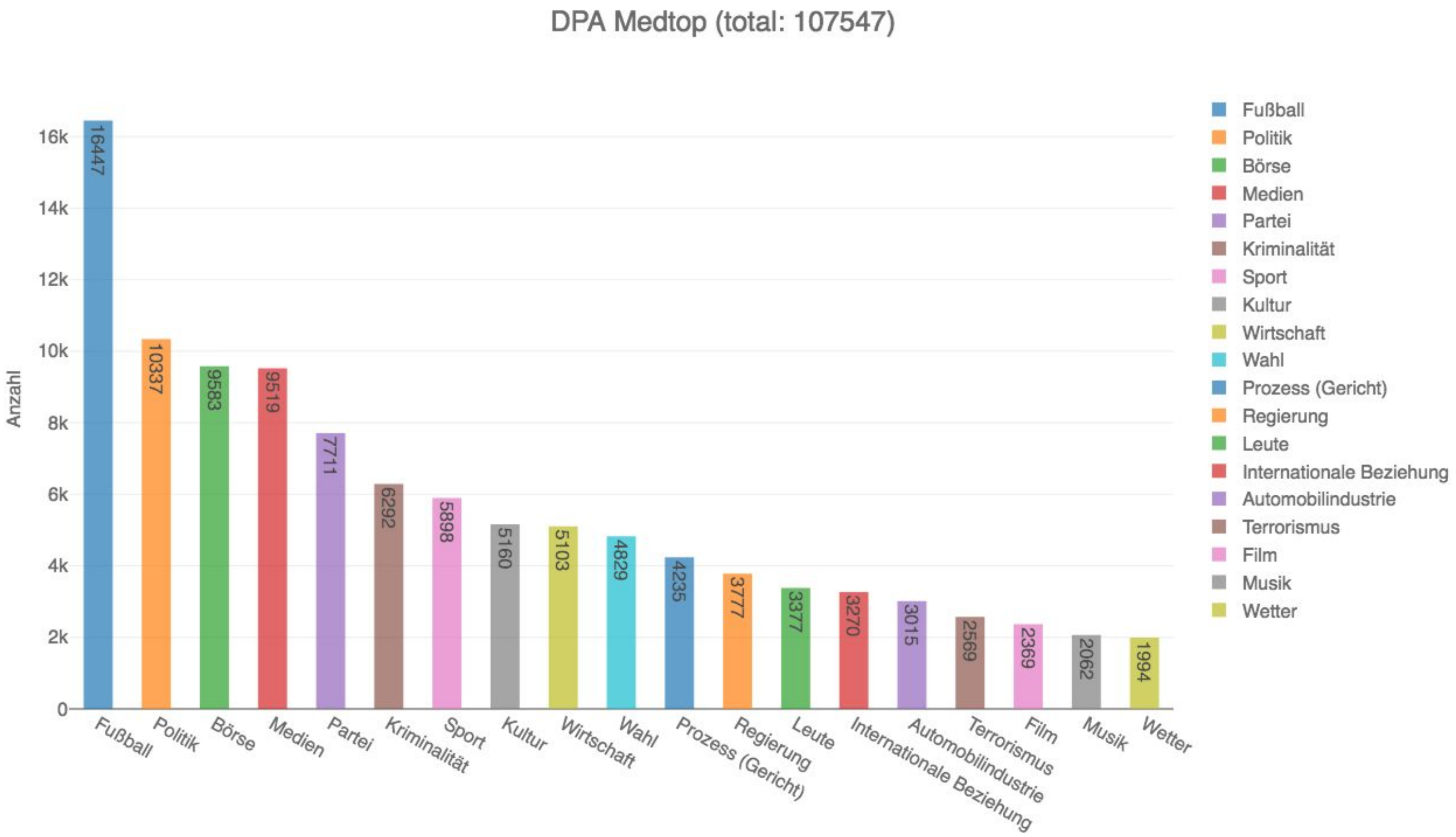
Contextual Neural News Summarization

By Matthias Nitsche

Table of content

Motivation

Motivation



DPA Medtop (total: 107547)
Anzahl
0
2k
4k
6k
8k
10k
12k
14k
16k
16447
10337
9583
9519
7711
6292
5898
5160
5103
4829
4235
3777
3377
3270
3015
2569
2369
2062
1994
Fußball
Politik
Börse
Medien
Partei
Kriminalität
Sport
Kultur
Wirtschaft
Wahl
Prozess (Gericht)
Regierung
Leute
Internationale Beziehung
Automobilindustrie
Terrorismus
Film
Musik
Wetter

Motivation

CSTI Cluster + Testing Machines (Ressources)

- Test Setup (2x Nvidia 1080 ti, 64GB RAM)
- Cluster setup (8x Nvidia P6000, 256GB RAM)
- Benchmarking on standard tasks
- “A deep learning pipeline at scale”

Research Question

How to create an abstractive multi document summarization (extended: dossier) system that leverages context and human input?

Dossiers

Goal: Dossier is a collection of papers or other sources, containing detailed information about a particular subject. It is a meta structure.

- Designed to fit a (topical) narrative / problem definition
