More Lexical Semantics

COMP-599

Oct 22, 2015

Outline

Approaches to detecting lexical semantic relationships

- Hearst's lexico-syntactic patterns
- Distributional semantics

From syntax to semantics

- Dependency trees
- Semantic roles and PropBank
- Frame Semantics and FrameNet

Last Class

Word sense disambiguation

Another lexical semantic task: detecting words that are in a certain lexical semantic relation

e.g., a rabbit is a mammal

Hearst Patterns (1992)

Pairs of terms that are in hyponym-hypernym relationships tend to occur in certain **lexico-syntactic** patterns:

The bow lute, such as the Bambara ndang, is plucked and has an individual curved neck for each string.

(Hearst, 1992)

What are the hyponym and hypernym in this passage?

Hearst's Original Patterns

NP such as {NP}* {and|or} NP

such NP as {NP .}* {or|and} NP

NP {, NP}* {,} or other NP

NP {, NP}* {,} and other NP

NP {,} including {NP, }* {or|and} NP

NP {,} especially {NP ,}* {or|and} NP

Exercise: label each NP as indicating hyponym or hypernym

How To Find Patterns?

Be smart and just think of them?

Hint: Think about our the idea of bootstrapping that we saw from last class

Other Relations

Using this approach, Hearst patterns have also been discovered and used for other relations between words, e.g., cause-effect relationships (Girju, 2002)

- e.g., *Earthquakes* ***cause*** *tidal waves*.
- NP-cause cause NP-effect

Other verbs:

- *induce, give rise (to), stem (from), etc.*

Synonymy

We've looked at the relationship between two words that co-occur, and their intervening words.

__Extinct birds__, <u>such as</u> __dodos__, __moas__, <u>and</u> __elephant birds__

What if the words don't tend to co-occur directly?

e.g., synonyms are supposed to be substitutes of each other

*The dodo **went extinct** in the 17th century.*

*The dodo **died out** in the 17th century.*

Another signal: the words that tend to co-occur with the target words

Distributional Semantics

You shall know a word by the company it keeps.

Firth, 1957

Understand a term by the distribution of words that appear near the term

Basic Idea

Go through a corpus of text. For each word, keep a count of all of the words that appear in its context within a window of, say, 5 words.

John Firth was an English linguist and a leading figure in British linguistics during the 1950s.

Term-Context Matrix

Each row is a vector representation of a word

	the	*was*	*and*	*British*	*linguist*
Firth	5	7	12	6	9
figure	276	87	342	56	2
linguist	153	1	42	5	34
1950s	12	32	1	34	0
English	15	34	9	5	21

Context words

Co-occurrence counts

Target words

Cosine Similarity

Compare word vectors A and B by

$$\text{sim}(A, B) = \frac{A \cdot B}{||A||\,||B||}$$

This corresponds to the cosine of the angle between the two vectors.

Range of values:

- -1 Vectors point in opposite directions
- 0 Vectors are orthogonal
- 1 Vectors point in the same direction

If vectors are positive (e.g., they're count vectors), similarity score is between 0 and 1.

Reasons Words Can Be Related

Cosine similarity gives you a lot more than synonymy!

Any words that tend to share context words will have high cosine similarity. What are some reasons for this?

- Synonymy or near-synonymy
- others?

Similarity vs. Relatedness

Similarity:

- Specifically about synonymy, hypernymy, hyponymy
- e.g., *chair* is similar to *furniture*
- *cat* is not similar to *scratching post*

Relatedness:

- Includes anything that might be associated
- *good* is related to *bad*

Confusingly, people often say similarity when they mean relatedness. e.g., what is cosine similarity a measure of?

Rescaling the Vectors

Instead of raw counts, people usually use a measure of how much two words are correlated with each other, above chance.

Pointwise mutual information (PMI)

$$\mathrm{pmi}(w_1, w_2) = \log \frac{P(w_1, w_2)}{P(w_1)P(w_2)}$$

- Numerator: probability of both words occurring (i.e., in each other's context)
- Denominator: probability of each word occurring in general

Pointwise Mutual Information Example

the occurs 100,000 times in a corpus with 1,000,000 tokens, of which it co-occurs with *linguistics* 300 times. *linguistics* occurs 2,500 times in total.

$$P(the, linguistics) = 0.0003$$
$$P(the) = 0.1$$
$$P(linguistics) = 0.0025$$

$$\text{pmi}(the, linguistics) = \log \frac{0.0003}{0.00025} = 0.26303 \text{ (base 2)}$$

If ratio is < 1, PMI is negative

- People often discard negative values → positive pointwise mutual information (PPMI)

Vector Space Evaluation

Word vectors have no objective inherent value

- Is the vector [0.4, 0.3, -0.2] better for the word *linguistics*, or [0.2, 0.5, 0.1]?
- Evaluate the similarity of vectors to each other instead
- Correlate against some gold standard. Many possible choices: http://wordvectors.org/suite.php

e.g., the WS-353 data set (Finkelsteinet al., 2002)

monk	oracle	5
cemetery	woodland	2.08
food	rooster	4.42
coast	hill	4.38
forest	graveyard	1.85
shore	woodland	3.08
monk	slave	0.92

Much More to This!

