

Inhaltsverzeichnis

1	Einleitung	2
1.1	Motivation	2
1.2	Problemstellung	2
1.3	Aufbau der Arbeit	2
2	Begriffverzeichnis	2
3	Daten und Vorverarbeitung	3
3.1	Datenbasis	3
3.2	Externe Programme und Hilfsmittel	3
3.3	Vorverarbeitung	3
4	Word2Vec	3
4.1	Parameter	3
4.2	CBOW	3
4.3	Skip-gram	3
4.4	Negative sampling	3
4.5	Hierarchical softmax	3
4.6	Distanz zwischen Vektoren im Word2Vec Modell	3
5	Wikipedia-Korpus	3
5.1	Gesamtkorpus	3
5.2	Teilkorpus	3
5.3	Testdaten	3
5.4	Vergleich und Analyse	3
6	Experimente	3
6.1	Synnonymsuche durch Rekursion	4
6.1.1	Beschreibung	4
6.1.2	Durchführung	4
6.1.3	Interpretation/Ergebnis	4
6.2	Konkretisierungen	4
6.2.1	Beschreibung	4
6.2.2	Durchführung	4
6.2.3	Interpretation/Ergebnis	4
6.3	Verallgemeinerungen	4
6.3.1	Beschreibung	4

6.3.2	Durchführung	4
6.3.3	Interpretation/Ergebnis	5
6.4	Unterschiedliche Beziehungen	5
6.4.1	Beschreibung	5
6.4.2	Durchführung	5
6.4.3	Interpretation/Ergebnis	5
6.5	Mehrdeutigkeit	5
6.5.1	Beschreibung	5
6.5.2	Durchführung	5
6.5.3	Interpretation/Ergebnis	5
7	Zusammenfassung	5
8	Ausblick	5
9	Fazit	5
10	Anhang	5
10.1	Testdaten	5

1 Einleitung

1.1 Motivation

1.2 Problemstellung

1.3 Aufbau der Arbeit

2 Begriffverzeichnis

Begriff	Erklärung
Ähnliche Worte	Im Word2Vec-Modell mit der Methode <i>most_similar()</i> erhaltene Worte.
SVM	Support Vector Machine
NBC	Naive Bayes Classifier

3 Daten und Vorverarbeitung

3.1 Datenbasis

3.2 Externe Programme und Hilfsmittel

3.3 Vorverarbeitung

4 Word2Vec

4.1 Parameter

4.2 CBOW

4.3 Skip-gram

4.4 Negative sampling

4.5 Hierarchical softmax

4.6 Distanz zwischen Vektoren im Word2Vec Modell

5 Wikipedia-Korpus

5.1 Gesamtkorpus

5.2 Teilkorpus

5.3 Testdaten

5.4 Vergleich und Analyse

6 Experimente

In diesem Kapitel sollen die unterschiedlichen Korpora (Gesamtkorpus¹ und Techkorpus²) untersucht werden. Dies soll durch ausgewählte Fragestellungen realisiert werden.

Die Fragestellungen beziehen sich immer auf die Ergebnisse, die aus den Testdaten³ erhaltenen ähnlichen Worten.

Jedes Experiment ist in drei Teile aufgeteilt Beschreibung, Durchführung und Interpretation/Ergebnis.

6.1 Synonymsuche durch Rekursion

6.1.1 Beschreibung

Es soll untersucht werden, ob man Synonyme zum Testwort erhält, wenn man die ähnlichen Worte dieses Testwortes erneut im Model mittels der Methode *most_similar()* sucht.

6.1.2 Durchführung

.

6.1.3 Interpretation/Ergebnis

.

6.2 Konkretisierungen

6.2.1 Beschreibung

.

6.2.2 Durchführung

.

6.2.3 Interpretation/Ergebnis

.

¹vgl. 5.1

²vgl. 5.2

³vgl. Testdaten unter 10.1

6.3 Verallgemeinerungen

6.3.1 Beschreibung

.

6.3.2 Durchführung

.

6.3.3 Interpretation/Ergebnis

.

6.4 Unterschiedliche Beziehungen

6.4.1 Beschreibung

.

6.4.2 Durchführung

.

6.4.3 Interpretation/Ergebnis

.

6.5 Mehrdeutigkeit

6.5.1 Beschreibung

.

6.5.2 Durchführung

.

6.5.3 Interpretation/Ergebnis

.

7 Zusammenfassung

8 Ausblick

9 Fazit

10 Anhang

10.1 Testdaten