Statistical Corpus-Based Word Sense Disambiguation: Pseudowords vs. Real Ambiguous Words

Tanja Gaustad
Alfa-Informatica
University of Groningen
The Netherlands
tanja@let.rug.nl
www.let.rug.nl/~tanja

ACL Student Research Workshop 2001

(General) Outline of the Problem

Word Sense Disambiguation (WSD): sense-tagged material needed for evaluation

No sense-tagged corpora for most languages

Solution: Use artificially created (evaluation) data

Question: Are the 'real' data and the artificial data *comparable*?

Overview

- Word Sense Disambiguation
- Pseudowords
- (Specific) Outline of the Problem
- Way of Proceeding
- Settings
  - Corpus
  - Classification Algorithm
  - Ambiguous Words / Pseudowords
- Results and Evaluation
- Conclusions and Future Work

Word Sense Disambiguation

Problem: Lexical semantic ambiguity

Goal: Recover correct sense (in given sentence/text)

Approach: Corpus-based
Means:

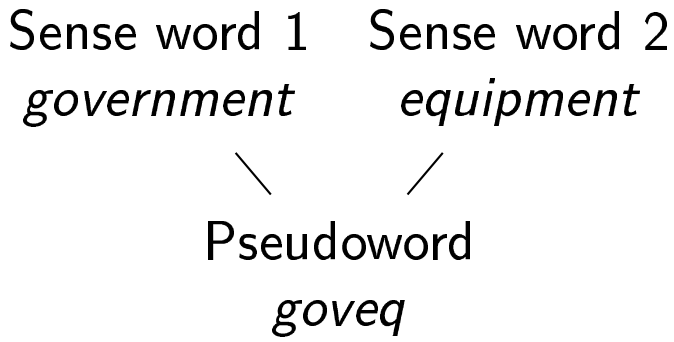
- Distributional information (frequency)
- Collocational information (context words)

Remaining Problem: No sense-tagged material
$\rightarrow$ Evaluation not possible

(Possible) Solution: Create evaluation data using pseudowords

Pseudowords

Technique: Introduce artificial ambiguity in corpus (Gale *et al.*, 1992; Schütze, 1992; 1998)



Example:

Original sentences:

(1) *Blah blahblahbblah blah* government *blah.*

(2) *Blahblah blahblah blahblahblah* equipment.

Ambiguous sentences:

(1) *Blah blahblahblah blah* goveq *blah.*

(2) *Blahblah blahblah blahblahblah* goveq.

Pseudowords (cont.)

Procedure:

1. Choose n sense words
2. Train algorithm on 80% of original (unambiguous) corpus
3. Create ambiguous test corpus (20%) by replacing sense words by pseudoword
4. Test on ambiguous corpus
5. Evaluate using original corpus

(Specific) Outline of the Problem

Fact 1: Evaluation needs sense-tagged material

Fact 2: Many languages without sense-tagged corpora

(Possible) Solution: Pseudowords

- Easily created
- Only raw text needed
- Any (supervised) algorithm can be tested

But:

Are the results on pseudowords comparable to results on real ambiguous words?

Way of Proceeding

Preliminary remarks:

Direct comparison of pseudowords and ambiguous words is not possible:

Ambiguous words can not be pseudowords
$\rightarrow$ no measure to express their differences

$\Downarrow$

Indirect comparison of results on same corpus with same algorithm

Way of Proceeding (cont.)

Indirect comparison:

1. Test real ambiguous words on SENSEVAL 1 corpus
2. Test pseudowords on SENSEVAL 1 corpus
3. Compare results

Iff results are similar,
WSD of pseudowords and WSD of
ambiguous words are comparable tasks.

Settings:
Corpus

SENSEVAL 1: Publicly available, widely used, (lexically) sense-tagged corpus

Training: Separate training file (30.000-70.000 words) per ambiguous word

Testing: Separate testing file (10.000-17.000 words) per ambiguous word

Reference: Adam Kilgarriff/Martha Palmer. 2000. Special Issue on SENSEVAL. *Computers and the humanities*, 34(1-2).

