Sentiment Mining of Nursing Notes for Mortality Prediction

Vineet Kumar (19BM6JP46)

Rohit Bajpai (19BM6JP56)

WorkinProgress | Project for (BM60130) Healthcare Analytics

Problem Statement

Using Unstructured Clinical Data (UCD) to predict mortality of patient

Types of UCD: Nursing Progress Note, Radiology Report etc.

Problem Statement

Using Unstructured Clinical Data (UCD) to predict mortality of patient

Types of UCD: Nursing Progress Note, Radiology Report etc.

Example of Nursing progress Note

Pt is a 57 y/p male with h/o afib, LE edema, who presented to cardiologist s office [**3-25**] c/p worsening LE edema and orthopnea. TTE showed R CHF and admitted hypoxic on admission and bradycardic in 50s-60s. Aggressively diuresed since admission, on lasix gtt up to 10/hour; spironolactone; diamox. He also has pulm. Hypertension. Triggered for somnolence on 24th started on 2L of O2. On the 25th demonstrated diminished mental status and hypoxia in setting of CPAP, so it was stopped. Last night patient intolerant of mask/autoset. Overnight and this morning had increasing hypercarbia and marked somnolence to point of being obtunded. Transferred to MICU for further management. Impaired Skin Integrity

Using Unstructured Clinical Data to predict **mortality of patient**

- In this work we are only considering Nursing Notes.
- Mortality of Patient — (i) 30 day mortality (ii) >30 days mortality

Dataset

- We are using MIMIC - III dataset [a]
- Access to dataset was obtained through proper channel and due compliance of ethics
- Dataset contains completely anonymised data of 40,000 patients who were admitted to the ICUs of the Beth Israel Deaconess Medical Center between 2001 and 2012

---

[a]Johnson AE, Pollard TJ, Shen L, Lehman LwH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Scientific data. 2016; 3. https://doi.org/10.1038/sdata.2016.35

Approach

- Uncompressed size of dataset is 6GB
- Data were extracted from the MIMIC database using **Google Big Query** with the help of dialect of Structured Query Language (SQL)
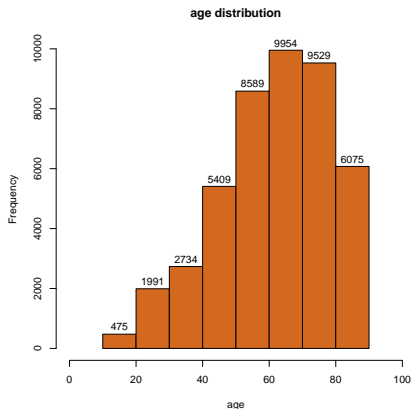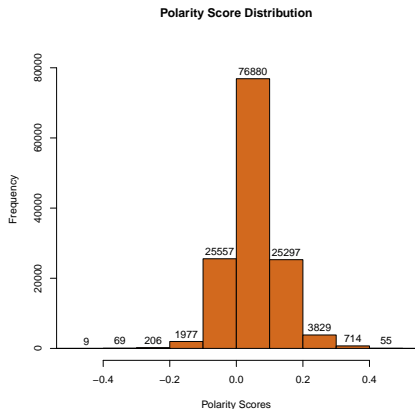- DB contains multiple tables
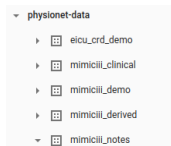
  using appropriate joins relevant patient notes and other information were extracted
- Using elementry techniques polarity scores were computed

Approach

1. Data Extraction from Database using Bigquery
2. Data Preprocessing (remove stop words etc.)
3. Sentiment Mining using LDA/Polarity and subjectivity
4. Using RNN for lower dimensional representation and unsupervised clustering
5. Try BERT for improved accuracy
6. (Possibly) survival analysis

Elementary Results

Polarity Score Distribution

age distribution

Daily codes pushed at
https://github.com/vntkumar8/HealthCareAnalytics

References

- Tran, Nam, and Joon Lee. "Using multiple sentiment dimensions of nursing notes to predict mortality in the intensive care unit." 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI). IEEE, 2018.
- McCoy, Thomas H., et al. "Sentiment measured in hospital discharge notes is associated with readmission and mortality risk: an electronic health record study." PloS one 10.8 (2015).
- Ghassemi, Mohammad M., Roger G. Mark, and Shamim Nemati. "A visualization of evolving clinical sentiment using vector representations of clinical notes." 2015 Computing in cardiology conference (CinC). IEEE, 2015.