# Time Series Analysis on real world datasets

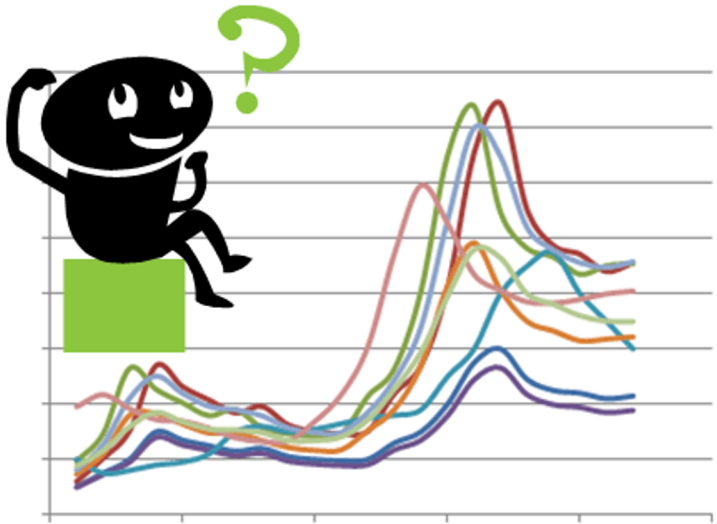
Supervisor : Christian Tomani
Presenters : Carlos Garcia Briones
Jonas Glotz
Thana Guetet

# Table of contents

# 1. Motivation

- Time dependent data analysis


- Uncertainty of Neural Networks



# 2. Goals

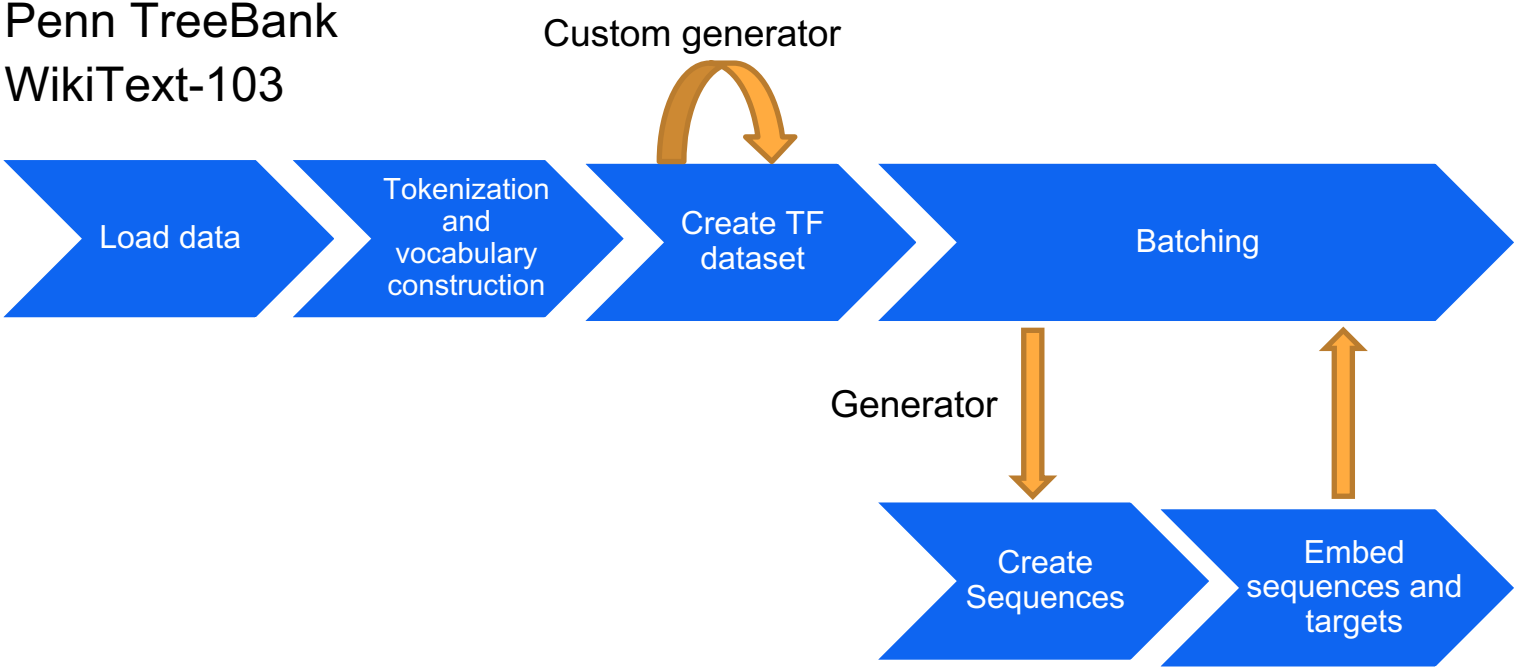
- Implement models with a relatively good performance.
    - Models: RNN, LSTM, GRU, IndRNN, Transformer-XL
- Create models able to deal with perturbed datasets.
- Understand and implement methods for increasing uncertainty awareness.

