

Semantische Klassifikation von Adverbien

TaCoS 2012 in Trier

Joachim Bingel

Universität Heidelberg
Institut für Computerlinguistik

2. Juni 2012

Überblick

- 1 Einleitung
- 2 Adverbien
- 3 Datensatz
- 4 Klassifizierung
- 5 Ergebnisse
- 6 Abschließend

Gegenstand der Arbeit

- Semantische Klassifikation von Adverbien

$$c : A \rightarrow C$$

A: Menge von Adverbien

C: Menge von semantischen Klassen

- Auf Typen- und Tokenlevel

Motivation

- Information Extraction: *wann, wo, wie* geschah etwas?
- *John [adverb] walked to the store*: Typ des Adverbs entscheidet beim Parsing über dessen syntaktische Position (Conlon und Evens 1992)
- Ressource für generelle semantische Sprachverarbeitung

Forschungsfragen

Wie gut können die folgenden Ansätze das Problem lösen?

1. Klassifizierung mit Distributioneller Semantik (auf Typenlevel)
2. Syntaxbasierte Klassifizierung (auf Tokenlevel)

Grundlegende Idee

1. Klassifizierung mit Distributioneller Semantik

- Distributionelle Hypothese: (Semantisch) ähnliche Wörter treten in ähnlichen Kontexten auf (Turney und Pantel 2010)
- ⇒ Adverbien der gleichen semantischen Klasse in ähnlichen Kontexten
- ⇒ Klassifizierung mittels Vektorraummodell

2. Syntaxbasierte Klassifizierung

- Verschiedene semantische Klassen weisen unterschiedliche syntaktische Eigenschaften auf
- ⇒ Syntaktische Eigenschaften von Adverbien im Satz geben Aufschluss über Klassenzugehörigkeit
- ⇒ Klassifizierung mittels maschinellem Lernverfahren mit syntaktischen Eigenschaften als Features

Grundlegende Idee

1. Klassifizierung mit Distributioneller Semantik

- Distributionelle Hypothese: (Semantisch) ähnliche Wörter treten in ähnlichen Kontexten auf (Turney und Pantel 2010)
- ⇒ Adverbien der gleichen semantischen Klasse in ähnlichen Kontexten
- ⇒ Klassifizierung mittels Vektorraummodell

2. Syntaxbasierte Klassifizierung

- Verschiedene semantische Klassen weisen unterschiedliche syntaktische Eigenschaften auf
- ⇒ Syntaktische Eigenschaften von Adverbien im Satz geben Aufschluss über Klassenzugehörigkeit
- ⇒ Klassifizierung mittels maschinellem Lernverfahren mit syntaktischen Eigenschaften als Features

Grundlegende Idee

1. Klassifizierung mit Distributioneller Semantik

- Distributionelle Hypothese: (Semantisch) ähnliche Wörter treten in ähnlichen Kontexten auf (Turney und Pantel 2010)
- ⇒ Adverbien der gleichen semantischen Klasse in ähnlichen Kontexten
- ⇒ Klassifizierung mittels Vektorraummodell

2. Syntaxbasierte Klassifizierung

- Verschiedene semantische Klassen weisen unterschiedliche syntaktische Eigenschaften auf
- ⇒ Syntaktische Eigenschaften von Adverbien im Satz geben Aufschluss über Klassenzugehörigkeit
- ⇒ Klassifizierung mittels maschinellem Lernverfahren mit syntaktischen Eigenschaften als Features

Grundlegende Idee

1. Klassifizierung mit Distributioneller Semantik

- Distributionelle Hypothese: (Semantisch) ähnliche Wörter treten in ähnlichen Kontexten auf (Turney und Pantel 2010)
- ⇒ Adverbien der gleichen semantischen Klasse in ähnlichen Kontexten
- ⇒ Klassifizierung mittels Vektorraummodell

2. Syntaxbasierte Klassifizierung

- Verschiedene semantische Klassen weisen unterschiedliche syntaktische Eigenschaften auf
- ⇒ Syntaktische Eigenschaften von Adverbien im Satz geben Aufschluss über Klassenzugehörigkeit
- ⇒ Klassifizierung mittels maschinellem Lernverfahren mit syntaktischen Eigenschaften als Features

Grundlegende Idee

1. Klassifizierung mit Distributioneller Semantik

- Distributionelle Hypothese: (Semantisch) ähnliche Wörter treten in ähnlichen Kontexten auf (Turney und Pantel 2010)
- ⇒ Adverbien der gleichen semantischen Klasse in ähnlichen Kontexten
- ⇒ Klassifizierung mittels Vektorraummodell

