

# 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?“



# 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



# 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

3

- ▶ NLTK
- ▶ PropBank

## Umsetzung

Features

Featureextraktion

Schwierigkeiten

## Evaluation

## Ausblick

## Literatur

## Referenzen

15



## Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

## Problemstellung

Anwendungsbereich

## Daten

Problemstellung

## Umsetzung

Features

Featureextraktion

Schwierigkeiten

## Evaluation

## Ausblick

## Literatur

## Referenzen

4

- ▶ 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 3.478.101 Wörtern
- ▶ 112.917 Sätze annotiert nach PropBank-Annotationsschema
- ▶ 292.975 Instanzen
- ▶ `wsj/00/wsjo001.mrg110goldpublish.01p` – – – `a10 : 0 – re/11 : 0 – ARG0`





# Penn Treebank

## Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

## Problemstellung

Anwendungsbereich

## Daten

6

## Problemstellung

## Umsetzung

Features

Featureextraktion

Schwierigkeiten

## Evaluation

## Ausblick

## Literatur

## Referenzen

```
( (S
  (NP-SBJ
    (NP (NNP Pierre) (NNP Vinken) )
    (, ,)
    (ADJP
      (NP (CD 61) (NNS years) )
      (JJ old) )
    (, ,) )
  (VP (MD will)
    (VP (VB join)
      (NP (DT the) (NN board) )
      (PP-CLR (IN as)
        (NP (DT a) (JJ nonexecutive) (NN director) ))
      (NP-TMP (NNP Nov.) (CD 29) )))
  (, .) ))
( (S
  (NP-SBJ (NNP Mr.) (NNP Vinken) )
  (VP (VBZ is)
    (NP-PRD
      (NP (NN chairman) )
      (PP (IN of)
        (NP
          (NP (NNP Elsevier) (NNP N.V.) )
          (, ,)
          (NP (DT the) (NNP Dutch) (VBG publishing) (NN group) )))))
  (, .) ))
```



# 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
- ▶ 197 unique
- ▶ 3966 distinct

15



# Path

## Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

## Problemstellung

Anwendungsbereich

## Daten

Problemstellung

## Umsetzung

### Features

Featureextraktion

Schwierigkeiten

## Evaluation

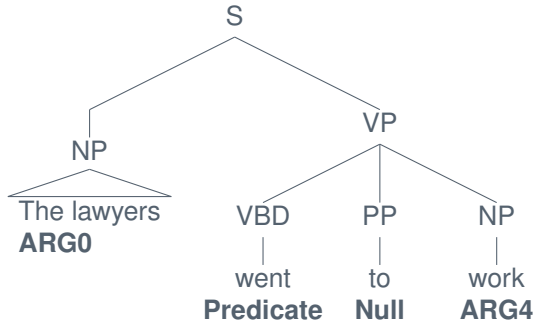
## Ausblick

## Literatur

## Referenzen

9

- ▶ beschreibt Pfad zwischen ARG und Predicate
- ▶ vereinfacht z.B. NP-SBJ  $\rightarrow$  NP
- ▶ extrahiert über Lowest Common Ancestor
- ▶ beispielsweise: NP $\uparrow$ S $\downarrow$ VP $\downarrow$ VBD
- ▶ 41737 distinct