# 3. Datasets

- NLP Datasets
  - Penn TreeBank
  - WikiText-103
  - Lambada
- Music Datasets
  - JSB Chorales
  - Nottingham
- Sensor Dataset
  - InsectWingBeat

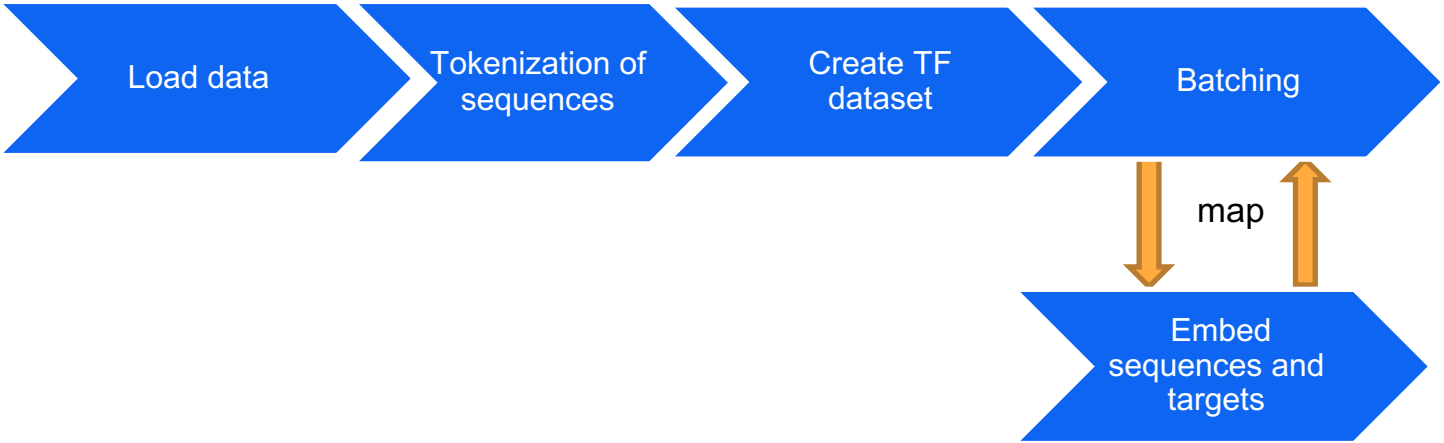
# 4. Workflow – Data Pipelines

- Penn TreeBank
- WikiText-103



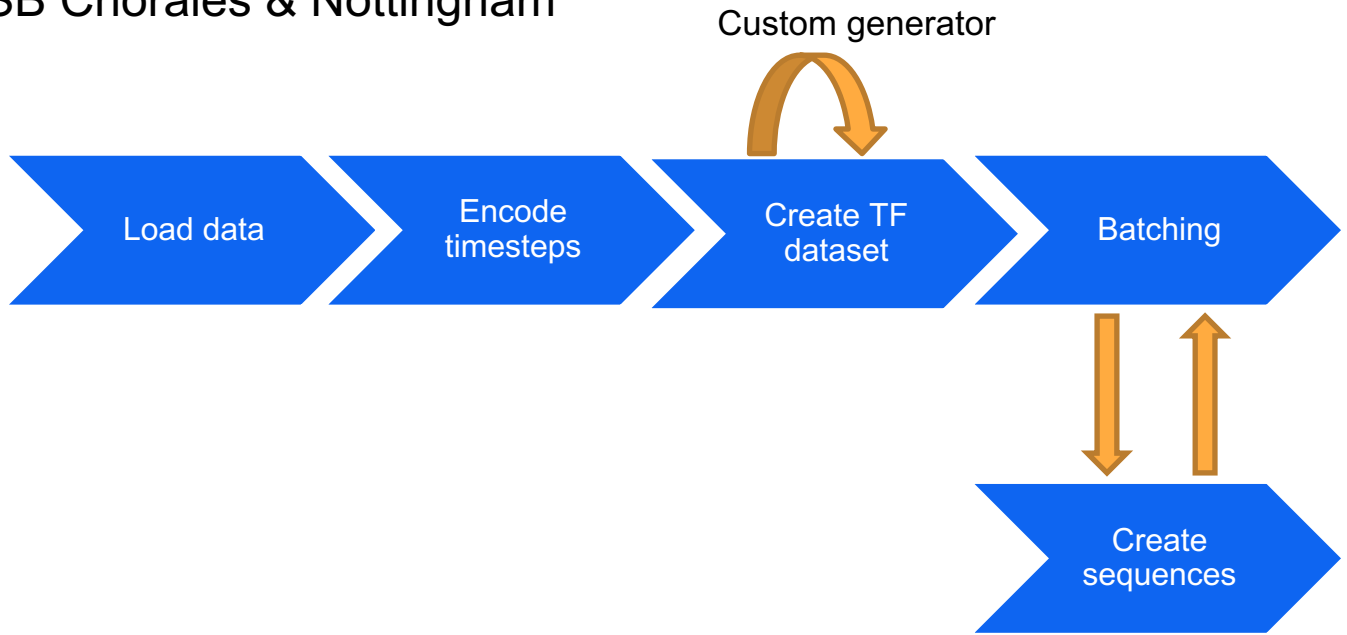
# 4. Workflow – Data Pipelines

- LAMBADA



# 4. Workflow – Data Pipelines

- JSB Chorales & Nottingham



