# Semantic Argument Classification

28. Januar 2015

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Institut für Computerlinguistik Universität Heidelberg





# Gliederung

### Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

1 TODICITISTOI

Daten & Tools

Umsetzun

Features

Featureextraktion

Schwierigkeite

Setup

Aushlick

Literatur

Referenze

Problemstellung

Daten & Tools

Umsetzung

Features

Featureextraktion

Schwierigkeiten

Experimente

Setup

Evaluation

Ausblick

Literatur



# Semantic Argument Classification

### Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

# Problemstellung Daten & Tools

\_ ....

Umsetzung

Features

Schwierigkeite

Experime

Evaluation

Ausblic

Literatur

Heterenzen

### 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 lt][Pred operates][Arg1 stores][ArgLoc mostly in lowa and Nebraska]

Institut für Computerlinguistik Ruprecht-Karls-Universität Heidelberg



### Daten & Tools

### Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

# Problemstellung Daten & Tools

### . . . .

Umsetzung

Egaturaaytraktir

Cabusianialarita

Schwierigkeite

Lxbeiii

Evaluation

Ausblick

Literatur

Referenzer

- ▶ Python3.4
- ► NLTK 3.0
- ► PropBank
- ▶ PennTreeBank
- ▶ Weka 3.7.12

Institut für Computerlinguistik Ruprecht-Karls-Universität Heidelberg



# PropBank

### Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellur

Daten & Tools

### . .

Umsetzung

Featureeytrakti

Schwierigkeiter

Experiment

Setup

Aushlick

Literatur

Referenze

- ▶ Argumentrollen sind für jedes Verb in Frames organisiert → weniger spezifisch

ARG0	proto-agent
ARG1	proto-patient
ARG2	instrument, benefactive, attribute
ARG3	starting point, benefactive, attribute
ARG4	ending point
ARGM	modifier



# PropBank

### Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

## Problemstellur Daten & Tools

### . . .

Umsetzung

Featureextraktion

Schwierigkeiter

Evporimente

Setup

Evaluatio

Ausblick

Literatur

Referenzer

- ► [ARG0 She][Predicate writes][ARG1 a program]
- ► [ARG0 She][Predicate writes][ARG2 about headbands]
- ► [argo She][Predicate writes][arg1 a program][arg3 for BApple<sup>TM</sup>]
- ► [ARGM-TMP Now][ARG0 she][Predicate writes][ARG1 a program]
- ▶ → verschieden Rollen

Institut für Computerlinguistik Ruprecht-Karls-Universität Heidelberg



### Penn Treebank

### Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellur

Daten & Tools

Umsetzung

Eastureaytrakti

Schwierigkeiter

Experiment

Evaluation

Ausblick

Literatur

 Subkorpus aus WSJ, bestehend aus ungefähr 1 Millionen Tokens

- ▶ 112.917 Prädikat-Argument Strukturen 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



# Klassenverteilung

### Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellung

### Daten & Tools

Umsetzung

Features

Schwierinkeiten

Experimen

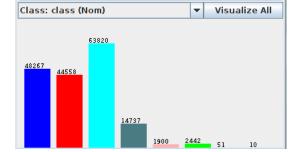
Setup

Ausblick

Literatur

Referenzer

Label	Count
ARG0	48267
ARGM	44558
ARG1	63820
ARG2	14737
ARG4	1900
ARG3	2442
ARG5	51
ARGA	10
	ARG0 ARGM ARG1 ARG2 ARG4 ARG3 ARG5





### Penn Treebank

### Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellu

### Daten & Tools

Umsetzung

Featureautraldi

reatureextrakt

Schwierigkeite

Setup

Evaluatio

Ausblick

Literatur

Referenzer

Institut für Computerlinguistik Ruprecht-Karls-Universität Heidelberg



### Features

### Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

### Problemstellung

Daten & Tools

#### OIIIOOtEt

### Features

Featureextrakti

Schwierigkeiter

Experimen

Control

Ausblick

Literatur

Referenzen

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



# Predicate

### Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellung

Daten & Tools

Umsetzuna

#### Features

Featureextraktio

....

Experimente

Evaluation

Ausblick

Literatur

Referenzer

► lemmatisierte Prädikat

▶ 3966 distinct feature values

Institut für Computerlinguistik Ruprecht-Karls-Universität Heidelberg



### Path

### Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

### Problemstell

Daten & Tools

#### Umsetzt

Features Featureextrakti

Schwierigkeiten

Experimente

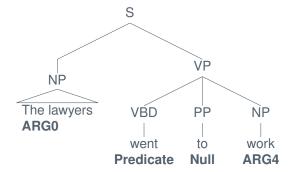
Evaluation

Ausblick

Literatur

Referenzer

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





# Phrase Type

### Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellun

Daten & Tools

Umsetzung

Features

Schwierigkeiter

Evporimont

Setup

Evaluation

Ausblick

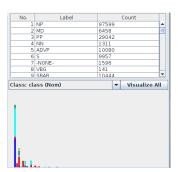
Literatur

Referenzer

beschreibt die Kategorie des Argument

► z.B: NP, MD, PP, SBAR

▶ 65 distinct feature values





### **Position**

### Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

### Problemstellung

Daten & Tools

#### Umaatauna

Features

### Schwieriekeite

Experimente

### Setup

Evaluatio

### Ausblick

Literatur

Referenzer

- Beschreibt, ob das Argument vor oder nach dem Prädikat steht
- ▶ Binäres Feature

No.	Label	Count	
	before	92712	
2	after after	83073	
Class: cl	ass (Nom)	▼ Visualize	ΑI
01ass: cl	lass (Nom)	▼ Visualize A	ΔI
	ass (Nom)	11344112	Al
	ass (Nom)	11344112	41
	ass (Nom)	11344112	Al
	iass (Nom)	11344112	AI
	lass (Nom)	11344112	All



### Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

### Problemstellun

Daten & Tools

#### Umcotzuna

Features

Featureextraktic

Schwierigkeiten

Experiment

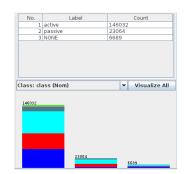
Setup

Ausblick

Literatur

Heierenzen

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





### Featureextraktion

Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellur

Daten & Tools

Umsetzun

Features

Featureextraktion

Schwieriakeite

Experimente

Setup

Ausblick

Literatur

Referenzer

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

Institut für Computerlinguistik Ruprecht-Karls-Universität Heidelberg



### Featureextraktion

### Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellun

Daten & Tools

Umsetzung

Eastures

Featureextraktion

Schwieriakeite

Experimente

Setup

Ausblick

Literatur

Referenzer

wsj/00/wsj\_0041.mrg 38 14 gold talk.01 vn-3a 0:1-ARGM-ADV 12:1-ARG0 14:0-rel 15:1-ARG1-about (S I(PP-LOC (IN Against) (NP (DT a) (NN shot)) ARGInstanceBuilder (PP (IN of) (NP (NNP Monticello))) (VBN superimposed) (NP (-NONE- \*)) (PP-CLR (IN on) (NP (DT an) (JJ American) (NN flag)))))) (NP-SBJ (DT an) (NN announcer)) (VBZ talks) (PP-CLR (IN about) (NP (DT the) ('' '') (JJ strong) (NN tradition)) ARGInstance processed features ARFFDocument attributes, data

Institut für Computerlinguistik Ruprecht-Karls-Universität Heidelberg



### Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellur

Daten & Tools

Umsetzung

Features Featureextraktion

Schwierigkeiter

- Conwierigkeite

Setup

Evaluatio

Ausblic

Literatur

Referenze

@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

Institut für Computerlinguistik Ruprecht-Karls-Universität Heidelberg



# Schwierigkeiten

### Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

### Problemstellt

Daten & Tools

#### Umsetzun

Features

Featureextrak

### Schwierigkeiten

Setup

Evaluation

Aushlick

Literatur

Referenzei

- ► PropBankChain- und PropBankSplitTreePointer
- Verwendung einer externen PennTreeBank
- ▶ einige Feature (bsp. path) nehmen sehr viele Werte an

Institut für Computerlinguistik Ruprecht-Karls-Universität Heidelberg



### Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

### Problemstellui

Daten & Tools

### Umsetzun

Featureextraktion

Calanteextraktio

Schwierigkeiter

### Setup

Evaluation

#### Ausblick

Literatur

Referenze

► 60% train, 20% dev, 20% test

▶ Baseline: ZeroR

► Naive Bayes, j48 tree, (libSVM)

bisher: Training auf train, Evaluierung mit dev

Institut für Computerlinguistik Ruprecht-Karls-Universität Heidelberg



# Ergebnisse

Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellung

Daten & Tools

Umsetzung

OmootEan

Featureextrakti

Schwierigkeite

Experimente

. . .

Evaluation

Ausblick

Literatur

Referenzen

	Precision	Recall	F-Measure
Baseline	0.132	0.364	0.194
Naive Bayes	0.771	0.778	0.770
j48 Tree	0.784	0.786	0.781



# Confusion Matrix (Naive Bayes)

Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellung

Daten & Tools

Features

Schwierigkeiten

Schwierigkeiter Experimente

Satur

Evaluation

Ausblick

Literatur

Referenzen

а	b	С	d	е	f	g	h	<- classified as
14497	348	361	272	0	4	0	1	a = ARG0
291	11394	2143	1009	189	48	0	1	b = ARGM
3119	707	17064	375	31	27	0	0	c = ARG1
180	792	1854	2163	29	23	0	0	d = ARG2
2	217	23	141	379	3	0	0	e = ARG3
37	289	144	147	170	99	0	0	f = ARG4
0	13	0	1	0	0	3	0	g = ARG5
5	0	0	0	0	0	0	0	h = ARGA
			1	1		'		

Institut für Computerlinguistik Ruprecht-Karls-Universität Heidelberg



# Feature Evaluation

Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellung

Daten & Tools

Umsetzung

-----

Featureextraktion

Schwierigkeiter

Experimente

Evaluation

Ausblick

Ausbiii

Literatur

Referenzen

	Precision	Recall	F-Measure	F-Measure Change
All Features	0.771	0.778	0.770	0
-voice	0.748	0.754	0.745	-0.025
-path	0.778	0.783	0.776	+0.006
-phraseType	0.735	0.747	0.733	-0.037
-position	0.758	0.773	0.757	-0.013
-predicate	0.717	0.732	0.716	-0.54
				•

Institut für Computerlinguistik Ruprecht-Karls-Universität Heidelberg



### **Ausblick**

### Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

### Problemstellui

Daten & Tools

#### Umsetzun

Featureextrakti

Schwierigkeiter

Evporimente

Setup

Evaluation

### Ausblick

Literatur

Referenze

- ► Path Feature überarbeiten
- ▶ HeadWord Feature implementieren
- genauere Evaluation
- ► SVM?
- ► Abschlussbericht schreiben

Institut für Computerlinguistik Ruprecht-Karls-Universität Heidelberg



### Quellen

### Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellur

Daten & Tools

Umsetzun

Featureextrakti

Schwierigkeiter

Setup

Ausblic

Literatur

Referenzen

- [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.

### Vielen Dank für Eure Aufmerksamkeit! Noch Fragen?