2. Syntaxbasierte Klassifizierung

- Verschiedene semantische Klassen weisen unterschiedliche syntaktische Eigenschaften auf
- ⇒ Syntaktische Eigenschaften von Adverbien im Satz geben Aufschluss über Klassenzugehörigkeit
- ⇒ Klassifizierung mittels maschinellem Lernverfahren mit syntaktischen Eigenschaften als Features

Grundlegende Idee

1. Klassifizierung mit Distributioneller Semantik

- Distributionelle Hypothese: (Semantisch) ähnliche Wörter treten in ähnlichen Kontexten auf (Turney und Pantel 2010)
- ⇒ Adverbien der gleichen semantischen Klasse in ähnlichen Kontexten
- ⇒ Klassifizierung mittels Vektorraummodell

2. Syntaxbasierte Klassifizierung

- Verschiedene semantische Klassen weisen unterschiedliche syntaktische Eigenschaften auf
- ⇒ Syntaktische Eigenschaften von Adverbien im Satz geben Aufschluss über Klassenzugehörigkeit
- ⇒ Klassifizierung mittels maschinellem Lernverfahren mit syntaktischen Eigenschaften als Features

Adverbien

Adverbien

Hochgradig heterogene Wortart:

- “variety of syntactic and semantic roles” (Jackendoff 1972)
- “catch-all term” (ibid.)
- “[in traditional grammar] the category to which words are assigned if they do not satisfy the more specific criteria for nouns, verbs, adjectives, prepositions and conjunctions” (Huddleston and Pullum 2002)
- “enjoy considerable mobility within the clause, [...] syntactically if not semantically optional” (Hoye 1997)
- “items are sometimes assigned for apparently no better reason than that they appear to relate less to any other class” (ibid.)
- “nobody seems to know exactly what to do with adverbs” (Ernst 2002)

Entscheidend: Adverbien erfüllen die Funktionen von *Adverbialen*, d.h. sie tragen Bedeutung bzgl. Zeit, Raum, Art und Weise etc.

Semantische Klassen

- Kein allgemein akzeptiertes Set an semantischen Klassen
- “Kernklassen”, die alle traditionellen Ansätze anerkennen (Haumann, 2007):
 - Art und Weise
 - Zeit
 - Raum
- Außerdem finden sich in der Literatur *Domänenadverbien*, *subjekt-orientierte* und *sprecher-orientierte* Adverbien, *konjunktive* Adverbien, *quantifizierende* und *evaluative* Adverbien, usw...

Klassenset für diese Arbeit

1. SPACE
2. TIME
3. MANNER
4. DOMAIN
5. DEGREE
6. SUBJECT-ORIENTED
7. SPEAKER-ORIENTED
8. DISCOURSE

Dazu feiner unterscheidende Subklassen

Ambige Adverbien

Was sind die semantischen Funktionen von z.B. *dreadfully* und *cleverly*?

Ambige Adverbien

She behaved dreadfully₁. Now she is dreadfully₂ sorry.

- *dreadfully₁* beschreibt die Art und Weise (*Manner*)
- *dreadfully₂* gibt einen Grad an (*Degree*)

John cleverly₁ answered their questions cleverly₂/stupidly.

- *cleverly₁* ist "subjektbezogen" (*Subject-Oriented*)
- *cleverly₂/stupidly* beschreibt die Art und Weise (*Manner*)

⇒ Häufige Ambiguität bezüglich Klassenzugehörigkeit auf Typenlevel

⇒ Ambiguität kann auf Tokenlevel resolviert werden, z.B. durch die syntaktische Position

Ambige Adverbien

She behaved dreadfully₁. Now she is dreadfully₂ sorry.

- *dreadfully₁* beschreibt die Art und Weise (*Manner*)
- *dreadfully₂* gibt einen Grad an (*Degree*)

John cleverly₁ answered their questions cleverly₂/stupidly.

- *cleverly₁* ist "subjektbezogen" (*Subject-Oriented*)
- *cleverly₂/stupidly* beschreibt die Art und Weise (*Manner*)

⇒ Häufige Ambiguität bezüglich Klassenzugehörigkeit auf Typenlevel

⇒ Ambiguität kann auf Tokenlevel resolviert werden, z.B. durch die syntaktische Position

Ambige Adverbien

She behaved dreadfully₁. Now she is dreadfully₂ sorry.