# 4. Workflow – Data Pipelines

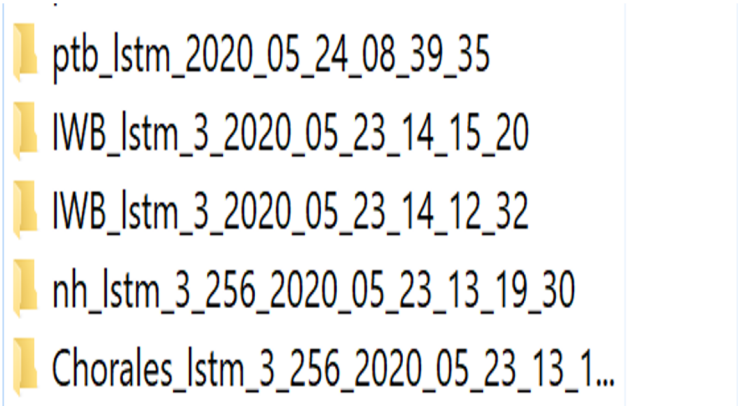
- InsectWingBeat

# 4. Workflow – Model Implementation

- Using Keras Sequential API to implement baseline Models
  - RNN
  - LSTM
  - GRU
- Json configuration files
  - Add flexibility, documentation and effectiveness on training models with different datasets and architectures.
- Building an automated experiment documentation

