

Semantic Argument Classification

28. Januar 2015

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Institut für Computerlinguistik
Universität Heidelberg



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386



Gliederung

Semantic Argument
Classification

Julian Baumann, Kevin
Decker, Maximilian
Müller-Eberstein

Problemstellung
Anwendungsbereich

Daten
Problemstellung

Umsetzung
Features
Featureextraktion
Schwierigkeiten

Evaluation

Ausblick

Literatur

Referenzen

Problemstellung
Anwendungsbereich

Daten
Problemstellung

Umsetzung
Features
Featureextraktion
Schwierigkeiten

Evaluation

Ausblick

Literatur



Semantic Argument Classification

Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellung

2

Anwendungsbereich

Daten

Problemstellung

Umsetzung

Features

Featureextraktion

Schwierigkeiten

Evaluation

Ausblick

Literatur

Referenzen

Was ist Semantic Argument Classification?

- ▶ Zuweisung bestimmter Rollen in einem Satz ⇒ „Wer tut wem was an?“
- ▶ It operates stores mostly in Iowa and Nebraska
- ▶ [Arg0 *It*][Pred *operates*][Arg1 *stores*][ArgLoc *mostly in Iowa and Nebraska*]



Daten

Semantic Argument Classification

Julian Baumann, Kevin
Decker, Maximilian
Müller-Eberstein

Problemstellung
Anwendungsbereich

Daten

Problemstellung

Umsetzung

Features
Featureextraktion
Schwierigkeiten

Evaluation

Ausblick

Literatur

Referenzen

3

- ▶ NLTK
- ▶ PropBank



Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellung

Anwendungsbereich

Daten

4

Problemstellung

Umsetzung

Features

Featureextraktion

Schwierigkeiten

Evaluation

Ausblick

Literatur

Referenzen

- ▶ versucht generalisierte Argumente zu verwenden → Parser
- ▶ Argumentrollen sind für jedes Verb in Frames organisiert → weniger spezifisch
- ▶ ARG0 = Proto-Agent
- ▶ ARG1 = Proto-Patient
- ▶ ARG2-ARG5 = Argumente mit steigender Intensität



Penn Treebank

Semantic Argument Classification

Julian Baumann, Kevin
Decker, Maximilian
Müller-Eberstein

Problemstellung

Anwendungsbereich

Daten

5

Problemstellung

Umsetzung

Features

Featureextraktion

Schwierigkeiten

Evaluation

Ausblick

Literatur

Referenzen

- ▶ Subkorpus aus WSJ und Brown Corpus, bestehend aus ungefähr 3.500.000 Wörtern
- ▶ 112.917 Sätze annotiert nach PropBank-Annotationsschema
- ▶ 292.975 Instanzen
- ▶ wsj/00/wsj_0001.mrg 1 10 gold publish.01 p—a 10:0-rel 11:0-ARG0



Penn Treebank

Semantic Argument Classification

Julian Baumann, Kevin
Decker, Maximilian
Müller-Eberstein

Problemstellung

Anwendungsbereich

Daten

Problemstellung

Umsetzung

Features

Featureextraktion

Schwierigkeiten

Evaluation

Ausblick

Literatur

Referenzen

6

```
((S
  (NP-MNR-SBJ
    (NP (DT The) (NN way) )
    (SBAR
      (WHADVP-1 (-NONE- 0) )
      (S
        (NP-SBJ (NNP MacArthur) )
        (VP (VBD said)
          (NP (PRP his) (NN line) )
          (ADVP-MNR (-NONE- *T*-1) )))))
  (: - -)
```

15



Features

Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellung

Anwendungsbereich

Daten

Problemstellung

Umsetzung

Features

Featureextraktion

Schwierigkeiten

Evaluation

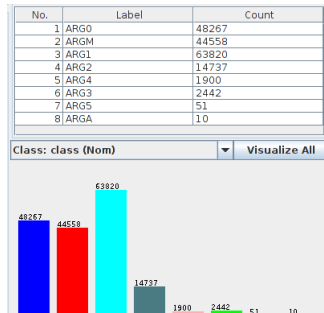
Ausblick

Literatur

Referenzen

- ▶ Predicate
- ▶ Path
- ▶ Phrase Type
- ▶ Position
- ▶ Voice

7



15



Predicate

Semantic Argument Classification

Julian Baumann, Kevin
Decker, Maximilian
Müller-Eberstein

Problemstellung

Anwendungsbereich

Daten

Problemstellung

Umsetzung

Features

Featureextraktion

Schwierigkeiten

Evaluation

Ausblick

Literatur

Referenzen

8

- ▶ lemmatisierte Prädikat
- ▶ 3966 distinct

15



Path

Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellung
Anwendungsbereich

Daten
Problemstellung

Umsetzung

Features
Featureextraktion
Schwierigkeiten

Evaluation

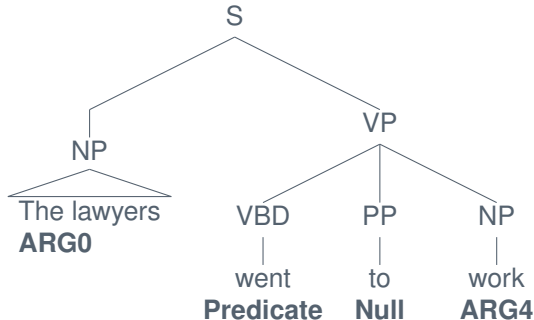
Ausblick

Literatur

Referenzen

9

- ▶ beschreibt Pfad zwischen ARG und Predicate
- ▶ vereinfacht z.B. NP-SBJ → NP
- ▶ extrahiert über Lowest Common Ancestor
- ▶ beispielsweise: NP↑S↓VP↓VBD
- ▶ 41737 distinct



