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

Problemstellu

Problemstellur

1 TODIETTISTETIO

Features

Features

Schwieriekeite

ochwierigkeiti

Experimente

Ausblick

Litorate

Referenzer

Problemstellung

Daten

Problemstellung

Umsetzung

Features

Featureextraktion

Schwierigkeiten

Experimente

Ausblick

Literatur

Institut für Computerlinguistik Ruprecht-Karls-Universität Heidelberg



Semantic Argument Classification

Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellung

Problemetellu

Problemstellur

Umsetzung

Featureovtral

Schwierigkeite

Experimente

Aushlick

Litoroti

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 /t][Pred operates][Arg1 stores][ArgLoc mostly in lowa and Nebraska]

Institut für Computerlinguistik Ruprecht-Karls-Universität Heidelberg



Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellung

Daten

Problemstellun

. . .

OTTIOOTEGI

Featureeytra

Schwieriakeite

Experimente

Ausblick

Literatur

Referenzer

► NLTK

► PropBank

Institut für Computerlinguistik Ruprecht-Karls-Universität Heidelberg



PropBank

Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellui

Daten

Problemstelli

Umsetzung

F-----

Featureextral

Schwierigkei

Experimente

Ausblick

Literatu

Referenzer

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

Institut für Computerlinguistik Ruprecht-Karls-Universität Heidelberg



PropBank

Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Daten

Experimente

Aushlick

► [ARG0 He][Predicate wrote][ARG1 a book]

- ► [ARG0 He][Predicate wrote][ARG2 about them]
- ► [ARG0 He][Predicate wrote][ARG3 a book for his children]
- ▶ → verschieden Rollen

Institut für Computerlinguistik Ruprecht-Karls-Universität Heidelbera



Penn Treebank

Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellu

Daten

Problemstellur

Ulasa atau asa

Features

Featureextrak

Schwierigkeite

Experimente

Ausblick

Literatu

Referenzer

- Subkorpus aus WSJ und Brown Corpus, bestehend aus ungefähr 1.000.000 Wörtern
- ► 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



Penn Treebank

Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellu

Daten

1 100/01/10/01

Umsetzung

F-----

Featureextrak

Schwierigkeit

Experimente

Aushlick

Literatu

Referenzen

Institut für Computerlinguistik Ruprecht-Karls-Universität Heidelberg



Features

Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellung

Broblomotollun

Features

Easturgeytraktion

Schwierigkeiten

Experimen

Ausblick

Literatur

neierenzen

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

No.	Labe	Count
1	ARG0	48267
	ARGM	44558
	ARG1	63820
	ARG2	14737
5	ARG4	1900
6	ARG3	2442
7	ARG5	51
8	ARGA	10
iass: Cl	ass (Nom)	▼ Visualize A
iass: Ci	ass (Nom)	▼ Visualize A



Predicate

Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellung

Daten

Problemstellur

Umsetzung

Features

Featureextraktion

Schwierigkeiten

Experimente

Ausblick

Literatur

Referenzer

► lemmatisierte Prädikat

▶ 3966 distinct

Institut für Computerlinguistik Ruprecht-Karls-Universität Heidelberg



Path

Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellun

Daten

FIODIEITISTEII

Umsetzung

Features

Featureextrakt

Schwierigkeite

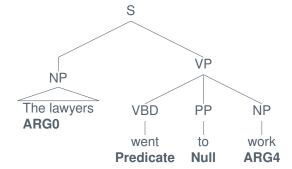
Experimente

Aushlick

Literatu

. .

