Good Neighbors Make Good Senses

Exploiting Distributional Similarity for Unsupervised WSD

Samuel Brody Mirella Lapata

School of Informatics
University of Edinburgh

Coling 2008

Outline

Outline

Motivation

Supervised WSD

- Most accurate WSD systems to date are supervised.
- Rely on sense-labeled training data to train standard classifiers.
  - – Acquiring sufficient labeled data is very expensive.
  - – Limits the use in new domains and languages.
  - – Makes supervised WSD unfeasible for many applications.

Unsupervised WSD

- \+ Independent of labeled data.
- \+ Most promising solution for large-scale use.
- – Much less accurate than supervised methods.

The Supervision Gap

Question:

How to achieve supervised accuracy, without sacrificing unsupervised advantages?

Answer: Automatically create labeled data.

- Circumvent the problematic, knowledge rich, step
- Go directly to the data

Outline

Multilingual Approach

Parallel Corpora (Gale et al., 1992; Ng et al., 2003)

- Assume different translations represent different senses.
- Label the instances according to their translation.

Disadvantages

- requires accurate alignment
- limited availability (restricted domains)
- limited sense distinctions
- not free of manual intervention

Linguistic Knowledge

Synonyms from a Lexical Resource (Leacock et al., 1998; Mihalcea, 2002)

- Obtain synonymous/related words for each sense.
- Use related words to learn indicators of sense.

Disadvantages

- disregards domain - synonyms may be irrelevant
- polysemy of synonyms degrades performance
- trade-off between specificity and amount of information

Our Approach

Distributional Neighbors

- Extension of McCarthy et al. (2004).
- Based on distributional similarity - words are related if used in similar contexts.
- Uses semantic similarity to associate neighbors with senses.

Method Advantages

+ relates directly to context cues
+ domain specific
+ polysemy restricted by similarity

Design Advantages

+ one model per target word
+ does not use a tailored classifier

Example

WordNet senses for the word *"sense"*:

1. A general conscious **awareness**.
   (e.g., *a sense of security*)
2. The **meaning** of a word.
   (e.g., *The dictionary gave several senses for the word*)
3. Sound practical **judgment**.
   (e.g., *I can't see the sense in doing it now*)
4. A natural appreciation or **ability**.
   (e.g., *keen musical sense*).

Example

Semantic Neighbors from WordNet

- **Neighbors of sense 1:** sentience, sensation, sensitivity, sensitiveness, sensibility, modality, module, knowingness, ...
- **Neighbors of sense 2:** signified, acceptation, signification, significance, meaning, import, symbolization, symbolisation,...
- **Neighbors of sense 3:** gumption, logic, sagacity, judgment, judgement, discernment, prudence, judiciousness, eye, ...
- **Neighbors of sense 4:** hold, grasp, appreciation

- few exact synonyms
- many related words
- neighbors are not “substitutable”
- neighbors are themselves polysemous

Neighbor Polysemy

Monosemous Semantic Neighbors

- **Neighbors of sense 1:** cognisance, self-awareness
- **Neighbors of sense 2:** signified, signification, nuance, moral, intention

- greatly reduced number of neighbors
- no monosemous neighbors for last two senses
- neighbors may be rare

Example

Distributional Neighbors

- **Neighbors of sense 1:** awareness, feeling, instinct, enthusiasm, sensation, vision, tradition, consciousness, anger, panic, loyalty
- **Neighbors of sense 2:** emotion, belief, meaning, manner, necessity, tension, motivation

- No neighbors for last two senses.
- Not prevalent in the corpus (corroborated by the test data).

Outline

Method

1. Acquire distributional neighbors (using Lin 1998 or InfoMap)
2. Associate neighbors with senses through semantic similarity measure (based on Lesk 1986)
3. Extract instance vectors for neighbors from large corpus
4. Label instances with associated sense
5. Use labeled data to train supervised classifier

*"... an attempt to state the **meaning** of a word"*

becomes

*"... an attempt to state the **sense** (s#2) of a word."*

Outline

Data

Corpus

The British National Corpus (BNC)

- cross-section of 20th century, written & spoken, British English.
- 100 million words

Evaluation

Nouns from Senseval 2 & 3 lexical samples

- instances from BNC
- coarse-grain senses

	# Words	Ambiguity	1st Sense
SE-2	25	3.28	65.96%
SE-3	20	4.35	60.90%

Tools

Distributional Neighbors

- dependency based (Lin, 1998)
- co-occurrence based (InfoMap)

Classifiers

Evaluated on a variety of classifiers, from different paradigms:

- SVM - multi-class bound-constrained SVC (Hsu and Lin, 2001)
- Maximum Entropy (Megam, Daumé III 2004)
- Label Propagation (SemiL, Zhu and Ghahramani 2002)

Outline

Results

Senseval 2	AllWN	MonoWN	Depend	Manual
SVM	48.12%	53.29%	64.38%	72.52%
MaxEnt	40.93%	52.11%	62.32%	71.91%
LP	42.67%	49.54%	63.32%	69.28%
McCarthy	59.98%			
Lesk	48.12%			