- Chronological / Historical / Hierarchical
- Transparent and comprehensible
- Presenting central (non-biased) arguments
- Shallow at first glance, deep at second look

Dossiers

Problem: Dossiers are too hard for a first "summarization" task. Dossiers can be viewed as a Multi Task Learning (MTL) problem on different objectives.

- Key arguments of entire collections with different topics / chronology
- Presenting narratives / presenting a discourse
- Implicit real world knowledge not in the source

Summarization

Goal: Automatic summarization of text, by shortening large amounts of documents keeping their original points.

- Single/ Multi document summarization
- Concise and fluent, based on facts (no fakes!)
- Extractive / Abstractive

Classic summarization

Extractive, e.g. reordering sentences and passages. (abstractive not possible)

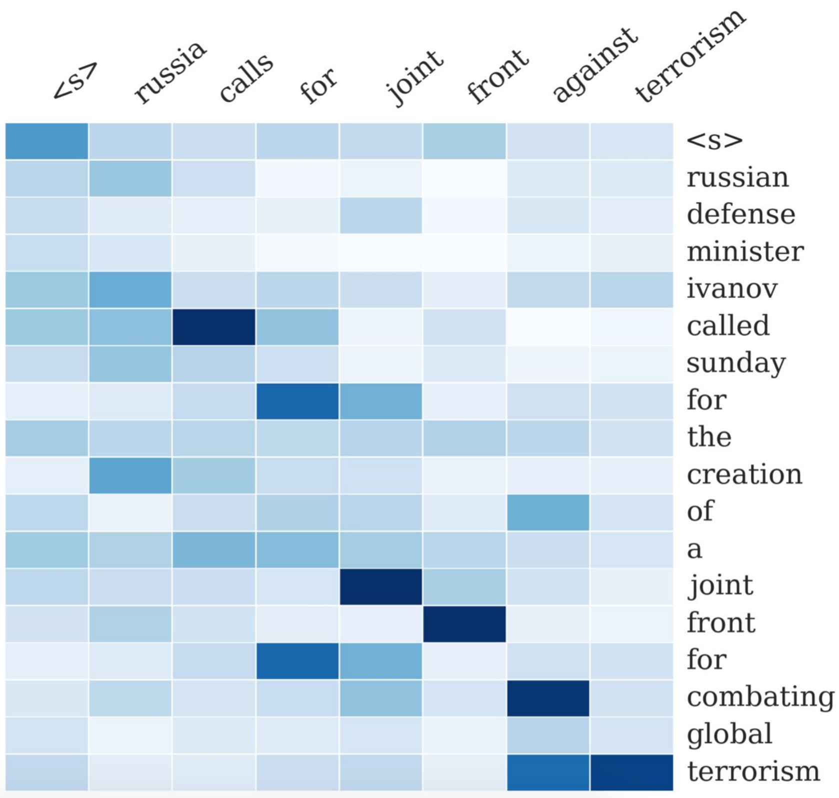
- Cue words - dictionary based
- Title - weighted average of title
- Location - beginning / end of text
- Indicator matrix - word/sentence importance
- Topic based (*LSI*, *LDA*, *TF-IDF*)