# Phrase Type

## Semantic Argument Classification

Julian Baumann, Kevin  
Decker, Maximilian  
Müller-Eberstein

### Problemstellung

Anwendungsbereich

### Daten

Problemstellung

### Umsetzung

Features

Featureextraktion

Schwierigkeiten

### Evaluation

### Ausblick

### Literatur

### Referenzen

10

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

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 extract.py





Julian Baumann, Kevin  
Decker, Maximilian  
Müller-Eberstein

## Problemstellung

### Anwendungsbereich

## Daten

## Umsetzung

## Features

## Featureextraktion

## Schwierigkeiten

## Evaluation

## Ausblick

## Literatur

## Referenzen

```
@relation SAC_All
```

```
@attribute predicate {join,publish,name,use,make,cause,expose,report,enter,show,say,stop,appear,bring,talk,hear,have,study,I
@attribute phraseType {NP,MD,PP,NN,ADVP,S,-NONE-,VBG,SBAR,NNB,RB,WHADVP,CC,VBP,UCP,WHNP,NNP,VP,NP=2,'',JJ,ADJP,RBR,CD,VBD,VE
@attribute position {before,after}
@attribute path {NP^S^VP^VP,MD^VP^S^VP^VP,NP^VP^VP^S^VP^VP,PP^VP^VP^S^VP^VP,NN^NP^NP^PP^NP^VP^S^VP^NP^PP^NP^NP,NP^S^VP^VP^S
@attribute voice {active,passive,NONE}
@attribute class {ARG0,ARGM,ARG1,ARG2,ARG3,ARG5,ARGA}
```

data  
join, NP, before, NP^S!VP!VP, active, ARG0  
join, MD, before, MD^VP^S!VP!VP, active, ARG0  
join, NP, after, NP^VP^VP^S!VP!VP, active, ARG1  
join, PP, after, PP^VP^VP^S!VP!VP, active, ARG0  
join, NP, after, NP^VP^VP^S!VP!VP, active, ARG0  
publish, NN, after, NN^NP^NP^PP^NP^VP^S!VP!NP!PP!NP!NP, active, ARG0  
name, NP, before, NP^S!VP!VP, passive, ARG1  
name, NP, after, NP^S^VP^VP^S!VP!VP, passive, ARG2  
use, ADVP, before, ADVP^RRC^NP^S^S!S!NP!RRC!VP, passive, ARG0  
use, NP, before, NP^NP^S^S!S!NP!RRC!VP, passive, ARG1  
use, S, after, S^VP^RRC^NP^S^S!S!NP!RRC!VP, passive, ARG2  
make, -NONE-, before, -NONE-^NP^S^VP^RRC^NP^S^S!S!NP!RRC!VP!S!VP!VP, active, ARG0  
make, NP, after, NP^VP^VP^S^VP^RRC^NP^S^S!S!NP!RRC!VP!S!VP!VP, active, ARG1  
cause, NP, before, NP^S^S!S!VP!VP, active, ARG0  
cause, NP, after, NP^VP^VP^S^S!S!VP!VP, active, ARG1  
expose, PP, after, PP^VP^RRC^NP^PP^NP^PP^NP^VP^S^S!S!VP!NP!PP!NP!PP!NP!RRC!VP, passive, ARG2  
expose, ADVP, after, ADVP^VP^RRC^NP^PP^NP^PP^NP^VP^S^S!S!VP!NP!PP!NP!PP!NP!RRC!VP, passive, ARG0  
expose, NP, before, NP^NP^PP^NP^VP^VP^S^S!S!VP!VP!NP!PP!NP!PP!NP!RRC!VP, passive, ARG1  
report, NP, before, NP^S!VP, active, ARG0  
report, S, before, S^S!VP, active, ARG1  
enter, NP, before, NP^S^SBAR^VP^S^S!S!VP!SBAR!S!VP, active, ARG0  
enter, NP, after, NP^VP^S^SBAR^VP^S^S!S!VP!SBAR!S!VP, active, ARG1  
cause, NP, before, NP^S^PP^VP^S^S!S!VP!PP!S!VP, active, ARG0  
cause, NP, after, NP^VP^S^PP^VP^S^S!S!VP!PP!S!VP, active, ARG1  
show, NP, before, NP^NP^VP^S^PP^VP^S^S!S!VP!PP!S!VP!NP!SBAR!S!VP, active, ARG1  
show, ADVP, after, ADVP^VP^S^SBAR^NP^VP^S^PP^VP^S^S!S!VP!PP!S!VP!NP!SBAR!S!VP, active, ARG0  
say, NP, before, NP^S!VP, active, ARG0  
say, S, before, S^S!VP, active, ARG1  
make, NP, after, NP^VP^S^SBAR^NP^NP^S!NP!NP!SBAR!S!VP, active, ARG1  
make, NP, before, NP^NP^NP^S!NP!NP!SBAR!S!VP, active, ARG0  
stop, NP, before, NP^S!VP, active, ARG0  
stop, VBG, after, VBG^VP^VP^S!VP, active, ARG1



# 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