Senseval 3	AllWN	MonoWN	Depend	Manual
SVM	53.16%	46.32%	57.47%	71.22%
MaxEnt	49.67%	44.85%	57.35%	71.75%
LP	47.41%	43.60%	60.60%	67.57%
McCarthy	57.14%			
Lesk	48.66%			

Results

Senseval 2	AllWN	MonoWN	Depend	Manual
SVM	48.12%	53.29%	64.38%	72.52%
MaxEnt	40.93%	52.11%	62.32%	71.91%
LP	42.67%	49.54%	63.32%	69.28%
McCarthy	59.98%			
Lesk	48.12%			

Senseval 3	AllWN	MonoWN	Depend	Manual
SVM	53.16%	46.32%	57.47%	71.22%
MaxEnt	49.67%	44.85%	57.35%	71.75%
LP	47.41%	43.60%	60.60%	67.57%
McCarthy	57.14%			
Lesk	48.66%			

Results

Senseval 2	AllWN	MonoWN	Depend	Manual
SVM	48.12%	53.29%	64.38%	72.52%
MaxEnt	40.93%	52.11%	62.32%	71.91%
LP	42.67%	49.54%	63.32%	69.28%
McCarthy	59.98%			
Lesk	48.12%			

Senseval 3	AllWN	MonoWN	Depend	Manual
SVM	53.16%	46.32%	57.47%	71.22%
MaxEnt	49.67%	44.85%	57.35%	71.75%
LP	47.41%	43.60%	60.60%	67.57%
McCarthy	57.14%			
Lesk	48.66%			

Results

Senseval 2	AllWN	MonoWN	Depend	Manual
SVM	48.12%	53.29%	64.38%	72.52%
MaxEnt	40.93%	52.11%	62.32%	71.91%
LP	42.67%	49.54%	63.32%	69.28%
McCarthy	59.98%			
Lesk	48.12%			

Senseval 3	AllWN	MonoWN	Depend	Manual
SVM	53.16%	46.32%	57.47%	71.22%
MaxEnt	49.67%	44.85%	57.35%	71.75%
LP	47.41%	43.60%	60.60%	67.57%
McCarthy	57.14%			
Lesk	48.66%			

Secondary Senses Accuracy

Senseval 2	Depend	
SVM	14.3%	(60.1%)
MaxEnt	6.3%	(66.9%)
LP	8.9%	(63.3%)

Senseval 3	Depend	
SVM	17.6%	(45.0%)
MaxEnt	8.5%	(55.0%)
LP	5.6%	(60.9%)

Secondary Senses Accuracy

Senseval 2	Depend	
SVM	14.3%	(60.1%)
MaxEnt	6.3%	(66.9%)
LP	8.9%	(63.3%)

Senseval 3	Depend	
SVM	17.6%	(45.0%)
MaxEnt	8.5%	(55.0%)
LP	5.6%	(60.9%)

Fine Grain

		Coarse Grain	Fine Grain	Decrease
Ambiguity	SE 2	3.28	5.6	+2.32
	SE 3	4.35	4.8	+0.45
Manual	SE 2	69.28%	53.90%	-15.38%
	SE 3	67.57%	61.54%	-6.03%
Dependency	SE 2	63.32%	47.71%	-15.61%
	SE 3	60.60%	52.80%	-7.80%

Outline

Conclusions

- unsupervised WSD method
- surpasses state-of-the-art unsupervised methods
- utility similar to supervised framework

Future Work

- scale to the all-words task
- examine confidence-based scoring
- integrate with supervised learning

Thank You!

Syntax-Free Neighbors

	Mono.	Co-occur	Depend.
Senseval 2	49.54%	58.61%	63.32%
Senseval 3	47.41%	59.92%	60.60%

Bibliography

Daumé III, Hal. 2004. Notes on CG and LM-BFGS optimization of logistic regression.

Gale, W., K. Church, and D. Yarowsky. 1992. A method for disambiguating word senses in a large corpus. In *Computers and the Humanities*. 26, pages 415–439.

Hsu, C. and C. Lin. 2001. A comparison of methods for multi-class support vector machines.

Leacock, Claudia, Martin Chodorow, and George A. Miller. 1998. Using corpus statistics and wordnet relations for sense identification. *Computational Linguistics* 24(1):147–165.

Lesk, Michael. 1986. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. In *Proceedings of the 5th SIGDOC*. New York, NY, USA, pages 24–26.

Lin, Dekang. 1998. Automatic retrieval and clustering of similar words. In *Proceedings of the 17th international conference on Computational linguistics*. Association for Computational Linguistics, Morristown, NJ, pages 768–774.

McCarthy, Diana, Rob Koeling, Julie Weeds, and John Carroll. 2004. Finding predominant senses in untagged text. In *Proceedings of the 42th ACL*.