

Beschreibung des Codes

Zur Bachelor Arbeit

Florian Omiecienski

1 Beschreibung

In diesem Ordner befindet sich der python3-Code, der die Experimente der Bachelor Arbeit umsetzt. Die Modelle und Experimente werden in dem Order *./modelle123/* behandelt. Der Code und die Programme zum Erstellen der Akteur-Embeddings befindet sich in dem Ordner *./entity-embedding-bootstrap/*. In beiden Ordnern liegt ein Dokument, welches den jeweiligen Ordnerinhalt beschreibt. Dort sind auch die jeweilige Installations- und Benutzungs-Hinweise zu finden.

Die Ordnerstruktur entspricht der Struktur die benötigt wird, um die Experimente durchzuführen. Teile der Daten mussten jedoch wegen ihrer Größe ausgespart bleiben. Alle Experiment-Ergebnisse sind enthalten wurden jedoch in zip-Dateien komprimiert. Es fehlen die trainierten Modell-Instanzen sowie die FastText-Daten. Zudem fehlen die Wikipedia-Artikel in *./wikipedia-daten/database/pages*, die Meta-Daten (Download-Zeitpunkt und Artikel-URL) sind aber noch vorhanden. Die DebateNet2.0-Daten wurden in eine Passwort-gesicherte zip-Datei komprimiert und müssen entpackt werden, bevor sie verwendet werden können. Das Passwort ist auf der CD der Bachelor-Arbeit gespeichert. Die Ergebnisse der zwei Experimente sind unter *./ergebnisse/* und *./ergebnisse-random/* zu finden. Der letzte der beiden Pfade enthält die Ergebnisse der Experimente, die mit zufälligen Akteur-Embeddings durchgeführt wurden. Die Ergebnis-Dateien können mit dem Programm *./modelle123/show-evaluation.py* betrachtet werden. Die trainierten und die zufällig erzeugten Akteur-Embeddings sind unter *./akteur-embeddings/* zu finden.

2 Benutzungs-Hinweis

Laden sie die deutschen FastText-Vektoren¹ in den Ordner *./data-fasttext-german/*. Durch ausführen der zwei Skripte *./entity-embedding-bootstrap/entity.bootstrap.sh* und *./modelle123/experimente.sh* können nun erst die Akteur-Embedding erstellt werden und dann die Modelle trainiert werden. Es ist zu beachten, dass das durchführen der Experimente bis zu 26 Stunden dauern kann.

¹Textversion der deutschen Vektoren <https://fasttext.cc/docs/en/crawl-vectors.html>

```
sh ./entity-embedding-bootstrap/entity_bootstrap.sh  
sh ./modelle123/experimente.sh
```