

Installationshinweise:

„Deep Learning für NLP mit Pytorch“ – Chi Nhan Nguyen

Hinweise des Trainers:

Google Colab:

Ich präsentiere den Workshop komplett mit Google Colab. Colab ist eine Cloud-basierte Web-Applikation, die auf Jupyter-Notebooks basiert. Die Notebooks laufen in einer virtuellen Linux-Umgebung mit vorinstallierten Data-Science-Paketen und GPU-Support (!):

https://colab.research.google.com/notebooks/basic_features_overview.ipynb

Dies brauchst du während des Workshops:

- Google Account
- Web-Browser (Chrome oder Firefox bevorzugt)
- Internetzugang

Das Tutorial besteht aus drei Notebooks und kann über diese Links kopiert werden (die endgültigen Versionen werden erst etwa zwei Tage davor fertig sein!):

1. 01_pytorch_de.ipynb

(<https://drive.google.com/open?id=1gRLaTgP0H6wxSTMZnBDNn4GPt0w6bxzH>)

2. 02_NER_de.ipynb (https://drive.google.com/open?id=1fenlk8FVa34uxJ5Q_TAar-0K4SgbBXqQ)

3. 03_seq2seq_de.ipynb

(https://drive.google.com/open?id=1D80u5dNollFk1DJ_TYfdLileAEOVFQQ5)

Zusätzliche Pakete und Daten werden während des Workshops auf die virtuelle Maschine heruntergeladen. Logge dich dazu mit deinem Account bei Google ein und gehe in deinen Browser zu den Links. Öffne das Dokument mit Colab und mache jeweils eine Kopie von ihnen in deinem GDrive. Gehe zur Kopie und schließe das Original. Mit deiner eigenen Kopie kannst du nun so viel rumspielen, wie du willst.

Lokale Installation:

Diese Anleitung ist für den Fall, dass du die Notebooks lokal auf deinem Laptop laufen lassen willst. Das empfehle ich dir nur bedingt, ist aber mit etwas Aufwand möglich!

Außerdem sind die Beispiele auf der Colab GPU getestet und werden auf deiner lokalen CPU wesentlich langsamer laufen (außer du hast eine GPU auf deinem Laptop). Also, wenn du den etwas größeren Aufwand nicht fürchtest und sogar eine NVIDIA GPU mit CUDA auf deinem Laptop installiert hast:

Voraussetzungen:

- Anaconda (<https://www.anaconda.com/download/>) mit Python 3.6
- Python-Pakete: numpy, pandas, nb_conda, scikit-learn, gensim, matplotlib, jupyterlab

Installation:

Installiere Anaconda und generiere eine Conda-Umgebung mit der Datei: environment.yml (<https://drive.google.com/open?id=14QXKVMK-fVUAf-RWKDDXw6h0xhZpX4hB>)

Führe folgenden Befehl im Terminal aus:

```
'conda env create -f environment.yml'
```

Und aktiviere die Umgebung 'pytorch':

```
'source activate pytorch' (unter Linux oder MacOS)
```

```
'activate pytorch' (unter Windows)
```

Danach:

- Kopiere die Notebooks von den Links oben und speichere sie auf deinen lokalen Rechner
- Kopiere folgende Dateien:
 - images.zip (und entpacke die zip-Datei)
(<https://drive.google.com/open?id=1hJhDlq6J9cmIMKItx3MBFQzLIJQ3HA0B>)
 - ner_dataset.csv
(<https://drive.google.com/open?id=1OuGw91HbHe6mAmv6ukh4ObcLpnnPqllS>)
 - gensim_en.txt
(<https://drive.google.com/open?id=1BtLeGG2djX9ptcwnvAcFyOzutC95Wdz5>)
 - gensim_de.txt (<https://drive.google.com/open?id=18idLZ78k39Mfaksg9t0YvUAIoZqq1Pon>)
 - eng-de.txt (https://drive.google.com/open?id=1nUmAsx_UZ82aCFg_g0TD3smr-Q8abSyP)
- Starte Jupyterlab (z.B. im Terminal mit 'jupyter-lab')
- Wechsel zum Kernel mit dem Namen 'pytorch' (oben generiert mit environment.yml)
- Lade die Notebooks
- Beachte, dass du die Google Colab Download- und Installationsanweisungen überspringen solltest

Das war's!

Bis zum Workshop,
Chi Nhan