Settings:
Classification Algorithm

Naive Bayes Classification:

- Fast, used often, performs well (in general)
- Uses distributional and contextual information

Training: Bayes rule

$$P(s_k|c) = \frac{P(c|s_k)}{P(c)} P(s_k)$$

s_k: sense k of ambiguous word w
c: context words within context window ($\pm$ 3)

Testing: Bayes decision rule

Decide s' if $P(s'|c) > P(s_k|c)$ for $s_k \neq s'$

Settings:
Ambiguous Words / Pseudowords

Ambiguous Words:

- Only nouns (15)
  - only not POS ambiguous (8)
  - only with training data (4)
- Coarse-grained senses from SENSEVAL dictionary entries (3 x 2 senses, 1 x 4 senses)

Pseudowords:

- Per ambiguous word: choose sense words to make up pseudowords
  - same frequency distribution of sense words as senses of ambiguous word
  - make arbitrary selection

Results and Evaluation

	Base.	Results	Diff. [Std. err.]
accident (2 s.)	92.88	84.45	**- 8.43**
pseudowords	92.38	91.98	- 0.40 [± 0.89]
behaviour (2 s.)	95.70	84.95	**-10.75**
pseudowords	94.85	94.56	- 0.29 [± 1.24]
excess (4 s.)	58.06	50.35	**- 7.71**
pseudowords	58.19	73.22	+15.03 [± 1.55]
shirt (2 s.)	58.98	57.50	**- 1.48**
pseudowords	58.38	72.48	+14.1 [± 6.60]

Table 1: Results (correct in %)

Results: Significantly different

Possible explanation:

Pseudowords have clearly distinct senses

Senses of ambiguous words are more related

Problem: Bad performance on ambiguous words

→ High baseline? Not enough data?

Conclusions and Future Work

Conclusion: Disambiguation of pseudowords is only in a limited way comparable to disambiguation of real ambiguous words

Other corpora: Try same experiment on other sense-tagged corpora

Too few sense-tagged corpora: esp. for languages other than English

Initial evaluation problem: Test WSD algorithms for languages without sense-tagged corpora not solved

Only solution: Hand-annotation (cost-, expertise- and time-intensive)

Questions

- Should we forget about pseudowords or should we keep using them for comparison of WSD algorithms on a relative basis?

- Is it an improvement to modify the technique of pseudowords?

	Base.	Results	Diff. [Std. err.]
accident	92.88	84.45	- **8.43**
timwe	91.90	91.56	- 0.34
yeatra	92.47	91.77	- 0.70
peolang	93.10	91.88	- 0.59
woan	92.12	93.44	+ 0.97
goveq	92.35	91.33	- 1.14
mean	92.38	91.98	- 0.40 [± 0.89]
behaviour	95.70	84.95	**-10.75**
peostan	93.40	92.99	- 0.41
tima	95.33	95.64	+ 0.31
yeagro	95.34	94.04	- 1.30
wodat	94.92	93.79	- 1.13
gopay	95.26	96.36	+ 1.10
mean	94.85	94.56	- 0.29 [± 1.24]
excess	58.06	50.35	- **7.71**
womuconba	58.62	71.86	+13.24
gopoemch	57.64	72.92	+15.28
dacipapro	57.71	73.98	+16.27
pemanora	58.64	73.00	+14.36
heterite	58.37	74.39	+16.02
mean	58.19	73.22	+15.03 [± 1.55]
shirt	58.98	57.50	- **1.48**
schoclu	59.02	72.79	+13.77
mastre	58.55	74.83	+16.28
cimon	58.04	78.69	+20.65
coufam	57.96	63.91	+ 5.95
wogia	58.33	72.22	+13.89
mean	58.98	72.48	+14.10 [± 6.60]

Table 2: Results (correct in %)