- *dreadfully₁* beschreibt die Art und Weise (*Manner*)
- *dreadfully₂* gibt einen Grad an (*Degree*)

John cleverly₁ answered their questions cleverly₂/stupidly.

- *cleverly₁* ist "subjektbezogen" (*Subject-Oriented*)
- *cleverly₂/stupidly* beschreibt die Art und Weise (*Manner*)

⇒ Häufige Ambiguität bezüglich Klassenzugehörigkeit auf Typenlevel

⇒ Ambiguität kann auf Tokenlevel resolviert werden, z.B. durch die syntaktische Position

Ambige Adverbien

She behaved dreadfully₁. Now she is dreadfully₂ sorry.

- *dreadfully₁* beschreibt die Art und Weise (*Manner*)
- *dreadfully₂* gibt einen Grad an (*Degree*)

John cleverly₁ answered their questions cleverly₂/stupidly.

- *cleverly₁* ist "subjektbezogen" (*Subject-Oriented*)
- *cleverly₂/stupidly* beschreibt die Art und Weise (*Manner*)

⇒ Häufige Ambiguität bezüglich Klassenzugehörigkeit auf Typenlevel

⇒ Ambiguität kann auf Tokenlevel resolviert werden, z.B. durch die syntaktische Position

Ambige Adverbien

She behaved dreadfully₁. Now she is dreadfully₂ sorry.

- *dreadfully₁* beschreibt die Art und Weise (*Manner*)
- *dreadfully₂* gibt einen Grad an (*Degree*)

John cleverly₁ answered their questions cleverly₂/stupidly.

- *cleverly₁* ist "subjektbezogen" (*Subject-Oriented*)
- *cleverly₂/stupidly* beschreibt die Art und Weise (*Manner*)

⇒ Häufige Ambiguität bezüglich Klassenzugehörigkeit auf Typenlevel

⇒ Ambiguität kann auf Tokenlevel resolviert werden, z.B. durch die syntaktische Position

Ambige Adverbien

She behaved dreadfully₁. Now she is dreadfully₂ sorry.

- *dreadfully₁* beschreibt die Art und Weise (*Manner*)
- *dreadfully₂* gibt einen Grad an (*Degree*)

John cleverly₁ answered their questions cleverly₂/stupidly.

- *cleverly₁* ist "subjektbezogen" (*Subject-Oriented*)
- *cleverly₂/stupidly* beschreibt die Art und Weise (*Manner*)

- ⇒ Häufige Ambiguität bezüglich Klassenzugehörigkeit auf Typenlevel
- ⇒ Ambiguität kann auf Tokenlevel resolviert werden, z.B. durch die syntaktische Position

Syntaktische Eigenschaften

- Große Vielfalt syntaktischer Rollen von Adverbien
 - Adverbien können im Satz fast überall auftreten
 - Aber: semantische Funktion teilweise abhängig von syntaktischer Position
- ⇒ Syntaktische Eigenschaften als generelle Indikatoren für Klassenzugehörigkeit, z.B.:
- Gradadverbien: meist Modifikation und direkte Präzedenz von NPs und ADVPs
 - Satzmodifikation (Sprecherorientiert) vs. VP-Modifikation (Art und Weise)
 - Lineare Position im Satz bezüglich *core constituents* (Jacobson 1964, Huang 1975)
 - Position in Tiefenstruktur (Cinque 1999)

Datensatz

Erstellen des Datensatzes

- Datengetriebene Klassifizierung als Lernproblem: annotierte Trainingsdaten benötigt
- Anforderungen an Daten und Korpora
 - Syntaxbasierter Ansatz benötigt syntaktisch annotierte Korpusdaten
 - Features umfassen Phrasenstruktur- und Dependenzinformationen
 - Jeder Satz muss in beiden Annotationen vorliegen
 - Phrasenstruktur aus Penn Treebank
 - Dependenzbäume extrahiert mit Penncoverter
 - Penn Treebank (Brown + WSJ): 56.152 Adverbtoken und 2.080 Typen (~ 27 Token/Typ)

Datenbeispiel

Originale Phrasenstruktur:

```
( (S
  (PP (TO To)
    (NP (PRP her) ))
  (NP-SBJ (NNS peanuts)
    (CC and)
    (NNS emeralds) )
  (VP (MD would)
    (VP (VB have)
      (VP (VBN been)
        (NP-PRD (RB just)
          (ADJP (RB so) (JJ much) )
          (NN blubber) ))))
  (. .) ))
```

