

# Deep Learning for NLP



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

## Lecture 3 – Word Classification

Nils Reimers

**Course-Website: [www.deeplearning4nlp.com](http://www.deeplearning4nlp.com)**

# Lasagne is sooo....

- .... last Monday.
- New recommendation: Keras (<http://keras.io>)
- Install by: *sudo pip install keras*

# Runtime comparison

- I ran my own Theano implementation, Lasagne and Keras on the NER task:
  - MyTheano implementation: 5.2 seconds / epoch
  - Keras: 9.5 seconds / epoch
  - Lasagne: 16 seconds / epoch
- float 32 vs. float 64
  - Edit ~/.theanorc and add:  
*[global]*  
*floatX = float32*
  - Lasagne runtime reduced from 29.5 seconds to 16 seconds / epoch

# Contribution in NLP in the Time of Deep Learning

- Question: *Does everyone need to build deep neural networks? Or can I just use an existent network, similar to Weka?*
- My hypothesis: In the future there are going to be 4 classes of contributions:
  - Generating human insights into tasks - no deep learning required
  - Invention / description of new tasks - no deep learning required, but SotA-approaches should still be applied
  - Adaption of existent network architectures to other tasks
    - Currently you can find tons of these contributions: Use network X for task Y
    - Fair to good understanding of deep learning required
  - Create a new network architecture for existent task to improve performance
    - These are the really cool contributions
    - Large expertise in deep learning required

*(A non-expert speculation by me)*



# What were your findings on the homework / on the MNIST dataset?

# Recommended Readings



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

- Watch Lecture 2 and Lecture 3 from Richard Sochers CS224d class
- Collobert, Weston et al., *Natural Language Processing (almost) from Scratch*
- Mikilov's papers on Word2Vec:
  - <https://code.google.com/p/word2vec/>
- Pennington et al., GloVe: Global Vectors for Word Representations
  - <http://nlp.stanford.edu/projects/glove/>
- Interesting comment on Word2Vec vs. Glove
  - <https://docs.google.com/document/d/1ydlujJ7ETSZ688RGfU5IMJJsbxAi-kRI8czSwpti15s/edit#heading=h.66rkmh7nd17u>
- Levy and Goldberg, 2014, *Dependency Based Word Embeddings*
  - <https://levyomer.wordpress.com/2014/04/25/dependency-based-word-embeddings/>

# Strong Reading Recommendation

---

- A Primer on Neural Network Models for Natural Language Processing
  - <http://u.cs.biu.ac.il/~yogo/nnlp.pdf>
  - Currently in a draft version

# Questions?



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

- Any questions on word2vec / word embeddings in general?



# Existent Tools

- Word2Vec
  - <https://code.google.com/p/word2vec/>
- Doc2Vec
  - Learns dense representations for phrases, sentences and documents
  - <https://groups.google.com/forum/#!topic/word2vec-toolkit/Q49FIrNOQRo> or Gensim
- GloVe:
  - <http://nlp.stanford.edu/projects/glove/>
- Levy and Goldberg adaption of word2vec:
  - <https://levyomer.wordpress.com/2014/04/25/dependency-based-word-embeddings/>

# Existent Word Embeddings

- German:
  - <https://www.ukp.tu-darmstadt.de/research/ukp-in-challenges/germeval-2014>
- English:
  - Word2Vec-Website
    - Trained on Google News (100 billion tokens)
    - Freebase-vectors (100 billion tokens)
  - GloVe-Website
    - Wikipedia (6 billion tokens)
    - Common Crawl (42 and 840 billion tokens)
    - Twitter (27 billion tokens)
  - Levy-Website
    - Dependency-based embeddings on English Wikipedia

# Impact of Word Embeddings

- In most networks for NLP, word embeddings are the basis
- Having good word embeddings, increases significantly your performance
- Which are the best is hard to tell
  - Try all available
  - Word2Vec & GloVe: Better in capturing semantic similarity
  - Levys dependency based embeddings: Better in capturing syntactic similarities
- Selecting of the corpus, possible preprocessing and fine tuning of the parameters can have a huge impact on the quality

# Develop a Neural Network for NER

- We will develop a Neural Network for Sequence Tagging
  - Can be used for any word classification and sequence tagging task
  - Can be used for POS, NER, Chunking, Parsing and Semantic Role Labeling
- Approach is known as SENNA
  - All details in the paper of Collobert and Weston, *Natural Language Processing (almost) from Scratch*
- In our example we will use the GermEval-2014 dataset (Named Entities in German)
- Pretrained word embeddings are provided
  - Good word embeddings have a significant impact on the performance
  - I spent a lot of time on the training of the word embeddings