Learning and Predicting Verb Argument Binding

software project winter term 2013/14

Damian Gorski, Angelika Kirilin,
Elisa Starke, Mengfei Zhou
Department of Computational Linguistics
Heidelberg University, Germany

Motivation

- coreferring arguments in pairs of co-occurring verbs can be arbitrarily linked:
  Mary$_i$ visited the new art gallery with Susann$_j$. She$_{i/j}$ loves art.
- meaning of the involved verbs can make one argument binding more probable/obligatory:
  Ed$_i$ shouted at Tim$_j$ because he$_j$ crashed the car.
  Ed$_i$ shouted at Tim$_j$ because he$_i$ was angry. (Rahman & Ng 2012)
- analysis of semantic properties of verbs reveal patterns
  → Certain verbs can stand in a particular (semantic) relation
  → How can we squeeze profit from these insights?

Goals

- Are there predictable **argument binding (coreference) patterns**?
- What are the possibly involved factors?
- Does big data help finding reliable patterns?
- Can we reliably predict such patterns on type-/token-level?

Corpora

I. **OntoNotes-4.0** (CoNLL 2011 subset)
- 10 Mbyte of text
- multi-layer annotation
- gold-standard annotation of i.a. coreference

II. **GigaWord 5th Edit.**
- ca. 26 Gbyte of Text
- automatic annotation (MATE parser)

Statistics

co-occurrences found within 2 sentences:

OntoNotes
- 11,703 (of 1,415 verb pairs)
- average co-occ. per verb pair: 32

GigaWord
- 42,857,885 (of 4,456 verb pairs)
- average co-occ. per verb pair: 4006

Feature Extraction

I. **pattern-based** (association measures):
coreferential probability
- in relation to all coref. patterns of verb x
  x = {v1,v2}; i,j = {subj, obj, inObj}:

$$\frac{C(coref(arg_{i(v1)}, arg_{j(v2)}))}{C(coref(arg_{i(x)}, *))}$$

- in relation to all patterns:

$$\frac{C(coref(arg_{i(v1)}, arg_{j(v2)}))}{C(coocc(arg_{i(v1)}, arg_{j(v2)}))}$$

II. **argument-based**:
e.g. *same word, string match, proper name*

III. **verb-based**:
e.g. *tempus, voice, verb distance*

Verb Pairs

I. **Tremper & Frank (2013)**
- verbs in **semantic relation**:
  entailment, presupposition, antonomy, temporal inclusion
- 200 pairs

II. **Chambers & Jurafsky (2009)**
- ‘narrative chains’
- 8,069 pairs

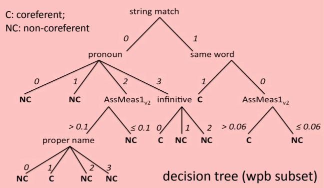
Argument Patterns

- dependency resolution for **subjects, dir. + ind. objects**
- coreference resolution (GW): extraction of **verb pairs with coreferring arguments** (= arg. pattern):
  I. **OntoNotes**: 2,603 patterns
  II. **GigaWord**: 3,354,693 patterns

Ed shouts at Tim. He overslept.
subj obj subj

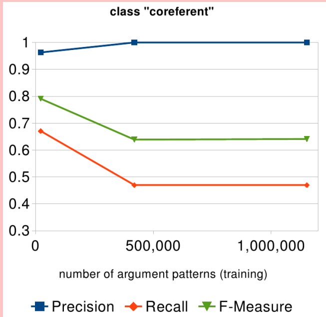
Classification (WEKA J48)

number of argument patterns	
training (GigaWord)	
WPB (1.3%)	43,726
+ XIN (25%)	838,750
+ AFP (68.6%)	2,303,020
testing	
OntoNotes	5,205



decision tree (wpb subset)

Results



more training data leads to:
- drop of learning curve → levels out at **0,64** (f-measure)
- **precision**: increase for coreferent and decrease for non-coreferent patterns; **recall**: vice versa
- smaller decision tree:
  less features: *asso1$_{v2}$, infinitive, pronouns* discarded
- **worst** classified argument pattern: subj$_{v1}$_subj$_{v2}$ (32.3% false)
- **best** classified argument pattern: inObj$_{v1}$_obj$_{v2}$ (9.8% false)

comparison of performances:

coreferent		
	SIEVE	presented
Precision	0.79	1
Recall	0.58	0.47

non-coreferent		
	SIEVE	presented
Precision	0.67	0.65
Recall	0.85	1

- similar alternation of performance with regard to classes
- though less class-dependent behavior
- generalization (SIEVE) vs. specialization (presented system)

Conclusion

- performance decreases on larger training sets
  - insufficient ability of features to generalize for coreference, though precise prediction:
  - hence better performance on rare patterns (i.e. inObj$_{v1}$_obj$_{v2}$) than on frequent (i.e. subj$_{v1}$_subj$_{v2}$)
  - thus overfitting on non-coreferent patterns, especially for cases
    1. in which one argument is not realized
       John eats a burger and __ drinks a coke.
    2. in which tokens of both arguments are identical
  - data is inconsistent/widespread: big data reduces determination of pattern strength and requires more robustness
- argument binding patterns can help improve coreference resolution systems (important features: include proper names, identical tokens and realization)
- to be examined:
  - suspicion of different performances of certain features (e.g. ‘ProperName’) at different argument patterns
  - precision recommends an implementation in elaborated coreference resolution systems
  - prior probability (resp. association) of coreferent argument bindings

Problems

- entire automatic annotation of GigaWord
- correct treatment of verbs containing prepositions/adverbs
- not realized arguments (evaluation): not recognized by SIEVE yet recognized by presented system
- gold-standard corefence information of OntoNotes: displaced word no.

References

- N. Chambers & D. Jurafsky (2008). Unsupervised learning of narrative event chains. In Proceedings of the ACL/HLT 2008 Conference.
- N. Chambers & D. Jurafsky (2009). Unsupervised learning of narrative schemas and their participants. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL.
- M. Gerber & J.Y. Chai (2012). Semantic Role Labeling of Implicit Arguments for Nominal Predicates. Computational Linguistics, 38 (1).
- A. Rahman & V. Ng (2012). Resolving Complex Cases of Definite Pronouns: The Winograd Schema Challenge. In Proceedings of EMNLP-CoNLL.
- G. Tremper & A. Frank (2013). A Discriminative Analysis of Fine-Grained Semantic Relations including Presupposition. Dialogue and Discourse 4 (2) 282-324.