



# Algorithmus zur Ontologisierung Semantischer Relationen

Both/Lohse/Weidler

Formale Semantik, WiSe 14/15, A. Frank

April 8, 2015

- 1 Ziel
- 2 Vorgehen
- 3 Ressourcen
- 4 Programme
  - Preprocessing
  - Main Algorithm
- 5 Evaluation
- 6 Erkenntnisse/Probleme
- 7 Literatur

Ziel: Automatische Auswahl der wahrscheinlichsten Senses für  
Verb-Relations-Paare



# Vorgehen

- 1 Ermittlung des Goldstandards
- 2 Korpusbasierte WSD:
  - Korpusassoziationsmaß
  - Lesk
  - Relation
- 3 Featurevektoren für alle möglichen Paare der Trainingsdaten ermitteln
- 4 Automatische Klassifikation durch Weka
- 5 Ermittlung der wahrscheinlichsten Sensepaare für die Testdaten
- 6 Evaluation: Vergleich mit dem Goldstandard

# Ressourcen

- WordNet
- Semantic Relations Data Set (Tremper/Frank)
- SemCor
- OntoNotes
- Weka



# Programm: Preprocessing

- Gold Standard Annotator
- Gold Standard Generator
- Semcor Verb Cooccurrence Extractor
- OntoNotes Verb Cooccurrence Extractor Cooccurrence Extraction Merger



# Programm: Main Algorithm

- **Verb Ontologization Class**  
Berechnet die Featurevektoren und baut aus Goldstandard ein File mit Trainingsvektoren
- **Weka Arff Builder**  
Schreibt die Trainingsvektoren in ein von Weka lesbares File
- **ARFF Data Multiplier**  
Passt die Proportion bzgl der Datenpunkte in + und - Klasse an



# Programm: Main Algorithm

- **Sense Chooser**

Berechnet die wahrscheinlichsten Sensepaare für ein gegebenes Verbpaar

- **Evaluator**

Berechnet Evaluationswerte für die im Sensechooser verwandten Klassifikationen durch abgleich mit Goldstandard





# Evaluation

**Bestimmter Kongruenzwert:** 7.6923076923076925

**Anteil einzelner richtig gewählter Senses:**

- Congruence: 0.20512820512820512
- Completely correctly predicted Pairs: 3
- Partially correctly predicted Pairs: 10



# Erkenntnisse und Probleme

- Verschiedene mögliche Sensepaare für ein Verbrelationspaar
- Maße/Features nicht ausreichend/ergiebig
- zu wenig Daten aus Korpora?
- Sinnhaftigkeit der + Klassen Vervielfachung
- Clustering > Classification



# Literatur

- [http://www.cl.uni-heidelberg.de/publications/papers/tremper\\_frank.pdf](http://www.cl.uni-heidelberg.de/publications/papers/tremper_frank.pdf)
  - Korpusassoziation, Klassifikation, Semantic Relations Data Set
- Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.
  - Weka-Benutzung
- Daniel Jurafsky James H. Martin (2008); Speech and Language Processing; Prentice Hall, Auflage: 2nd edition.
  - Lesk, WSD, Kontextassoziation
- Semantikfolien 15-20