RNN - Time To Event

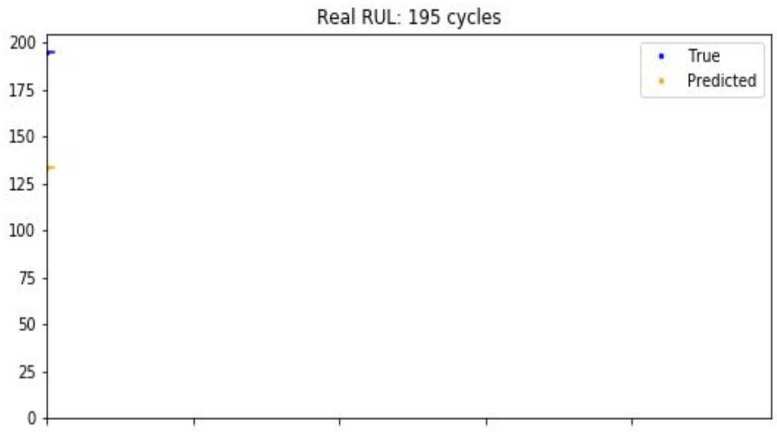
Using Recurrent Neural Networks to predict the time to an event

Manel Maragall

The Time to an Event is the primary outcome of interest in **many fields**

The Time To Failure

- Positively skewed
- Subject to censoring
- Usually long sequences explained by time varying variables
- Useful for Predictive Maintenance in the Pharmaceutical Industry
- Statisticians have been studying it for a long time



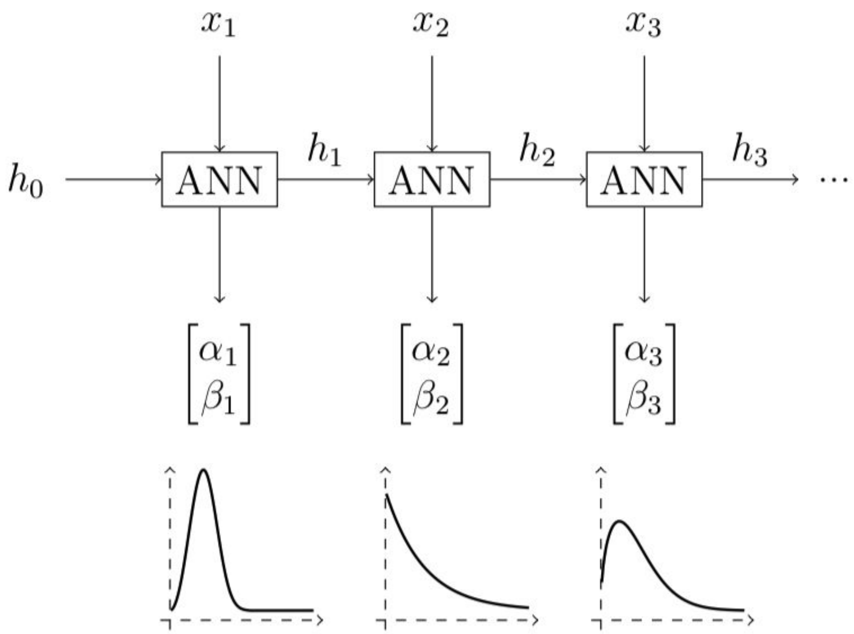
What if we use Recurrent Neural Networks to predict an **statistic distribution**?

WTTE-RNN

Use Recurrent Neural Networks to predict the shape and scale of the Weibull Distribution

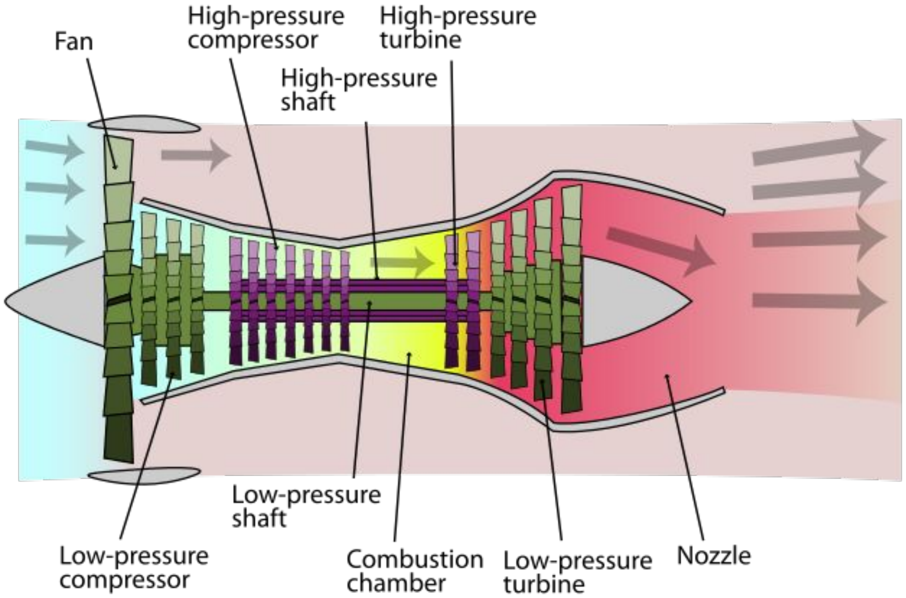
$$\underset{w}{maximize} \ln(\mathcal{L}(w,y,u,x)) := \sum_{t=0}^{T}(u_t \cdot [exp[\left(\frac{y_t+1}{\beta_t}\right)^{\alpha_t} - \left(\frac{y_t}{\beta_t}\right)^{\alpha_t}] - 1] - \left(\frac{y_t+1}{\beta_t}\right)^{\alpha_t})$$