Neural Summarization

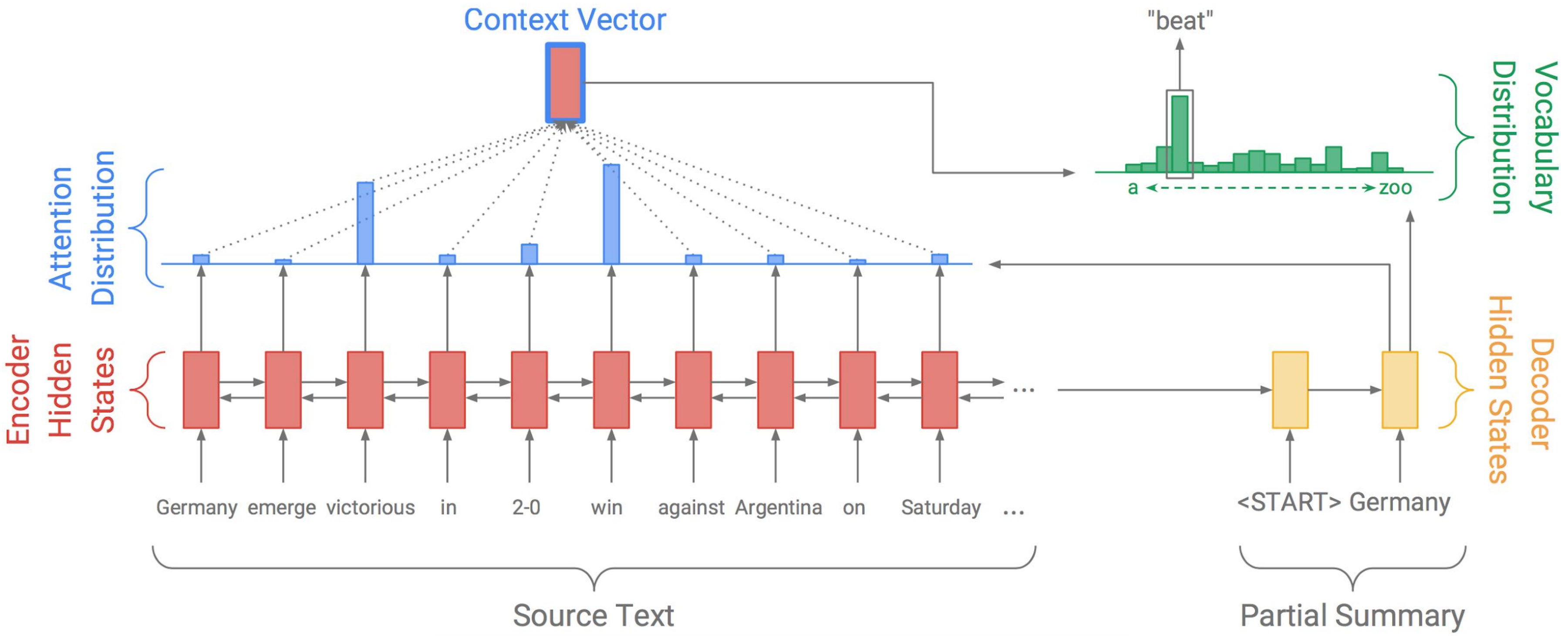
- End-to-end learning (large corpora + vocab problematic)
- Mostly abstractive (autoencoders)
- Window based methods
- Hierarchical e.g. sentence to word to character
- WSD and OOV
- Learning weights for factual checking and repetition
- Multi document summarization is “new”

Rush, et al. (2015)

