

Natural Language Systems Coursework

Requirements

- Python 3 (developed using Python 3.7.7)

Setup

1. Create a Python virtual environment (e.g. `conda create -name nls python=3.7.7` or `virtualenv env`)
2. Activate the environment (e.g. `conda activate nls` or `source ./env/bin/activate`)
3. Install dependencies: `pip install -r requirements.txt`
4. Install PyTorch (`pip install`):
 - Windows (CUDA):
`https://download.pytorch.org/whl/cu101/torch-1.4.0-cp37-cp37m-win_amd64.whl`
 - Windows (CPU):
`https://download.pytorch.org/whl/cpu/torch-1.4.0%2Bcpu-cp37-cp37m-win_amd64.whl`
 - Linux (CUDA):
`https://download.pytorch.org/whl/cu101/torch-1.4.0-cp37-cp37m-linux_x86_64.whl`
 - Linux (CPU):
`https://download.pytorch.org/whl/cpu/torch-1.4.0%2Bcpu-cp37-cp37m-linux_x86_64.whl`
 - MacOS:
`https://download.pytorch.org/whl/cpu/torch-1.4.0-cp37-none-macosx_10_9_x86_64.whl`
5. Create a `data` directory in the project root and extract all required data (corpus and target words) to separate folders in it. Alternatively, you can specify where to load the data from in the coursework scripts (`src/coursework{1/2}.py`).
6. Run the program from the root of the project: `python src/coursework{1/2}.py`

Project structure

- `src` - All the code written for the courseworks
 - `coursework1.py` - Entry of the program for coursework 1; should be run when evaluating it
 - `coursework2.py` - Entry of the program for coursework 2; should be run when evaluating it
 - `corpus.py` - Contains operations on the corpus
 - `classifier.py` - Contains the trainable bag-of-words sentiment classifier
 - `train.py` - Contains the training and evaluation script for the classifier in `classifier.py`
 - `utilities.py` - Contains utility methods, which are used by other scripts in the project