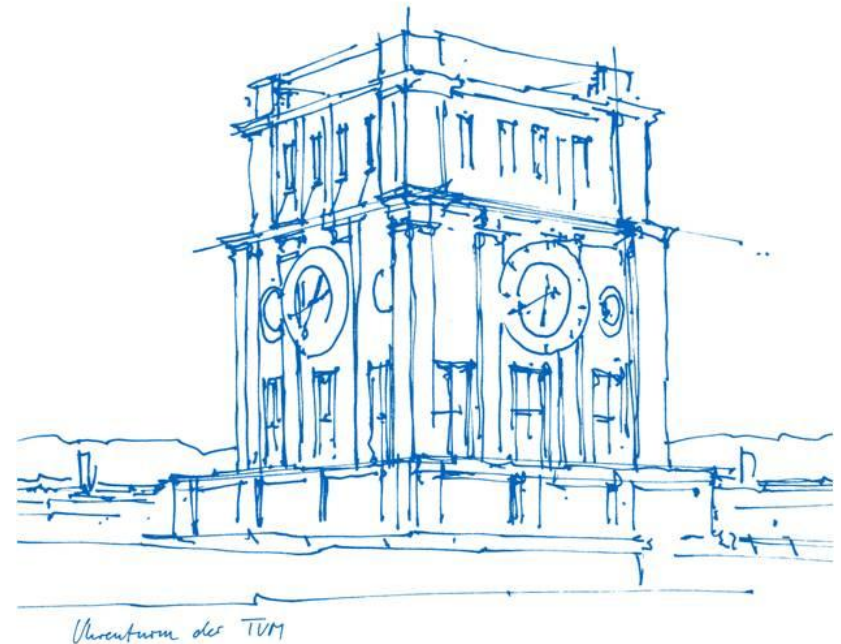# Comparing **Sentiment Analysis** of Discourse Units and Sentences

Group 9: Li Canchen, Hendrik Pauthner, Tim Pfeifle
Technische Universität München
Informatics
Research Group Social Computing
Munich, 5. June 2019

Progress Report

Amazon Review Dataset (~ 450k reviews) → EDUs

Problems:
- extremely large → run it on external machine, parallelize it on 12 cores
- very RAM-heavy → sub-divide it into even smaller parts
- Multiprocessing failed → instead, run segmentation script multiple times in parallel

Progress Report

Existing code in Lua

→ **Do not use Lua!**

→ Implementation in Keras on Tensorflow

Setup on Desktop-Machine (GTX 1070) instead of Colab for development
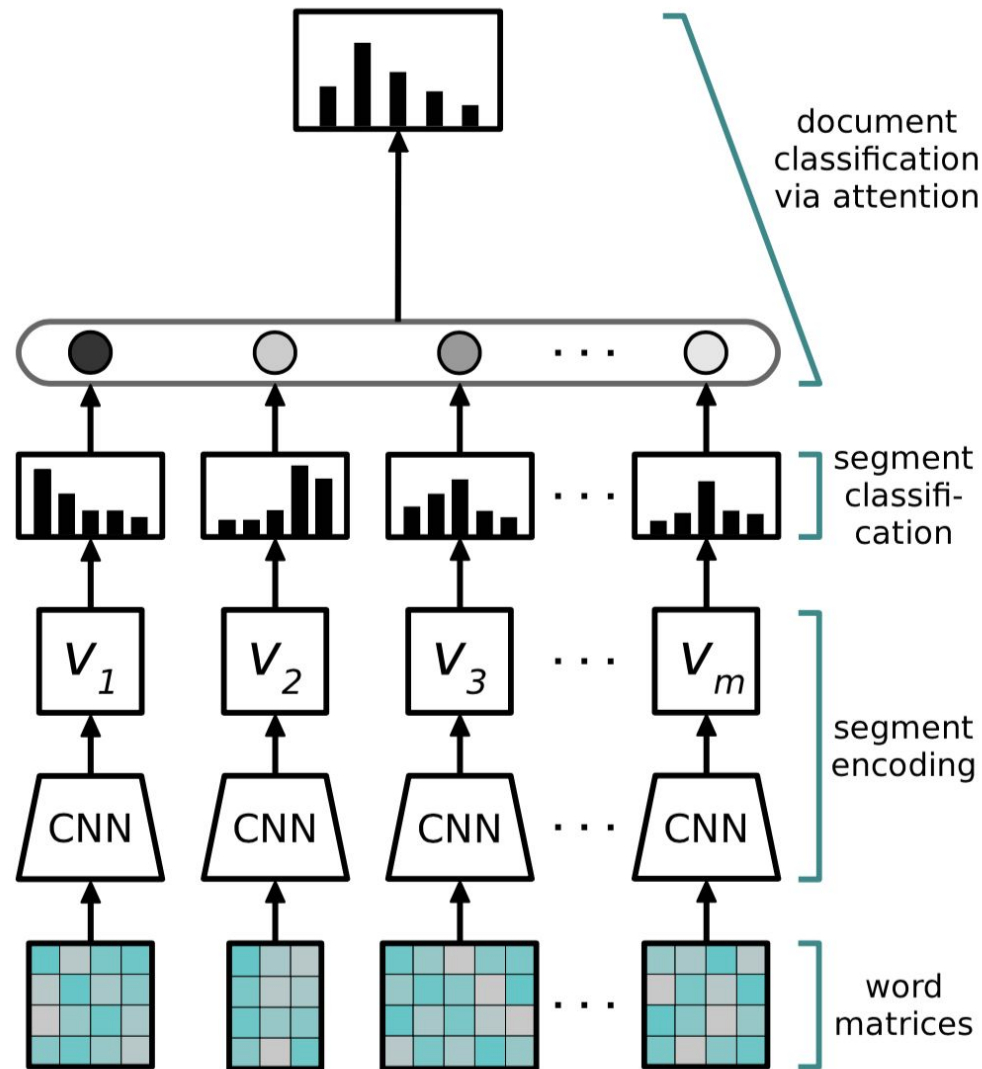
→ Same speed, but easier to work with

