

README

This repository contains the models for the paper *A Neural Network Multi-Task Learning Approach to Biomedical Named Entity Recognition* by Gamal Crichton, Sampo Pyysalo, Billy Chiu and Anna Korhonen. There are several files in the models folder:

- `baseline.py`: The MLP model used as a baseline for the experiments.

Example Usage: `python baseline.py 'path/to/dataset' 'path/to/vectorfile'`

- `baseline_config.py`: The configurable variables and their values for the MLP baseline model (`baseline.py`).
- `config.py`: The configurable variables and their values for the convolutional models.

- `MT-dependent.py`: The multi-task Dependent Model.

Example usage: `python MT-dependent.py 'path/to/data-files' 'dataset-1,...,dataset-n' 'path/to/vectorfile'`

- `multi-output_MT.py`: The multi-output multi-task model.

Example usage: `python multi-output_MT.py 'path/to/data-files' 'dataset-1,...,dataset-n' 'path/to/vectorfile'`

- `multi-output_MT-var-dataset.py`: The model used in the multi-task experiments which investigated the effect of multi-task learning on datasets of various sizes. Specify the *percent-keep* command to determine how much of the training examples of dataset whose size you wish to vary to randomly keep. This **must** be the first dataset specified, all other datasets will train with full training data.

Example usage: `python multi-output_MT-var-dataset.py --percent-keep 0.5 'path/to/data-files' 'path/to/reduced-dataset,path/to/whole-dataset' 'path/to/vectorfile'`

- `single_task.py`: The single task model.

Example usage: `python single_task.py 'path/to/dataset' 'path/to/vectorfile'`

Note: The experiments in the paper applied the Viterbi algorithm to the outputs. Use the `--viterbi` flag to replicate this.

License: The code is provided under MIT license.