Datenbeispiel

Abgeleiteter Dependenzbaum:

1	To	-	TO	-	-	6	ADV	-	-
2	her	-	PRP	-	-	1	PMOD	-	-
3	peanuts	-	NNS	-	-	6	SBJ	-	-
4	and	-	CC	-	-	3	COORD	-	-
5	emeralds	-	NNS	-	-	4	CONJ	-	-
6	would	-	MD	-	-	0	ROOT	-	-
7	have	-	VB	-	-	6	VC	-	-
8	been	-	VCN	-	-	7	VC	-	-
9	just	-	RB	-	-	12	NMOD	-	-
10	so	-	RB	-	-	11	AMOD	-	-
11	much	-	JJ	-	-	12	NMOD	-	-
12	blubber	-	NN	-	-	8	PRD	-	-
13	.	-	.	-	-	6	P	-	-

Annotation

- Bedingungen
 - Adverbtypen für Training durch min. 30 Token repräsentiert
 - Ausreichende Repräsentation aller Klassen
- 6 Annotatoren (Studenten der Computerlinguistik/Anglistik)
- Zuweisung der richtigen Klasse im Kontext (Tokenlevel) für je 30 Kontexte von 241 Typen in interaktivem Programm mit Korrekturmöglichkeit
- Ein Adverbtyp gehört einer Klasse an, wenn in mehr als 50% der Kontexte das entsprechende Klassenlabel vergeben wurde

Ergebnis der Annotation

- Doppelte Annotation von 50 Typen: IAA 69,8%, Kappa 0.617
 - 224 von 241 Adverbtypen verwendbar
 - Gleichmäßige Verteilung von ambigen Adverbtypen über Training-, Development- und Testset
 - Typ ist ambig gdw. nicht immer das gleiche Label vergeben wurde
 - Heuristische Akquise von zusätzlichen Kontexten für klar nicht-ambige Adverbtypen
- ⇒ Insgesamt 13.960 Tokens ($\sim 62,3$ Tokens/Type)

Verteilung der Adverbtypen über die Klassen

	Train _{70%}	Dev _{15%}	Test _{15%}	Σ	ambig
SPACE	23	5	5	33	39%
TIME	26	6	6	38	21%
MANNER	24	6	6	36	50%
DOMAIN	9	3	3	15	26%
DEGREE	35	7	7	49	53%
SUBJ-OR	6	2	2	10	90%
SPEAKER-OR	20	5	5	30	53%
DISCOURSE	9	2	2	13	31%
Σ	152	36	36	224	42%

Klassifizierung

VSMs

- Häufig benutzt um semantische Ähnlichkeit von Wörter, Dokumenten etc. zu messen (Turney und Pantel 2010)
- Erstellen eines Vektor von Kookkurrenz-Frequenzen für jedes Konzept
 - Dimensionen des Vektors entsprechen kookkurrierenden Wörtern
 - Menge der Vektoren bildet *semantischen Raum* in Form einer *Matrix*
- Zwei Konzepte sind sich ähnlich gdw. ihre Vektoren ähnlich sind
- Ähnlichkeitsmaß: z.B. Kosinus

Klassifizierung mit VSM

- Erstellung von Vektorraummodellen mit DependencyVectors
- Erlaubt Extraktion von Targets über bestimmte Pfade in einem Dependenzbaum (Padó 2007)
- VSMs erstellt jeweils auf Penn Treebank und BNC
- Adverbien als Targets, 10.000 Dimensionen

Klassifizierung mit VSM

Klassifizierung:

1. Berechne Zentroidenvektor \vec{c} für die Trainingstypen jeder Klasse c
2. Vergleiche Vektoren zu klassifizierender Typen mit allen \vec{c}
 - 2a. One-Vs-All. Trifft nacheinander binäre Entscheidungen: c oder $\neg c$
→ 0 bis $|C|$ mögliche Ergebnisse
 - 2b. Multiclass. Berechnet Ähnlichkeit für jedes c und erstellt Rangfolge über diese.
⇒ Beide Methoden geben Vektoren von Konfidenzwerten (für die Klassen) zurück. Auswahl der informativen Klassen durch Entropy Selection (Hartung und Frank 2010).

