multimodal fusion model that can leverage both structural as well as unstructured free form text

Future Directions

- Incorporate other **structural data** (which is mostly time series such as patient measurements, etc.), **sentiments** in our model; and build a simple multimodal fusion model that can leverage both structural as well as unstructured free form text

- Prediction of mortality remains a complex problem, however by incorporating more information extracted from unstructured data, like nursing notes it might improve the overall performance.
- Code is available in github
  `https://github.com/vntkumar8/clinical-notes-mining`

References

- Alistair EW Johnson et al. Mimic-iii, a freely accessible critical care database. Scientific data, 3:160035, 2016.
- Alvin Rajkomar et al. Scalable and accurate deep learning with electronic health records. NPJ Digital Medicine, 1(1):18, 2018.
- Ian E. R. Waudby-Smith, et al. Sentiment in nursing notes as an indicator of out-of-hospital mortality in intensive care patients. PLOS ONE, 13(6):1 – 11, 06 2018. doi: 10.1371/journal.pone.0198687. URL https://doi.org/10.1371/journal.pone.0198687.
- Bonggun Shin et al. Multimodal ensemble approach to incorporate various types of clinical notes for predicting readmission. In 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), pages 1 – 4. IEEE, 2019.

Thankyou!