CinC 2019 – Singapore

# Abstract of Presentation

Optimal feature selection leads to enhanced efficiency and accuracy when developing both supervised and unsupervised machine-learning models. In this work, a new Signature-based regression model is proposed to automatically identify a patient's risk of sepsis based on physiological data streams and to make a positive or negative prediction of sepsis for every time interval since admission to the intensive care unit. The gradient boosting machine algorithm that uses the features at the current time-points and the signature features extracted from the time-series to model the longitudinal effects of sepsis yields the utility function score of 0.360 (officially ranked 1st, team name: ‘Can I get your Signature?’) on the full test set. The signature method shows a systematic and competitive approach to model sepsis by learning from health data streams.

# Notes from attended talks

## Approaches

* LSTM encoder/decoder used, with the information in the encoded layer attached to a fully connected layer to make predictions. This allows optimization of both the L1 and L2 reconstruction losses.
* A missing matrix that encodes information about how long a certain variable has been missing for. A decay factor is computed that depends on the length of time since a variable was observed.
* Using a weighted loss function rather than a regression would improve the predictions.
* Have an additional feature for each variable with missingness that notes how long a feature has been missing for.

One presentation showed that they trained an LSTM and got a mediocre prediction, trained a gradient boosting algorithm and got a better one, then included the LSTM prediction as a feature into the gradient boosting model which gave them the best performance. This is something I have not tried and would be interested to see the difference between doing this and using a voting classifier. The ideas behind doing something like this with signatures is also interesting, as one method may uncover different high order relationships that are compressed onto the predictive value and can then be used to get even higher order information through the second classifier.