

Experiment 1 (Untersuchung der Merkmale)

Veränderte unabhängige Variablen :

Merkmalsmenge, Art der Trainingsmenge

Aufbau:

Testmenge : 818# manuelle Annotationen

Trainingsmenge :

1. Durchgang auto Samples M--G; M--S ----> M ist immer besser
2. Durchgang auto Samples M--G; ----> M ist immer besser
3. Durchgang auto Samples M--S, ----> M ist immer besser

Merkmalsmenge:

Durchgang 1:

alle einzeln:

(bag_of_pos, bag_of_words, tf_idf, readability, variety, perfect_tense, passive, adjective, sentence_start, modal_verb)

Durchgang 2:

Kombination mit IR-Merkmalen und besten diversen Merkmalen:

bop, bow;

bop, readability;

bop, tf_idf;

bow, readability;

bow, tf_idf;

tf_idf, readability

Durchgang 3:

Kombination alle diversen Merkmalen und IR:merkmale mit allen diversen Merkmalen:

bag_of_pos, all diverse; bag_of_words, all diverse; tf_idf, all diverse; all diverse

Dieses Experiment wurde jeweils mit Normalisierung und ohne Normalisierung durchgeführt

Experiment 2 (Entwicklung nach Trainings_Samples)

Veränderte unabhängige Variablen :

Merkmalsmenge, Anzahl der Trainingsmenge

Aufbau:

Testmenge : 818# manuelle Annotationen

Trainingsmenge : Alle Durchgänge auto Samples M--G; M--S ---> M ist immer besser

Merkmalsmenge: Alle Durchgänge bag_of_pos; bag_of_words; tf_idf; all diverse

Anzahl der Trainingsmenge:

13 Durchgänge mit durchgehender Steigerung der Trainingsmenge.

2-5-10-50-100-200-500-1.000-2.000-5.000-10.000-20.000-40.000

Experiment wurde mit Normalisierung durchgeführt

Experiment 3 (Entwicklung nach Test_Samples)

Veränderte unabhängige Variablen :

Merkmalsmenge, Anzahl der Testmenge

Aufbau:

Testmenge : 818# manuelle Annotationen + steigende Anzahl aus M--G; M--S

Trainingsmenge : Alle Durchgänge auto Samples M--G; M--S ---> M ist immer besser

Merkmalsmenge: Alle Durchgänge bag_of_pos; bag_of_words; tf_idf; all diverse

Anzahl der Testmenge:

13 Durchgänge mit durchgehender Steigerung der Testmenge -->

2 - 5 - 10 - 50 - 100 - 200 - 500 - 1.000 - 2.000 - 5.000 - 10.000 - 20.000 - 40.000

Experiment wurde mit Normalisierung durchgeführt

Experiment 4 (Vorhersage der Auto_Samples)

Veränderte unabhängige Variablen :

Merkmalsmenge, Trainingsmenge, Testmenge

Aufbau:

Testmenge : # 818

1. Durchgang auto Samples M--G; M--S ---> M ist immer besser
2. Durchgang auto Samples M--G; ---> M ist immer besser
3. Durchgang auto Samples M--S, ---> M ist immer besser

Trainingsmenge : # 10k

1. Durchgang auto Samples M--G; M--S ---> M ist immer besser
2. Durchgang auto Samples M--G; ---> M ist immer besser
3. Durchgang auto Samples M--S, ---> M ist immer besser

Merkmalsmenge:

Durchgang 1:

alle einzeln:

(bag_of_pos, bag_of_words, tf_idf, readability, variety, perfect_tense, passive, adjective, sentence_start, modal_verb)

Durchgang 2:

Kombination mit IR-Merkmalen und besten diversen Merkmalen:

bop, bow;
bop, readability;
bop, tf_idf;
bow, readability;
bow, tf_idf;
tf_idf, readability

Durchgang 3:

Kombination alle diversen Merkmalen und IR:merkmale mit allen diversen Merkmalen:

bag_of_pos, all diverse; bag_of_words, all diverse; tf_idf, all diverse; all diverse

Experiment wurde mit Normalisierung durchgeführt