# 4. Workflow – Model Implementation

- Example of documented experiments

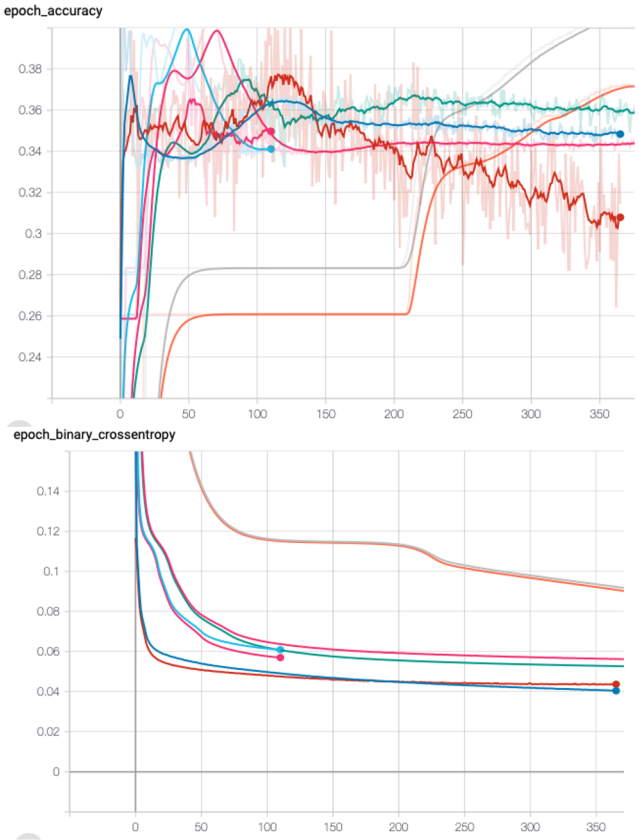



- Example of a Configuration file

```
{
  "dataset": "InsectWingBeat",
  "dataset_args": {...},
  "model": "RNN",
  "model_args": {
    "recurrent_units": [256, 128, 128],
    "linear_hidden_units": [
      {"units": 128, "activation": "relu"},
      {"activation": "softmax", "name": "output"}
    ]
  },
  "learning_rate": 1e-3,
  "compile": {
    "loss": "categorical_crossentropy",
    "optimizer": "adam",
    "metrics": ["accuracy"]
  },
  "logdir": "log",
  "experiment_name": "some_experiment",
  "fit": {"epochs": 10, "verbose": 2}
}
```

# 4. Workflow – Trained Models

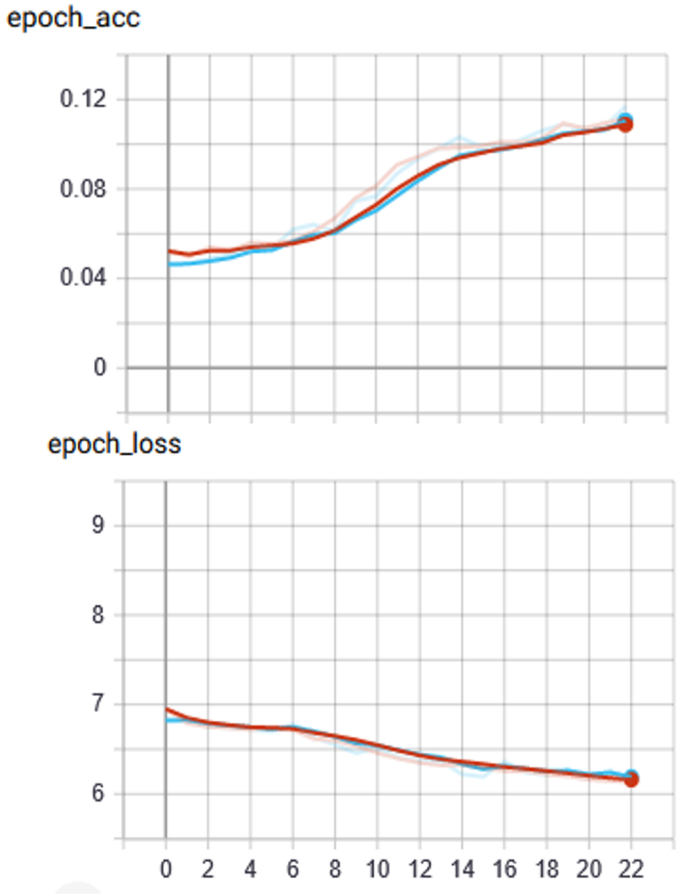
- Nottingham dataset
- Lstm model
- Hyperparameters:
  - Layers: 1,2
  - Hidden units: {256, 500}
  - Learning rate: {1e-4, 1e-5}



# 4. Workflow – Trained Models

- Penn TreeBank Dataset
- Validation perplexity of ~67.7
  (= validation loss of 6.0804)

- perp. = 2^H(x)
- H is the (cross-)entropy,
  i.e. the loss



# 5. Next steps

1. Implement the Transformer-XL and Independent Neural Network models.
2. Implement models that accept sequences with varying lengths to use with the Lambada dataset.
3. Use an informed grid search to determine better suitable hyperparameters and train the baseline models.
4. Train the newly implemented models on the chosen datasets.
5. Expand training script:
    a. Use tensorboard to log other metrics (e.g. F1 Score)

# 6. Potential challenges

- Train models as big as Transformer-XL on the provided resources with limited time.
- The NLP datasets take a long time (>30 min per epoch) to train.
    - Specially hard to optimize with limited resources
- The Lambada dataset is fairly small in comparison to the other NLP datasets, while being the most challenging one

Thank you for listening!