- Keeps state
- Remote Interpreter in PyCharm → easy debugging, …
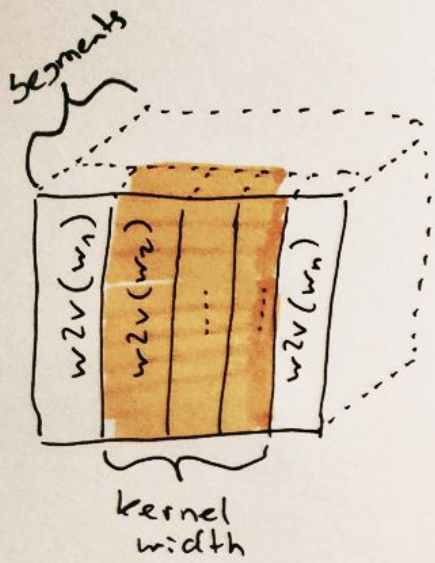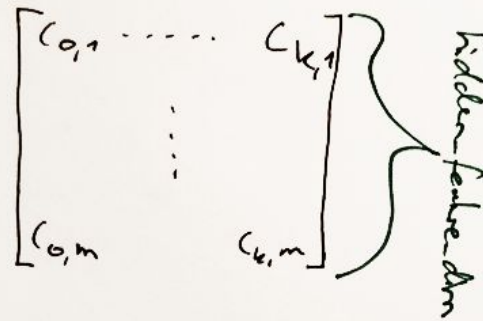
Setup Tensorboard

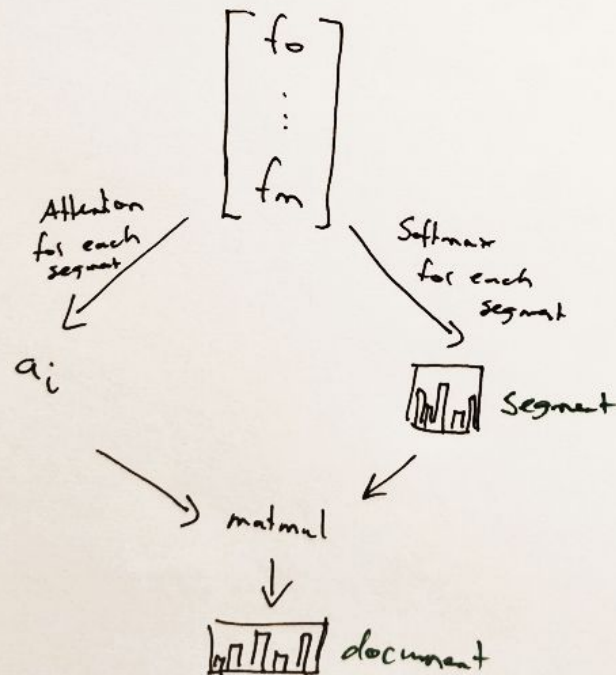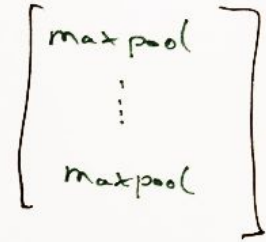[1] MILNET: https://github.com/stangelid/oposum

Model

document
classification
via attention

segment
classifi-
cation

segment
encoding

word
matrices

[1] MILNET: https://github.com/stangelid/oposum

segments

w2v(w_1) w2v(w_2) ... w2v(w_n)

kernel width

$k = n - \text{kernelwidth} + 1$

Segment Encoding

$$\begin{bmatrix} c_{0,1} & \cdots\cdots & c_{k,1} \\ & \vdots & \\ c_{0,m} & & c_{k,m} \end{bmatrix}$$ hidden feature dim

Batch Norm ReLU

$$\begin{bmatrix} \text{maxpool} \\ \vdots \\ \text{maxpool} \end{bmatrix}$$

$$\begin{bmatrix} f_0 \\ \vdots \\ f_m \end{bmatrix}$$

$$\begin{bmatrix} f_0 \\ \vdots \\ f_m \end{bmatrix}$$

Attention for each segment

a_i

Softmax for each segment

Segment

matmul

document

Next Steps

- Training on preprocessed Amazon Review Dataset
- Hyperparameter search
    - Different parameters in paper and in reference implementation
- Comparison of architectures

Muchas Gracias!

Li Canchen, Hendrik Pauthner, Tim Pfeifle

Munich, 22. May 2019