$$\underset{w}{maximize} \ln(\mathcal{L}(w,y,u,x)) := \sum_{t=0}^{T}(u_t \cdot [\alpha_t \cdot \ln\left(\frac{y_t}{\beta_t}\right) + \ln(\alpha_t)] - \left(\frac{y_t}{\beta_t}\right)^{\alpha_t})$$



Turbofan Engine Dataset

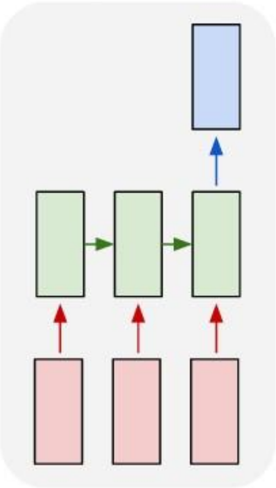
- Simulated with C-MAPSS at NASA
- 6 operating conditions
- 2 failure modes
- Original training set (100) is split into train (80) and validation (20)
- Engines in train are monitored from the start (normal) to the end (failure)
- Test sequences (100) are censored
- 26 variables



Two ways to model the problem

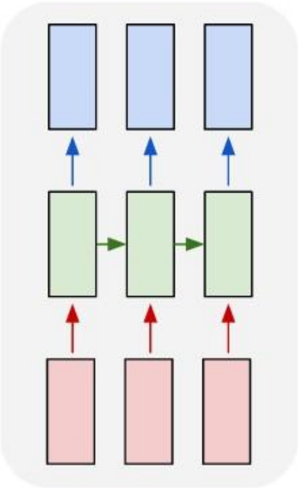
Rolling Window

- Split in sequences of "lookback period"
- Return the output state of the **last unit**
- Sequences are independent



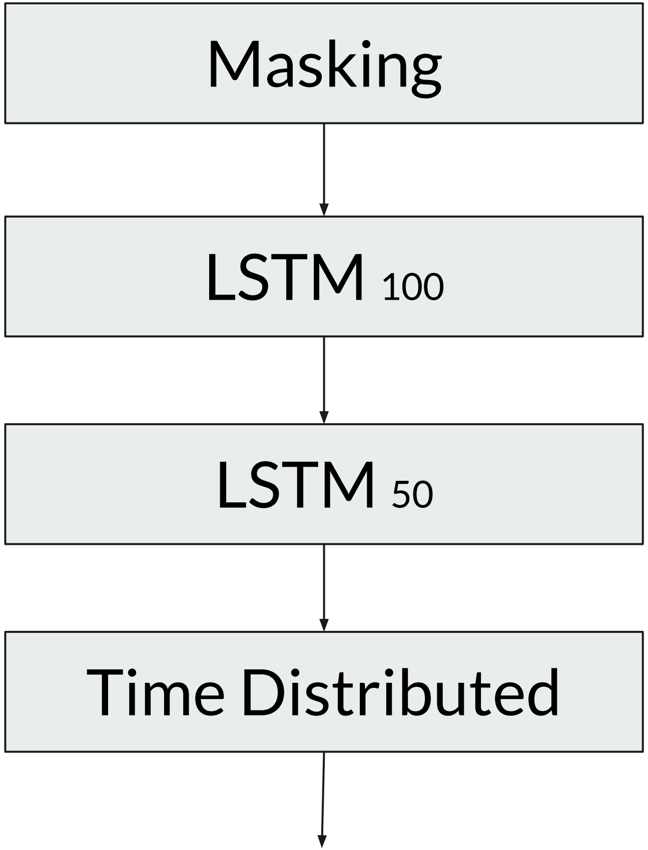
Batch Mode

- Organize the data in batches
- Return the output of **all the units**
- State is preserved
- Easy to shuffle