15



Phrase Type

Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellung

Anwendungsbereich

Daten

Problemstellung

Umsetzung

Features

Featureextraktion

Schwierigkeiten

Evaluation

Ausblick

Literatur

Referenzen

- beschreibt die Kategorie des Argument
- z.B: NP, MD, PP, SBAR
- 65 distinct

No.	Label	Count
1	NP	97599
2	MD	6458
3	PP	29042
4	NN	1311
5	ADVP	10080
6	S	9957
7	-NONE-	1596
8	VBG	141
9	SBAR	10444

Class: class (Nom) Visualize All

10

15



Position

Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellung

Anwendungsbereich

Daten

Problemstellung

Umsetzung

Features

Featureextraktion

Schwierigkeiten

Evaluation

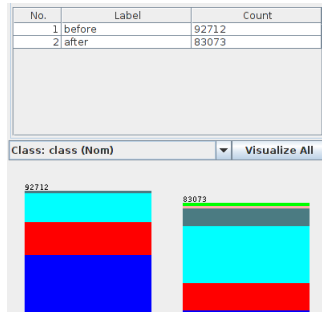
Ausblick

Literatur

Referenzen

11

- Beschreibt, ob das Argument vor oder nach dem Prädikat steht
- berechnet mithilfe von WordNum
- 2 distinct



15



Voice

Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellung

Anwendungsbereich

Daten

Problemstellung

Umsetzung

Features

Featureextraktion

Schwierigkeiten

Evaluation

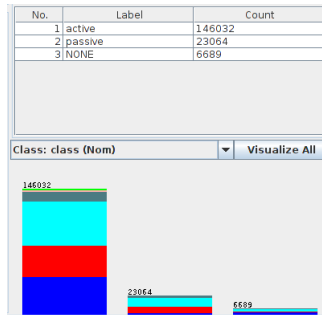
Ausblick

Literatur

Referenzen

12

- ▶ gibt an, ob das Prädikat aktiv oder passiv ist
- ▶ größtenteils annotiert
- ▶ 3 distinct: active, passive, unknown



15



Featureextraktion

Semantic Argument Classification

Julian Baumann, Kevin
Decker, Maximilian
Müller-Eberstein

Problemstellung

Anwendungsbereich

Daten

Problemstellung

Umsetzung

Features

Featureextraktion

Schwierigkeiten

Evaluation

Ausblick

Literatur

Referenzen

13

```
featureList = [...] # zu extrahierende Features
for pbInstance in pbInstances :
    for pbArg in pbInstance.arguments :
        features = []
        for feature in featureList :
            featureList.append(extFeature(feature, pbArg,
pbInstance))
# write features to file in ARFF
```

15



ARFF

Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellung
Anwendungsbereich

Daten
Problemstellung

Umsetzung
Features
Featureextraktion
Schwierigkeiten

Evaluation

Ausblick

Literatur

Referenzen

14

@relation SAC_All

@attribute predicate {join,publish,name,use, make, cause, ...}

@attribute phraseType {NP, MD, PP, NN, ADVP, S, ...}

@attribute position {before, after}

@attribute path {NP^S!VP!VP, MD^VP^S!VP!VP,...}

@attribute voice {active, passive, NONE}

@attribute class {ARG0, ARGM, ARGa, ARG1, ...}

@data

join, NP, before, NP^S!VP!VP, active, ARG0

join, MD, before, MD^VP^S!VP!VP, active, ARGM

join, NP, after, NP^VP^VP^S!VP!VP, active, ARG1

join, PP, after, PP^VP^VP^S!VP!VP, active, ARGM

join, NP, after, VP^VP^S!VP!VP, active, ARGM

15



Quellen

Semantic Argument Classification

Julian Baumann, Kevin
Decker, Maximilian
Müller-Eberstein

Problemstellung
Anwendungsbereich

Daten
Problemstellung

Umsetzung
Features
Featureextraktion
Schwierigkeiten

Evaluation

Ausblick

Literatur

Referenzen

15

- [1] Omri Abend und Roi Reichart. *Unsupervised Argument Identification for Semantic Role Labeling*.
- [2] Jean Carletta. "Assessing agreement on classification tasks: the kappa statistic". In: *Computational Linguistics* (1996), S. 249–254.
- [3] Daniel Gildea. "Automatic labeling of semantic roles". In: *Computational Linguistics* 28 (2002), S. 245–288.
- [4] Alessandro Moschitti und Cosmin Adrian Bejan. "A Semantic Kernel for Predicate Argument Classification". In: *IN CONLL 2004*. 2004, S. 17–24.
- [5] Sameer Pradhan u. a. *Support Vector Learning for Semantic Argument Classification*. 2005.

15

Vielen Dank für Eure Aufmerksamkeit!
Noch Fragen?



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386