- Single sentence abstractive summarizations
- Attention glueing encoder / decoder
- Fixed length summaries
- Handles OOV words with a pointer parameter



Rush, et al. (2015)

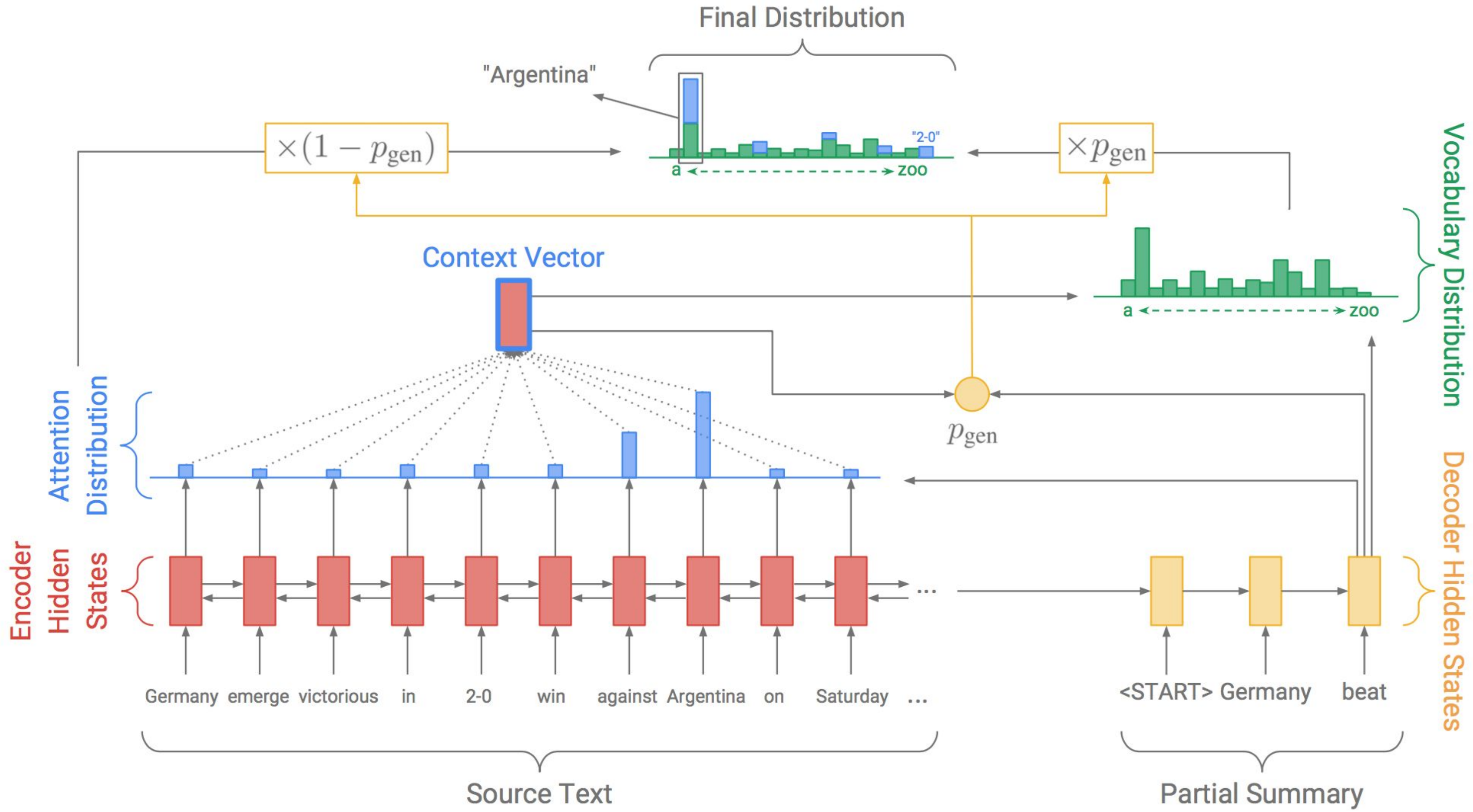


Source: A. See, et al. (2017)

