Introduction to Corpus-based Computational Semantics:

Semantic relations between words

Alexander Koller
Columbia University

Sabine Schulte im Walde
University of Stuttgart

Introductory Course
ESSLLI 2007
Trinity College, Dublin, Ireland
August 6-10, 2007

Where are we?

- Word sense inventories with semantic relations, e.g. Wordnet
  - » word sense disambiguation
- Semantic similarity between words
  - » distributional hypothesis
  - » association measures
  - » clustering

Now what?

- Semantic relations between words ← today
  - » Can we learn Wordnet relations automatically?
  - » Word relations vs. sense relations
- Similarity between phrases ← tomorrow
  - » Paraphrase: Two semantically similar word sequences
  - » Extended distributional hypothesis
- We will mostly ignore word sense ambiguity today (but see e.g.: Snow et al. 2006)

Learning semantic relations

The big picture

- Word relations are useful.
  - » hyponymy (is-a)
    - information retrieval etc.: find more specific or approximate matches
  - » meronymy (part-of)
    - answer questions about parts of objects
    - metonymy
  - » various semantic relations between verbs
    - happens-before: reconstruct sequence of events

The big picture

- Developing Wordnet-like resources by hand is expensive.
  - » new languages
  - » new domains

- So let's learn them from corpora!

Semantic relations in corpora

- Problem: Commonsense relations like hyponymy are so obvious that they're rarely expressed explicitly (Grice maxims).
- Need to be sneaky.
  - » The bow lute, such as the Bambara ndang, is plucked ... (Hearst 92)
  - » XML and other markup languages
  - » the car's windshield

Using surface patterns

- Idea: Surface patterns encode semantic relations.
- Methods:
  - » come up with surface patterns for the relation that we find interesting
  - » search for occurrences of this pattern in a large corpus (or the Internet)
- Criteria for patterns:
  - » occur frequently and in many different text genres
  - » highly specific
  - » easy to recognize

Hearst 92: Patterns for hyponymy

- NP such as NP
- such NP as NP
  - » works by such authors as Herrick, Goldsmith, and Shakespeare
- NP or/and other NP
  - » bruises, wounds, broken bones or other injuries
  - » temples, treasuries, and other important civic buildings
- NP, including NP
  - » all common-law countries, including Canada and England, ...
- NP, especially NP
  - » most European countries, especially France, England, and Spain

How much of an NP to use?

- Often the matched noun phrase will contain words that we don't want to consider:

  temples and other important civic buildings

  delete "other", "some", etc. without changing hyponymy

- What words should be deleted depends on the application.

How much of an NP to use?

- Often the matched noun phrase will contain words that we don't want to consider:

  temples and other important civic buildings

context-dependent meaning
-- better remove this

- What words should be deleted depends on the application.

Example hyponyms

fabric	silk
grain	barley
disorders	epilepsy
businesses	nightclub
crimes	kidnappings
countries	Brazil India Israel
vegetables	broccoli
games	checkers
regions	Texas
assets	stocks
jurisdictions	Illinois
companies	Volvo Saab
institutions	Tufts
airlines	Pan USAir
agencies	Clic Zoli
companies	Shell

crops	milo
wildlife	deer raccoons
conditions	epilepsy
conveniences	showers microwaves
perishables	fruit
agents	bacteria viruses
felonies	shootings stabbings
euphemisms	restrictees detainees
goods	shoes
officials	stewards
geniuses	Einstein Newton
gifts	liquor
disasters	AIDS
materials	glass ceramics
partner	Nippon

- Grolier's Encyclopedia (8M words), "such as" pattern.

Sometimes, things go wrong

- Metonymy:
  The commission expected Berlin -- and other countries in the bloc -- to take further action ...
- Underspecified hypernyms:
  fiction devices such as plot, character, setting, and language
- Context and point-of-view dependent pairs:
  warheads used against less protected targets such as aircraft

- Also: Only 152 N-N pairs in 8.6 million words!

Finding new patterns

- Patterns are very specific, so we want lots of them.
- If we do it purely by hand, we're sure to miss some.
- Semi-automatic procedure:
  - » pick some word pairs in the semantic relation
  - » search corpus for close-by occurrences of these words
  - » generalize over occurrences to find patterns
- In principle, applicable to arbitrary semantic relations; but manual generalization step is unpleasant.
- Some experiments with automatic generalization (e.g. Ravichandran & Hovy 02; Pantel & Pennachiotti 06).

Meronymy

- We can try to use exactly the same method for other semantic relations, e.g. meronymy (part-of).
- Patterns from Berland & Charniak (1999):
  - » N's N
    "building's basement"
  - » N of/in the/a Adj* N
    "basement of/in a building"
  - » N-PL of/in N-PL
    "basements of/in buildings"
- Also do some statistical analysis on results.