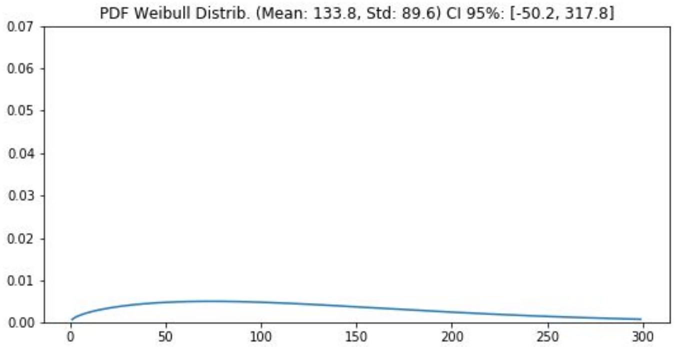
Baseline

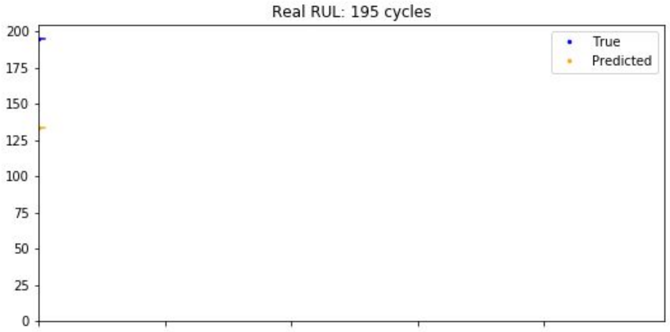
- Batch Mode
- Right Padding
- Recurrent Dropout of 20%
- Early Stopping of 30 epochs
- Shuffle batches
- Exponential activation function



Adapting to WTTE-RNN

- Discrete log-likelihood loss
- Exponential activation for the scale parameter
- Sigmoid activation for the shape parameter
- Initialize scale parameter
- Gradient clipping
- Scale factor in GRU variant

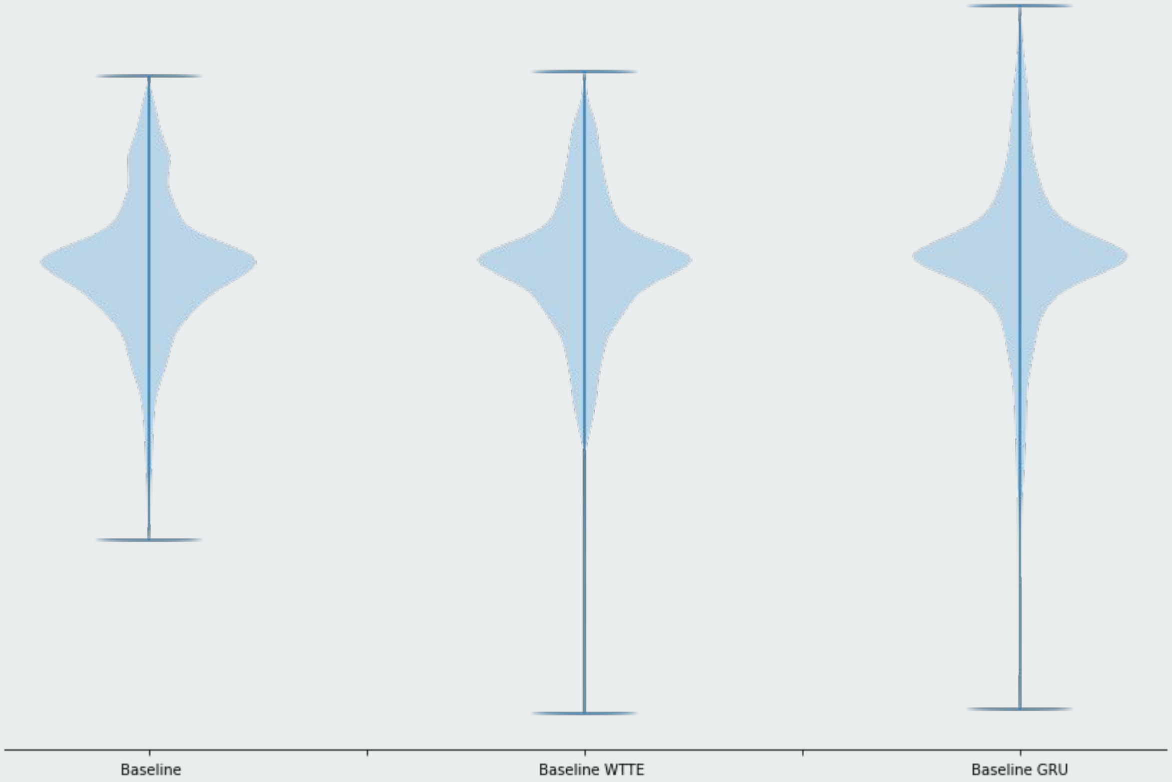




95%

Of the failures from the validation set **can be warned** with an anticipation between 40 and 60 cycles just by triggering an alarm when the standard deviation is below 10 cycles.

MAE	17.4	17.8	18.3
RMSE	24.0	25.3	27.5
R^2	0.87	0.85	0.82



The WTTE-RNN model is **just as good** as the regressor, but it has many **interesting attributes** for Predictive Maintenance

Thanks!

manelmc13@gmail.com
@manelmc
@manelmc1