A. See, et al. (2017)

- OVV handling via pointer network - generate or copy?
- Learning a coverage vector, attending less to frequently attended ideas over time
- Handling documents, not sentences
- Handles facts better and removes more nonsensical claims

A. See, et al. (2017)



More concepts

- POS / NER / TF-IDF embeddings - *R. Nallapati, et al.* (2016)
- Hierarchy - learning sentences down to words (using CNNs with LSTMs) - *R. Nallapati, et al.* (2016)
- Use side information during attention - *S. Narayan, et al.* (2017)
- Factual checking, reducing invalid claims - *Z. Cao, et al.* (2018)
- Topic awareness - *L. Wang, et al.* (2018)

Context and Knowledge

Context and Knowledge

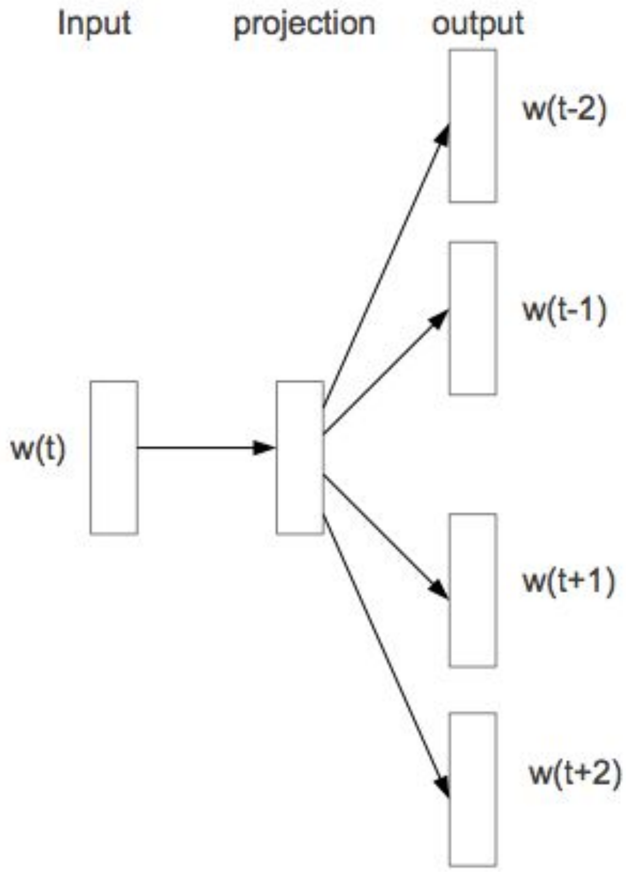
Transfer Learning: Store knowledge conditioned on a domain *Ds* and apply it successfully to a different domain *Dt*.

- Word2Vec, GloVe, Doc2Vec etc. are such models
- In image recognition any model that is able to detect "shapes"

Word2Vec (Mikolov et al. 2013)

