



Maschinelle Sprachverarbeitung

Übung

Aufgabe 4: Evaluationsergebnisse

Mario Sängner

mario.saenger@informatik.hu-berlin.de

Übung 4

- Implementierung eines Klassifikationsprogramms zur automatischen Analyse von Filmbewertungen
 - Bereitstellung eines Datensatzes mit Trainingsbeispielen
 - Auswahl einer beliebigen Klassifikationsmethode
 - Durchführung einer 10-fach Kreuzvalidierung
- Implementierung in Python oder Java möglich
 - Breites Spektrum an Lösungsmöglichkeiten
- Wettbewerb: Die Lösung mit der höchsten Genauigkeit gewinnt!

Lösungsübersicht

Gruppe	Ansatz		
	Sprache	Framework	Verfahren
AS_JG	Python	Keras	Neuronales Netzwerk
Batorurowaiaru	Python	Scikit-Learn	Logistic Regression
Bitdiggers	Python	Scikit-Learn	SVM
Dennis	Python	Scikit-Learn	Logistic Regression
DieBesteGruppe	Java	Eigene Impl.	K-Nearest-Neighbor
LidlEffort	Python	Scikit-Learn	Ada Boost Classifier
NotWorkingAsIntended	Python	Keras	RNN (Bi-GRU)
Pi	Python	Scikit-Learn	SVM
TraubeMinze&EinTee	Python	Eigene Impl.	Regression
WWGDS	Python	Scikit-Learn	SVM

Evaluationsergebnisse

Gruppe	Ansatz	10-KV	Test	Rang
AS_JG	NN	72.55	85.72	8
Batorurowaiaru	LogReg	88.87	88.34	3
Bitdiggers	SVM	90.90	90.24	2
Dennis	LogReg	87.00	88.21	4
DieBesteGruppe	KNN	73.21	68.15	9
LidlEffort	AdaBoost	88.44	86.94	7
NotWorkingAsIntended	Bi-GRU	89.04	87.50	6
Pi	SVM	91.15	90.44	1
TraubeMinze&EinTee	Regression	63.30	62.78	10
WWGDS	SVM	82.00	87.64	5

Evaluationsergebnisse

Gruppe	Ansatz	10-KV	Test	Rang
AS_JG	NN	72.55	85.72	8
Batorurowaiaru	LogReg	88.87	88.34	3
Bitdiggers	SVM	90.90	90.24	2
Dennis	LogReg	87.00	88.21	4
DieBesteGruppe	KNN	73.21	68.15	9
LidlEffort	AdaBoost	88.44	86.94	7
NotWorkingAsIntended	Bi-GRU	89.04	87.50	6
Pi	SVM	91.15	90.44	1
TraubeMinze&EinTee	Regression	63.30	62.78	10
WWGDS	SVM	82.00	87.64	5

Variation der Trainingsdatengröße

Gruppe	Ansatz	100%	25%	10%
AS_JG	NN	85.72	83.89 (-2.14%)	77.28 (-9.85%)
Batorurowaiaru	LogReg	88.34	86.63 (-1.93%)	85.02 (-3.76%)
Bitdiggers	SVM	90.24	88.19 (-2.27%)	85.74 (-4.99%)
Dennis	LogReg	88.21	85.93 (-2.58%)	82.87 (-6.05%)
DieBesteGruppe	KNN	68.15	67.10 (-1.54%)	66.35 (-2.64%)
LidlEffort	AdaBoost	86.94	85.38 (-1.79%)	82.85 (-4.70%)
NotWorkingAsInt	Bi-GRU	87.50	84.77 (-3.12%)	69.96 (-20.04%)
Pi	SVM	90.44	88.57 (-2.07%)	85.90 (-5.03%)
TraubeMinzeTee	Regression	62.78	69.00 (+9.91%)	70.13 (+11.72%)
WWGDS	SVM	87.64	86.44 (-1.37%)	84.70 (-3.36%)

State-of-the-Art (SOTA)

- SOTA-Klassifikationsverfahren: <https://nlpprogress.com>

Model	Accuracy	Paper / Source
ULMFIT (Howard and Ruder, 2018)	95.4	Universal Language Model Fine-tuning for Text Classification
Block-sparse LSTM (Gray et al., 2017)	94.99	GPU Kernels for Block-Sparse Weights
oh-LSTM (Johnson and Zhang, 2016)	94.1	Supervised and Semi-Supervised Text Categorization using LSTM for Region Embeddings
Virtual adversarial training (Miyato et al., 2016)	94.1	Adversarial Training Methods for Semi-Supervised Text Classification
BCN+Char+CoVe (McCann et al., 2017)	91.8	Learned in Translation: Contextualized Word Vectors

Road-Map für die letzten Wochen

- Do, 24.01.2019
 - Abgabe 5. Übungsaufgabe
 - Ausgabe 6. Übungsaufgabe
- Do, 31.01.2019
 - Besprechung der 5. Übungsaufgabe
- Do, 07.02.2019
 - Optionaler Q&A-Termin
- So, 10.02.2019
 - Abgabe 6. Übungsaufgabe
- Do, 14.02.2019
 - Besprechung 6. Übungsaufgabe
 - Abschluss, Feedback, Siegerehrung