Research Essay with Joel

Initial Ideas

* Patient similarity metrics, personalized mortality prediction, electronic medical record data
* What machine learning tools are the most useful for clinicians/clinical translation?

What I’d like to explore

* Patient similarity
* RNNs

Research questions/Problems

* Outcomes
  + Mortality
  + Risk of readmittance

To read/or just look at tbh:

Joel’s latest work on patient similarity metrics and electronic medical data

Personalized Mortality Prediction Driven by Electronic Medical Data and a Patient Similarity Metric:

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0127428>

 Personalized mortality prediction for the critically ill using a patient similarity metric and bagging

<https://ieeexplore.ieee.org/document/7455902?arnumber=7455902&tag=1>

Patient similarity for precision medicine: A systematic review

<https://www.ncbi.nlm.nih.gov/pubmed/29864490>

 Real-time mortality prediction in the Intensive Care Unit

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5977709/>

Joel’s student Thesis: Towards Precision Medicine in Intensive Care: Leveraging Electronic Health Records and Patient Similarity

<https://uwspace.uwaterloo.ca/bitstream/handle/10012/14726/Sharafoddini_Anis.pdf?sequence=4&isAllowed=y>

RNNs and electronic health records

Scalable and accurate deep learning with electronic health records:

<https://www.nature.com/articles/s41746-018-0029-1/>

General information of analytics in big data:

From Big Data to Precision Medicine

<https://www.frontiersin.org/articles/10.3389/fmed.2019.00034/full>

Big data analytics for personalized medicine:

<https://www.sciencedirect.com/science/article/pii/S0958166918301903#bib0260>

Option?

-          Collaborator at St Michaels to get a different data source

<https://www.chartdatascience.ca/the-data>

<https://www.stmichaelshospital.com/media/detail.php?source=hospital_news/2018/1207>

<https://medium.com/intel-student-ambassadors/predictions-using-the-mimic-iii-database-part-1-1aa58475da40>

First Reading: Joel’s latest work on patient similarity metrics and electronic medical data

* Prediction while training/testing? on a subset of the data – individuals which are “similar”
* “similarity” is defined by cosine similarity

Cosine similarity: <https://en.wikipedia.org/wiki/Cosine_similarity>

* + Cosine similarity =
  + Ranges from -1,1 where -1 means exact opposite, and 1 meaning exactly the same, 0 is orthogonal
  + Meaning is how similar are the orientations of the two vectors
  + Ex when measuring term frequencies for documents, cosine similarity ranges from 0 to 1 because all values are positive.
* They use patient similarity successful
* Drawback: too strong of a threshold results in small sample effects

Personalized Mortality Prediction Driven by Electronic Medical Data and a Patient Similarity Metric:

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0127428>

First meeting with Joel: Jan 8th 2020

Approaches to take to a project on Patient Similatity according to Joel

* How to define Similarity
* Should individuals be weighted differently
  + Ie more similar individuals have a higher weighting
* Supervised or unsupervised?
  + Should the similarity metric take into account the outcome
* Feature slection in the similarity metric
  + Are the features the same in the similarity metric as in the final predictive model?
  + Are the features selected biologically or medically motivated?
  + How are they picked in a general setting?

Data

* eICU data
  + less used than Mimic
  + more informative
  + see its publication, nature scientific data
  + includes data from multiple hospitals (200 hospitals)
* getting access to eICU
  + to do
  + may take 3 hours ~
  + request access at eicu-crd.mit.edu
    - here also there are tables/data definitions
* nuisance of the eICU dataset
  + for a patient which has 2 ICU stays in the same hospitals in the same year, the order of the stays are not identifiable
    - you will know if on one of the stays, the patient dies
  + convention in the lab: analyse the first unit stay

Final Model/similarity notes

* because we are testing a methodology, we can run a model/individual selector on a subset of features to make it faster

Important Predictor/feature: APACHE IV Severity of Illness score(SOI)

* may be good to include in model/similarity
* score calculated in the first 24 hours of a patient stay
* proprietary
  + we don’t know the underlying model of this score