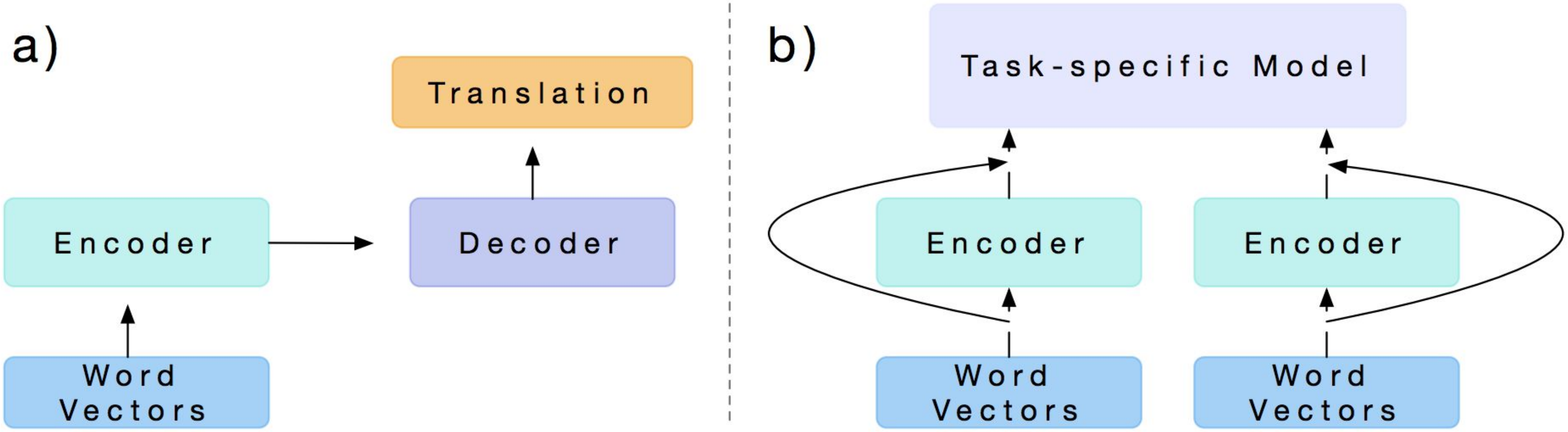
- Unsupervised Word embedding
- *Skip-gram* with *sliding window* as context of center word
- *Negative sampling* with noise contrastive estimation
- Essentially a PMI Matrix
- Related: *GloVe*, *Dep2Vec*, *Dict2Vec*

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-c\leq j\leq c, j\neq 0}\log p(w_{t+j}|w_t)$$



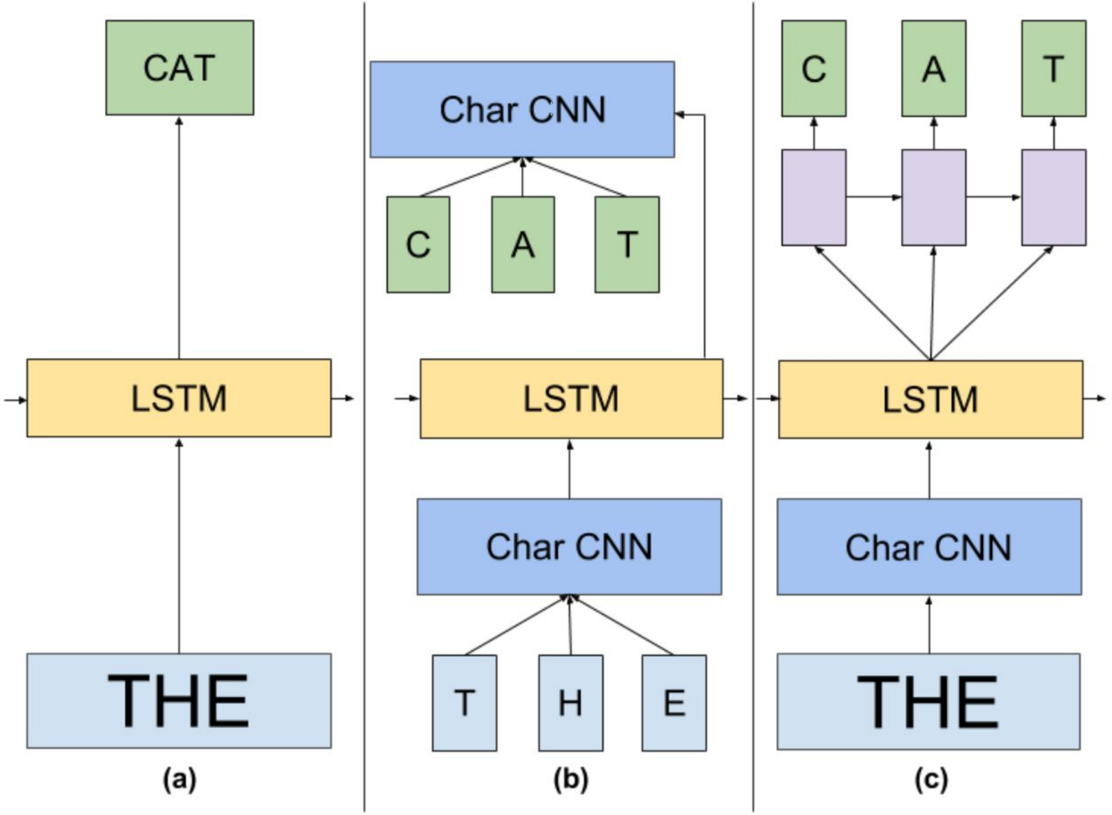
CoVe (B. McCann, et al. 2017)

