

Spezifikationsvortrag

Fabian Düker, Uli Steinbach

Universität Heidelberg, Institut für Computerlinguistik
Softwareprojekt, SoSe 2018

Prof Dr. Katja Markert

12.06.2018

Übersicht I

Übersicht

Aufgabe

Inhaltliche Spezifikation

inh. Spezifikation

Modularisierung und Aufgabenverteilung

Modularisierung und Aufgabenverteilung

Programmarchitektur, Datenstrukturen

Programmarchitektur und Datenstrukturen

Übersicht

Autom. Erstellung eines Lexikons für die Erkennung von
Abusive Words

Anwendung auf Germeval 2018 Task I \Rightarrow Binäre Klassifikation von
5000 Tweets

Problemstellung

- ▶ Problem: Hatespeech ist in ständiger Veränderung begriffen (Neologismen, Ambiguität, Kontext/Domäne)
- ▶ Wiegand et al. 2016: Erstellung eines englischen Lexikons mit guten Ergebnissen auf cross-domain Evaluation
- ▶ SentiWS: Lexikon mit negativen Wörtern für das Deutsche

Lösungsansatz

- ▶ Erstellung Baselexikon aus SentiWS neg. Sentiment-Lexikon
- ▶ halbautomatische Erweiterung des Baselexikons mit deutschen Schimpfwörtern
- ▶ autom. Erweiterung mittels graphbasiertem Label-Propagation-Algorithmus
- ▶ Anwendung auf Germeval 2018 Datenset und Evaluation

halbautom. Erweiterung mit deutschen Schimpfwörtern

- ▶ Genius API: Erstellung eines Deutschrappkorpus
- ▶ Deutschrapp: zeitgemäße Verwendung von Schimpfwörtern (genrespezifisch, aber auch politisch + rassistisch)
- ▶ Autom. Extraktion von Kandidaten mittels syntaktischer Pattern
- ▶ Beispielpattern: Du [NN] , Du [ADJ]* [NN]
- ▶ Manuelle Bereinigung der extrahierten Daten und Auswahl von eindeutigen Schimpfwörtern

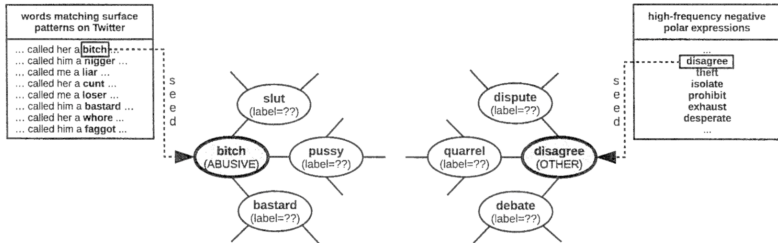
Auszug aus Korpus

► here

Graph-basierter Ansatz für autom. Erweiterung des Baselexikons

- ▶ Erstellung von pos + neg seed-Liste mit annotierten Schimpfwörtern aus Baselexikon (+) und häufigsten Wörtern (-)
- ▶ Graph mit Kanten zw. Wörtern auf Basis von Kosinusähnlichkeit zwischen Word Embeddings Vektoren (auf Twitter Korpus trainiert)
- ▶ Propagierung der Seed-Labels auf ungelabelte Knoten/Wörter mittels graphbasiertem Label-Propagation-Algorithmus (Adsorption Algorithmus, Talukdar et al. 2008)

Graph-basierter Ansatz für autom. Erweiterung des Baselexikons



Evaluation

- ▶ Baseline 1: Unigram und Bigram SVM:
- ▶ Baseline 2: Feature Selection (Mutual Information) SVM
- ▶ Allg. Preprocessing: Autosarkasmus-SP (SoSe 2016): Alle Tweets wurden tokenisiert, normalisiert und pos-getagged. Zusätzlich Lemmatisierung und stopword removal.
- ▶ 10-Fold Cross Validation mit kontrollierbarem random-seed für bessere Vergleichbarkeit.

Evaluation

- ▶ Unigram und Bigram SVM: Es wurde ein SVM Klassifizierer mit Standard-Parametern (Regularisierungs-Parameter $C = 1.0$, linear Kernel) trainiert. Input-Features sind die auf der Dokument-Term Matrix berechneten tf-idf Werte für Uni- bzw. Bigramme.
- ▶ Feature Selection Algorithmus: Berechnung des mutual information score (Manning et. al 20??) zwischen Label und Wort \Rightarrow 1500 Wörter mit den höchsten mi-scores wurden als Input-Features für den SVM Klassifizierer benutzt ($C=1.0$, linear-kernel).

Evaluation

Table: Baseline: tf-idf unigram SVM

[illegible]

Evaluation

Table: Baseline: tf-idf bigram SVM

[illegible]

Evaluation

Table: Baseline: Feature-Selection m. Mutual Information

[illegible]

Modul		Uli	Fabian
Erstellung Baselexikon	Integration SentiWS		x
	Extraktion neg. Wörter in SentiWS		x
	PAT-basierte Extraktion von NNs/ADJs	x	
	Skript für Erstellung eines Annotations-Testsets	x	
Erstellung Baselines	Implementation manuelle 10-Fold Cross Validation	x	
	Integration Autosarkasmus-Tweet Preprocessing	x	
	Erweitertes Preprocessing (Lemmatisierung, Stopwörter)	x	
	Implementation Unigram/Bigram SVM Baseline	x	
	Implementation Mutual Information (Manning et. al 2011)	x	
	Implementation Feature Selection (MI) SVM Baseline	x	
	Evaluation und Output	x	
Erstellung Word-Similarity Graph			
	Word Embeddings auf Twitter Daten	x	x
	Erstellung des Wortähnlichkeitsgraphen auf Basis von Kosinusähnlichkeiten	x	x
	Unknown Words Handhabung (character-level embeddings?)	x	x
	Erstellung der Seed Listen (pos + neg)	x	x
Label Propagation			
	Implementation Adsorption Algorithmus (Talukdar 2008)	x	x
	Erweiterung des Baselexikons mit Output	x	x
Anwendung und Evaluation			
	Test auf Germeval Daten	x	x
	Verbesserungen und Erweiterungen	x	x
	Visualisierung des Outputs	x	x
	Präsentation der Ergebnisse	x	x
	Abschlussbericht	x	x

Aufgabenverteilung

Zeitplan

Programmarchitektur

Datenstrukturen

Literatur