

Diskussion Recherche / Zwischenstand Projekte

HS Experimentelles Arbeiten in der Sprachverarbeitung

Nils Reiter

`nils.reiter@uni-koeln.de`

December 8, 2022

Recherche Forschungsstand “Fake News Detection”

- ▶ Drei Dimensionen
 - ▶ Eigener Workflow
 - ▶ Stil der Papiere
 - ▶ Inhalt der Papiere

Recherche Forschungsstand “Fake News Detection”

- ▶ Drei Dimensionen
 - ▶ Eigener Workflow
 - ▶ Stil der Papiere
 - ▶ Inhalt der Papiere
- ▶ Was ist Ihnen aufgefallen?
- ▶ Was hat Sie überrascht?
- ▶ Was fanden Sie leicht/schwer?

Patronizing and condescending language (PCL)

1. Preprocessing

2. Train / test-Split

3. Modell-Definition (Layer)

4. Evaluation

Features

Humor and offensiveness

- binary: Humor / NichtHumor

1. Preprocessing

2. Split

3. Tokenization

4. Vektorizing / Embedding ^{Bag of words}

5. LogReg.

6. Evaluati-

Historical NER

- Wortweise Klassifikation 1. NE? 2. Weder A

Wort	
Petr	B
Milch	1
ist	0
ein	0
Glase	0
.	0

Features

Textklassifikation

- Bag of words
 - Vektoring
- } =

Wortklassifikation / Sequence Labeling

- Worttestwörter

Peter => 7
Michael => 3
ist => 225
:
:

	the	buy	
-	{ the: 5, buy: 1, ... }		
-	{ the: 3, buy: 5, ... }		

ist	7	3	⊕	100D Embeddy
en	3	225		

How To NLP Experiment – in Practice

1. Get to know the phenomenon
 - ▶ Theory: Read literature about it
 - ▶ Practice: Annotate it (e.g., on paper!)

How To NLP Experiment – in Practice

1. Get to know the phenomenon
 - ▶ Theory: Read literature about it
 - ▶ Practice: Annotate it (e.g., on paper!)
 2. Establish gold standard
-

How To NLP Experiment – in Practice

1. Get to know the phenomenon
 - ▶ Theory: Read literature about it
 - ▶ Practice: Annotate it (e.g., on paper!)
2. Establish gold standard
3. Establish experimental setup
 - ▶ Data structures to represent prediction result
 - ▶ Write a dummy model as a start
 - ▶ Write evaluation code and test it on the dummy model

How To NLP Experiment – in Practice

1. Get to know the phenomenon
 - ▶ Theory: Read literature about it
 - ▶ Practice: Annotate it (e.g., on paper!)
2. Establish gold standard
3. Establish experimental setup
 - ▶ Data structures to represent prediction result
 - ▶ Write a dummy model as a start
 - ▶ Write evaluation code and test it on the dummy model
4. Establish a baseline and evaluate it
 - ▶ May require statistical data analysis

How To NLP Experiment – in Practice

1. Get to know the phenomenon
 - ▶ Theory: Read literature about it
 - ▶ Practice: Annotate it (e.g., on paper!)
2. Establish gold standard
3. Establish experimental setup
 - ▶ Data structures to represent prediction result
 - ▶ Write a dummy model as a start
 - ▶ Write evaluation code and test it on the dummy model
4. Establish a baseline and evaluate it
 - ▶ May require statistical data analysis
5. The core: Train/develop your own model for prediction
 - ▶ Evaluate and analyse errors made by own model
 - ▶ Improve model and repeat

How To NLP Experiment – in Practice

1. Get to know the phenomenon
 - ▶ Theory: Read literature about it
 - ▶ Practice: Annotate it (e.g., on paper!)
2. Establish gold standard
3. Establish experimental setup
 - ▶ Data structures to represent prediction result
 - ▶ Write a dummy model as a start
 - ▶ Write evaluation code and test it on the dummy model
4. Establish a baseline and evaluate it
 - ▶ May require statistical data analysis
5. The core: Train/develop your own model for prediction
 - ▶ Evaluate and analyse errors made by own model
 - ▶ Improve model and repeat
6. Document findings and publish them