- Train a NMT system and use the encoder as a language model
- Initialize with *GloVe*
- Use the encoder in a specific language task generically
- Related: *Doc2Vec*



Chars (R. Jozefowicz, et al. 2016)

- Word *LSTM* (a) baseline
- Char CNN (*b*) and Char CNN with softmax *LSTM* (c) outperform
- Fewer parameters
- Handling OOV out of the box
- *Perplexity* down from 50 to 20 (impressive)



News and Knowledge

Goal: Create language models based on all the information and combine them through multi task learning

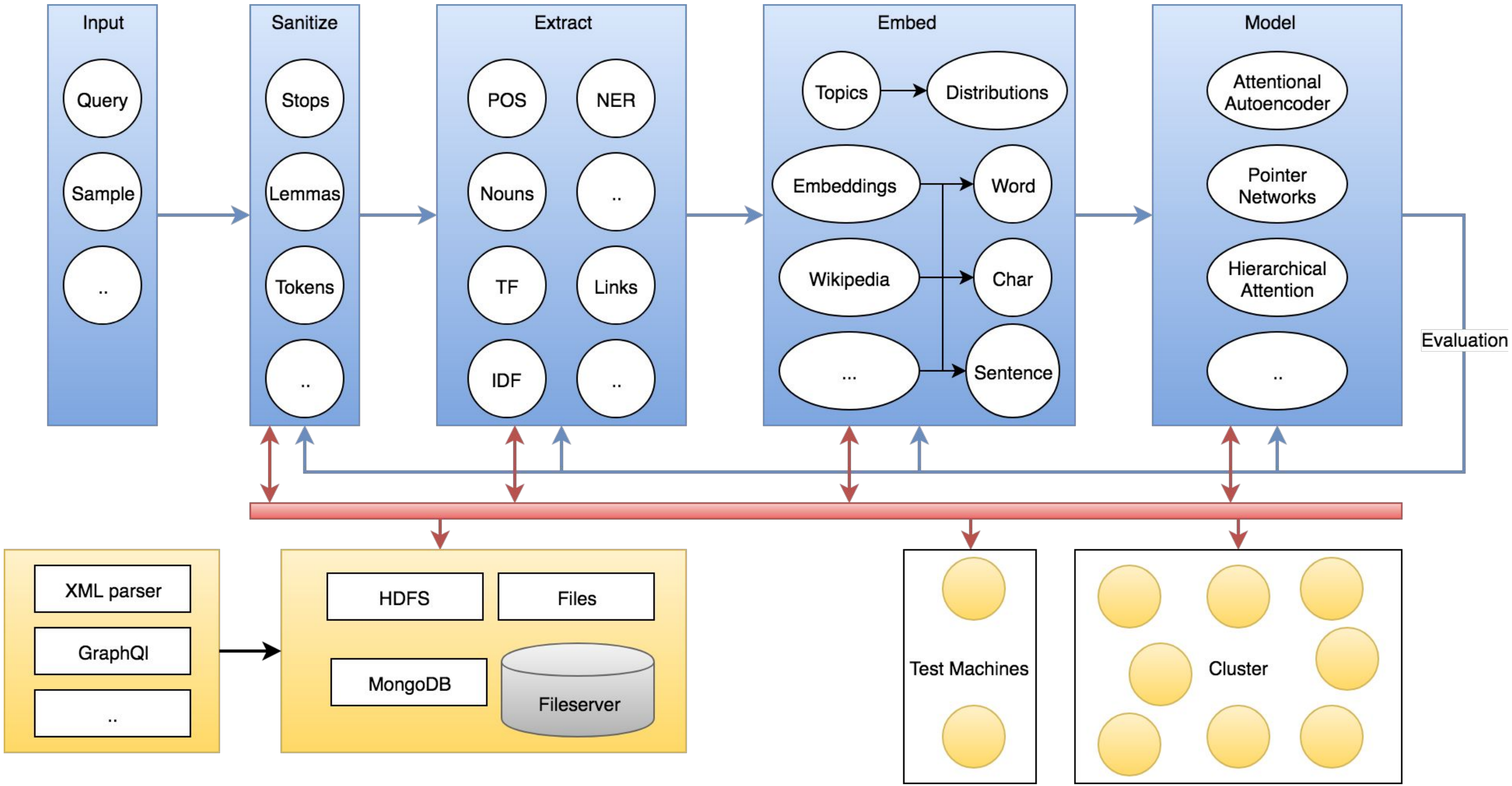
- News present topics over time (event based)
- Curated categories and keyword hierarchies
- Articles are well written (if news source is good)
- Descriptions, headlines, images etc.
- HTML and hyperlinks
- Editor notes and known authors
- Geo locations and publishing locations
- Named entities