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



Institut für Computerlinguistik Ruprecht-Karls-Universität Heidelberg



Phrase Type

Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellu

Broblometellun

Froblemstellum

Umsetzun

Features

O-busin-i-luita-

ochwierigkeiten

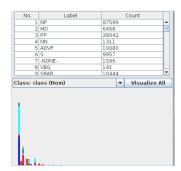
Experimente

Ausblick

Literatu

Referenzen

- 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

Doton

Problemstellun

Umsetzuno

Features

-eatureextraktion

Ochwierigkeiten

Experimente

Ausblick

Literatur

Referenzen

 Beschreibt, ob das Argument vor oder nach dem Prädikat steht

▶ Binäres Feature

No.	Label		Count
	before	92712	
2	after	83073	
	ass (Nom)	~	Visualize All
92712	ass (Nom)		Visualize All
	ass (Nom)	83073	Visualize All
	ass (Nom)		Visualize All
	ass (Nom)		Visualize All
	ass (Nom)		Visualize All
	ass (Nom)		Visualize All
	ass (Nom)		Visualize All



Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellun

Problemstellun

1 TODIOTTIOLOIIGI

Features

Featureextraktion

Schwierigkeiten

Experimente

Ausblick

Literatu

Referenzen

gibt an, ob das Prädikat aktiv oder passiv ist

- ▶ größtenteils annotiert
- ▶ 3 distinct feature values: active, passive, unknown

No.	Label	Count
1 ac	tive	146032
	ssive	23064
3 NO	NE	6689
Class: class	(Nom)	▼ Visualize A
146032		
	23064	6689



Featureextraktion

Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellur

Daten

Problemstell

Umsetzun

Features

Featureextraktion

Experimente

Ausblick

Literatu

neierenzen

```
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

FIODIEITIST

Deter

Problemstellu

O and a second

Eastures

Featureextraktion

Schwierigkeite

Experimente

Ausblick

Literatu

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

Problemstellui

Daten

Problemstellui

Umsetzung

Feature

Featureextraktion

Experimente

Aushlick

7 10000110

Litorate

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

Problemstellur

Daten

Problemstellu

. .

Easture

Easturaaytral

Schwierigkeiten

Experimente

Aushlick

I Second

Referenzer

- PropBankChain- und PropBankSplitTreePointer
- Verwendung einer externen PennTreeBank
- ▶ einige Feature (bsp. path) nehmen sehr viele Werte an
- Speicherbedarf der Algorithmen J48 und SVM

Institut für Computerlinguistik Ruprecht-Karls-Universität Heidelberg



Setup

Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellur

Daten

Problemstellur

Umsetzung

Omsetzui

reatures

Featureextrak

Schwierigkeit

Experimente

Aushlick

Literatu

Referenzer

► 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

Problemstellun

1 100/01/10/01/01/01

Continue

Featureextraktion

Schwierigkeiter

Experimente

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



Feature Evaluation

Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Experimente

Aushlick

а	b	С	d	е	f	g	h	

 \cap

f = ARG4q = ARG5h = ARGA

<- classifie

a = ARG0

b = ARGM

c = ARG1

d = ARG2

e = ARG3

Institut für Computerlinguistik Ruprecht-Karls-Universität Heidelbera



Confusion Matrix (Naive Bayes)

Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellung

Daten

Problemstellung

.....

Features

reatureextrai

Schwierigkeite

Experimente

Ausblick

Literatur

Heterenzen

		Precision	Recall	F-Measure	F-Measure Ch
١.		1 1603011	riccan	1 -IVICASUIC	1 -Ivicasure Off
	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
		1			!

Institut für Computerlinguistik Ruprecht-Karls-Universität Heidelberg



Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

Problemstellur

Daten

Problemstellu

Uma a stance a

Feature

Featureeytrak

Schwierinkeite

Experimente

Ausblick

_iteratu

Referenzer

► Path Feature überarbeiten

- ► HeadWord Feature implementieren
- genauere Evaluation
- ► SVM?
- ► Abschlussbericht schreiben



Quellen

Semantic Argument Classification

Julian Baumann, Kevin Decker, Maximilian Müller-Eberstein

i iobiemstenu

Problemstellu

...

Features

Featureextraktion Schwierigkeiten

Experimente

Ausblick

Literatu

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?