Syntaxbasierte Klassifizierung

- Extraktion von 8 syntaktischen Features für jedes der 13.960 Token
 1. POS des Heads
 2. Lexikalische Form des Heads
 3. Dependenzrelation zum Head
 4. Distanz zum Head
 5. Distanz zum Root-Element
 6. Lineare Position im Satz nach Jacobsen (1964)
 7. Domäne (S, VP, NP, S-BAR,...)
 8. In ADVP (True/False)

just	NN	blubber	NMOD	-3	3	End	NP	False	DEGREE
------	----	---------	------	----	---	-----	----	-------	--------

Syntaxbasierte Klassifizierung

Klassifizierung

1. Trainiere Klassifikatoren (BayesNet, J48 Decision Tree) in WEKA
2. Berechne WSK für jedes zu klassifizierende Adverbtoken für jede Klasse und ermittle die wahrscheinlichste(n) Klasse(n)
 - 2a. One-Vs-All oder
 - 2b. Multiclass

Ergebnisse

Ergebnisse - Distributioneller Ansatz

- Bestes Modell auf Penn Treebank mit One-Vs-All Klassifizierung
- Korrekte Klassifizierung von 69,4% der Testtypen

	SPACE	TIME	MANNER	DOMAIN	DEGREE	SUBJ-OR	SPEAKER-OR	DISCOURSE
Recall	1	0.83	0.67	0	0.86	0	0.60	0
Precision	1	0.56	1	0	0.60	0	0.43	0
f-Measure	1	0.67	0.8	0	0.71	0	0.50	0

Ergebnisse - Syntaxbasierter Ansatz

- Bestes Modell mit BayesNet und Multiclass-Klassifizierung
- Korrekte Klassifizierung von 68,2% der Testinstanzen

	SPACE	TIME	MANNER	DOMAIN	DEGREE	SUBJ-OR	SPEAKER-OR	DISCOURSE
Recall	0.696	0.831	0.512	0.024	0.677	0	0.260	0.155
Precision	0.386	0.857	0.557	0.111	0.647	0	0.319	0.127
f-Measure	0.496	0.844	0.534	0.040	0.662	0	0.287	0.140

Problematik der Ergebnisse

- Spärliche Klassenstärke bei der typbasierten Klassifizierung führt zu weniger aussagekräftigen Evaluationsergebnissen
- Unbalancierter Datensatz
- Teilweise schwierig separierbare Klassen
- Umgang mit Ambiguität bei der Evaluation

Abschließend

- Kombiniertes Ansatz brachte keine Verbesserung
- Adverbklassifikation ist zu einem gewissen Grade möglich
- Komplette Lösung des Problems scheint nicht erwartbar
- Distributionelles Modell mit leicht besserem Ergebnis, operiert aber nur auf Typenlevel, wenige Evaluationsinstanzen
- Syntaktisches Modell (auf Tokenlevel) auch mit gutem Ergebnis, aber abhängig von bestimmtem Korpusformat
- Größerer und balancierterer Datensatz kann eventuell zu besseren Ergebnissen führen

Referenzen

- Cinque, Guglielmo (1999): *Adverbs and Functional Heads*. Oxford University Press.
- Conlon, S. P.-N. and Evens, M. (1992). *Can computers handle adverbs? In Proceedings of the 14th International Conference on Computational Linguistics (COLING)*, Nantes, France.
- Ernst, Thomas (2002): *The Syntax of Adjuncts*. Cambridge University Press.
- Haumann, Dagmar (2007): *Adverb Licensing and Clause Structure in English*. Benjamins Publ., Amsterdam.
- Hartung, M. and Frank, A. (2010). *A Structured Vector Space Model for Hidden Attribute Meaning in Adjective-Noun Phrases. In Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC)*, Valletta, Malta.
- Hoyer, Leo (1997): *Adverbs and Modality in English*. Longman.
- Huang, Shuan-Fan (1975): *A Study of Adverbs*. Mouton, The Hague.
- Huddleston, R. and Pullum, G. (2002): *The Cambridge Grammar of The English Language*. Cambridge University Press.
- Jackendoff, Ray (1972): *Semantic Interpretation in Generative Grammar*. MIT Press.
- Jacobson, Sven (1964): *Adverbial Positions in English*. Studentbok, Stockholm.
- Padó, Sebastian and Lapata, Mirella (2007): *Dependency-Based Construction of Semantic Space Models. In Computational Linguistics 33*, pp. 161–199.
- Turney, Peter and Pantel, Patrick (2010): *From Frequency to Meaning: Vector Space Models of Semantics. In Journal of Artificial Intelligence Research 37*, pp. 141–188.