Conclusion

- Use language models on the domain data
- Learn a broad task through auxiliary tasks
- Connect different models with multi task learning techniques
- Characters actually make sense, so do hierarchies of words and sentences
- Use side information via attention

Outlook

KDD - Summarization



KDD - Technologies

Languages

Python3, Golang, Bash

Numeric

Numpy, Scipy

Data Munging

Pandas, Apache Arrow

Pipeline

Apache Airflow

File handling

HDFS, MongoDB, Apache Parquet

Learning

Keras, SpaCy, Tensorflow

Visuals

Jupyter, Matplot, Bokeh, Plotly

Deployment

Nvidia-Docker, Docker

GPU

CUDA / CUDNN

Outlook

Risks	*Chances*
• Extremely broad and complex • Improving area, quality is okay • Generating dossiers is “harder” than generating summaries • No specific literature • Real world data	• A lot of ground to cover • Real world data • German models are rare • Multiple interacting DL systems • Minimize risk by defining clear subtasks

Bibliography

N. Hälker (2015), Teilautomatisierte Erstellung von Dossiers auf der Basis von Textmining-Verfahren. HAW Hamburg

A. Rush, et al. (2015), A Neural Attention Model for Abstractive Sentence Summarization.

M. Abadi, et al. (2016), TensorFlow: A system for large-scale machine learning.

I. Goodfellow, et al. (2016), Deep Learning. MIT Press.

R. Nallapati, et al. (2016), Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond.

R. Jozefowicz, et al. (2016), Exploring the Limits of Language Modeling.

M. Allahyari, et al. (2017), Text Summarization Techniques: A Brief Survey. University of Georgia.

A. See, et al. (2017), Get To The Point: Summarization with Pointer-Generator Networks.

Bibliography

S. Narayan, et al. (2017), Neural Extractive Summarization with Side Information.

B. McCann, et al. (2017), Learned in Translation: Contextualized Word Vectors.

Z. Cao, et al. (2018), Faithful to the Original: Fact Aware Neural Abstractive Summarization.

L. Wang, et al. (2018), A Reinforced Topic-Aware Convolutional Sequence-to-Sequence Model for Abstractive Text Summarization.

J. Zhang, et al. (2018), Towards a Neural Network Approach to Abstractive Multi-Document Summarization.

A. Cohan, et al. (2018), A Discourse-Aware Attention Model for Abstractive Summarization of Long Documents.

Thanks!

Any questions ?

- matthias.nitsche@haw-hamburg.de