

SW05 - Gruppe 1

Natural Language Processing Basics 3

M. Nebroj, S. Hauri, S. Ineichen, R. Schwarzentruher

2019-10-22

Testatübung SW05

Aufgabe 1

Führen sie die einzelnen Schritte aus. Zunächst soll das entsprechende Modell mit folgenden Parametern erstellt werden: Features: 200 Context Window Size: 5

```
model = Word2Vec(sg=1, # 1 for skip-gram; otherwise CBOW
                  size=300, # num of features
                  window=5, # window size
                  workers=4) # count cpu
```

Zudem sollen folgende Fragen beantwortet werden:

1. Wie ähnlich sind sich Jon und Ygritte?

```
model.wv.similarity('jon', 'ygritte')
```

```
[output]
0.5669464
```

2. Wie sieht der Vektor aus zu Arryn?

```
model.wv.distances('arryn')
```

```
[output]
[0.69882214 0.7793826 0.64857626 ... 0.5807413 0.66005266 0.66054785]
```

3. Welche 7 Wörter sind am ähnlichsten zu Lannister?

```
model.wv.most_similar('lannister', topn = 7)
```

```
[output]
[('pays', 0.6897309422492981), ('tywin', 0.6743762493133545), ('jaime', 0.6510382890701294),
↪ ('debts', 0.650812029838562), ('kingslayer', 0.6498616933822632), ('kevan',
↪ 0.6497187614440918), ('stafford', 0.6121554374694824)]
```

4. Was ergibt: Stark + Winterfell - Dragons?

```
model.wv.most_similar(positive=['stark', 'winterfell'], negative=['dragons'], topn=1)
```

```
[output]
[('eddard', 0.6262251138687134)]
```

5. Was passt nicht „Winterfell, Riverrun, Jaime“?

```
model.wv.doesnt_match('winterfell riverrun jaime'.split())
```

```
[output]
'jaime'
```

6. Wie ähnlich sind sich folgende Sätze:

Satz 1: "Hodor That was all he ever said" Satz 2: "Hold the door"

```
text1 = 'Hodor That was all he ever said'.lower().split()
text2 = 'Hold the door'.lower().split()
similarity = model.wv.wmdistance(text1, text2)
print (similarity)
```

[output]

3.5015299231818906