# CTC LSTMs

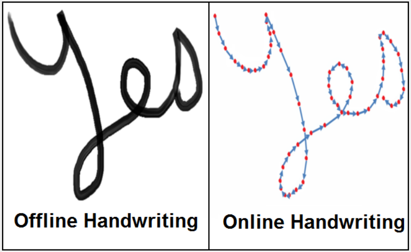
Seminar: Spoken word recognition

Marvin Borner

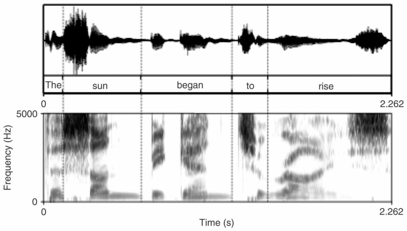
03.07.2023

# Motivation

**on-line** handwriting recognition



**on-line** spoken word recognition



# Introduction

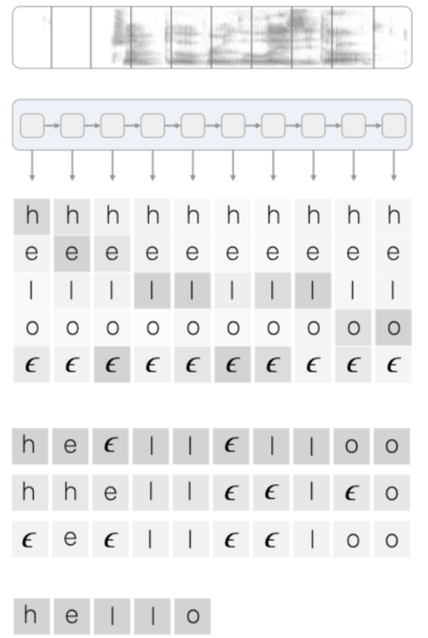
- Networks for sequential data (e.g. RNNs like LSTMs)
- What about variable timings?
    - Traditional models require alignment (e.g. text-audio)
    - Rate of speech/writing varies individually
- **CTC LSTMs can classify sequential data with variable timings**

# CTC LSTM

- **CTC**: Connectionist temporal classification
- **Input**: On-line observations (unaligned)
- **Output**: Continuous probability distribution over all possible labels
- **Training**: Output distribution should fit the probability of each label
- **Loss**: Maximize probability for correct answer

$\Longrightarrow$ Training using normal backpropagation

# Example



0. Train with alphabet $\{h, e, l, o, \epsilon\}$
1. Input: Spectogram (on-line)
2. Feed into LSTM (or other RNN)
3. Returns probability distribution
4. Compute probability of all sequences
5. Merge repeated tokens, remove $\epsilon$

# Conclusion

- Better than most methods (e.g. Markov chains)
- Probably replaced by attention models (transformers)
- **CTC LSTMs can classify sequential data with variable timings**

# thanks

Example code:
`github.com/marvinborner/ctc-lstm`

# Resources

- https://distill.pub/2017/ctc/
- https://towardsdatascience.com/intuitively-understanding-connectionist-temporal-classification-3797e43a86c
- https://www.assemblyai.com/blog/end-to-end-speech-recognition-pytorch/