Problems with meronymy

- But this works much less well than for hyponymy.
- "book":

Ocr.	Frame	Word	x/5
853	3069	author	5*
23	48	subtitle	4*
114	414	co-author	4*
7	16	foreword	5*
123	963	publication	2
5	10	epigraph	3*
9	32	co-editor	4*
51	499	cover	5*
220	3053	copy	2
125	1961	page	5*

"car":

Ocr.	Frame	Word	x/5
92	215	trunk	4*
27	71	windshield	5*
12	24	dashboard	5*
13	30	headlight	5*
70	318	wheel	5*
9	21	ignition	4*
43	210	hood	5*
119	880	driver	1
6	13	radiator	5*
4	6	shifter	1

- Results are much worse for less concrete nouns.

Why is meronymy harder?

- Patterns are much less specific than for hyponymy.
- positive:

The horn is *part of* the car.
The table *has* four legs.
The girl*'s* mouth is sensual.
The eyes *of* the baby are blue.
Each *door knob* was made of silver.
It was the girl *with* blue eyes.

negative:

He is *part of* the game (PARTICIPANT–EVENT)
Kate *has* four cats. (POSSESSION)
Mary*'s* brother is cute. (KINSHIP)
The dress *of* my niece is blue. (POSSESSION)
Dallas is a modern *Texas city*. (LOCATION)
The woman *with* triplets received a lot of attention. (KINSHIP)

- This means that B&C derive lots of false positives.

A possible way out

- Girju et al. 2005:
  - » Equip each pattern with a constraint on the semantic classes of the NPs it connects
  - » only accept pattern instance as indicating meronymy if the NPs belong to this semantic class
- Learn semantic class constraints from word sense annotated corpora using a decision tree learner:
  - » communication#2 of the communication#2 - YES
    ("act of the play")
  - » process#2 of the person#1 - NO
    ("growth of the child")

Evaluation of Girju et al.

- Improved recognition of meronyms.
- However, their approach requires extensive manual preparation of corpora etc.

Learning semantic verb relations

- Pattern-based approach can also be used to learn semantic relations between verbs: VerbOcean system (Chklovski & Pantel 04).
- Method:
  - » Hearst-style manual pattern induction for verb relations
  - » Hearst-style pattern matching, but on the Web
  - » Establish semantic relations using Mutual Information
- Can live with less specific patterns because statistical analysis will (hopefully) filter out random hits.

Relations between verbs

- similarity (produce / create):
  “X, i.e. Y”
- strength (wound / kill):
  “not just Xed but Yed”
- enablement (fight / win):
  “Xed * by Ying the”
- antonymy (open / close):
  “either Xed or Yed”
- happens-before (marry / divorce):
  “to X and then Y”

Estimating relatedness

- Association strength between verbs V1 and V2 and the pattern p (based on mutual information):

$$S_p(V_1, V_2) = \frac{P(V_1, p, V_2)}{P(p) \times P(V_1) \times P(V_2)}$$

- Estimate all probabilities using Google queries:
  P(V1): count Google hits for “to divorce”
  P(V1,p,V2): count hits for “married and then divorced”
  P(p): count in a large offline corpus
- Verb pair counts as related if association via a pattern for this relation exceeds some threshold.

Evaluation

Table 5. Accuracy of each semantic relation.

Semantic Relation	*System Tags*	*Tags Correct*	*Preferred Tags Correct*
similarity	41	63.4%	40.2%
strength	14	75.0%	75.0%
antonymy	8	50.0%	43.8%
enablement	2	100%	100%
no relation	35	72.9%	72.9%
happens before	17	67.6%	55.9%

- Two human judges, 100 random verb pairs.
- NB: Low recall.

Some results

- VerbOcean finds quite a few correct relations:

  adore [opposite-of] detest :: 18.491557
  idolize [stronger-than] adore :: 19.314744
  blacken [similar] scorch :: 19.286283
  indict [happens-before] convict :: 19.121721
  abhor [stronger-than] dislike :: 19.877318

- But also some pretty crazy ones:

  adore [opposite-of] marry :: 12.108326

Mutual information: Problems

- Also, pointwise mutual information is known to overestimate the importance of rare events.
- The first few pairs, by association strength:

soft-pedal [happens-before] overarch :: 26.451366
rear-end [happens-before] sideswipe :: 25.786826
deep-fry [stronger-than] caramelize :: 25.718227
air-drop [happens-before] airdrop :: 25.456703
misread [can-result-in] double-check :: 25.409792

Conclusion

- Pattern-matching methods for learning semantic relations work best if unambiguous patterns are available
  - » hyponymy works better than meronymy, verb relations
- Even for difficult relations, they find more correct pairs than expected by chance
  - » but low precision: how useful for applications?
- Attempts at dealing with ambiguous patterns
  - » learning constraints on related pairs
  - » statistical analysis
- Rely on redundancy: Web vs. corpora.