Singular Value Decomposition (SVD) – a method that can be used to reduce the dimensionality of vectors by approximately reconstructing the term-context matrix

- Removes noise and prevents overfitting of model

Neural network models – train vector space representation of word to predict words in context

- word2vec (Mikolov et al., 2013)

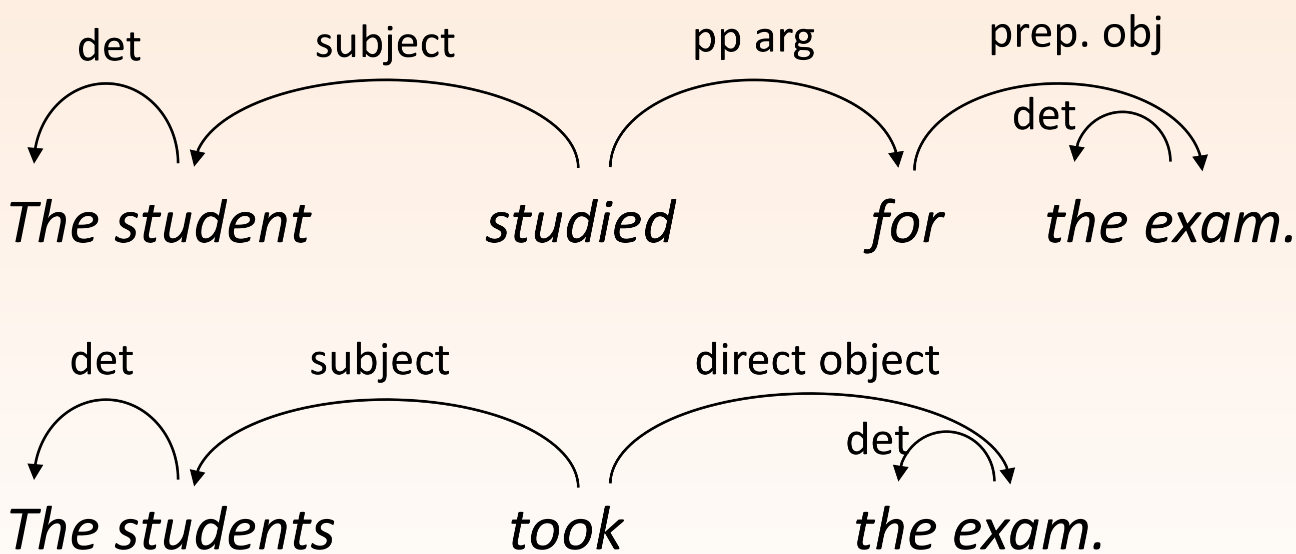
Abrupt Shift

Let's go back to looking at syntax and how it interacts with semantics.

- Dependency trees
- Semantic roles
- Frame Semantics

Dependency Structure

Dependency trees encode syntactic information.



Active-Passive Alternation

These groups of sentences have the same literal meaning, but are syntactically different.

The students took the exam.

The exam was taken by the students

The kind professor gave all the students As.

The kind professor gave As to all the students.

Semantic Roles

Instead of talking about subjects and (direct) objects, we can talk about agents and themes

[*The students*$_{\text{AGENT}}$] *took* [*the exam*$_{\text{THEME}}$].

[*The exam*$_{\text{THEME}}$] *was taken by* [*the students*$_{\text{AGENT}}$].

[*The kind professor*$_{\text{AGENT}}$] *gave* [*all the students*$_{\text{RECIPIENT}}$] [*As*$_{\text{THEME}}$].

[*The kind professor*$_{\text{AGENT}}$] *gave* [*As*$_{\text{THEME}}$] *to* [*all the students*$_{\text{RECIPIENT}}$].

Another Naming Confusion

Various names for semantic roles, with various implicit theoretical assumptions and nuances

- thematic relations
- thematic roles
- theta roles

In the NLP community, **semantic role** is the most common name.

Also, various schemes for what set of roles there should be with different names.

- THEME or PATIENT?

PropBank (Palmer et al., 2005)

Penn Treebank trees augmented with semantic role annotations

Labels semantic roles of *verbal* arguments:

A0 AGENT

A1 THEME/PATIENT

A2 RECIPIENT

etc.

https://verbs.colorado.edu/~mpalmer/projects/ace.html

Another project, NomBank, annotates the nominal arguments (Meyers et al., 2004)

Connection to Real World

Why stop at normalizing these simple alternations, which all have the same head word?

Should also normalize for different expressions within the same real-world context

BUYER *bought* GOOD

BUYER *purchased* GOOD

BUYER *went shopping for* GOOD

The purchase of GOOD *by* BUYER

The acquisition of GOOD *by* BUYER

Frame Semantics (Fillmore, 1976)

Word meanings must relate to **semantic frames**, a schematic description of the stereotypical real-world context in which the word is used.

buy →

FrameNet (Baker et al., 1998)

A collection of semantic frames

- Name of frame (e.g., Commerce_buy)
- Frame elements
  - **Core**: (BUYER, GOODS)
  - **Non-core**: (MONEY, SELLER, MANNER, PLACE, PURPOSE, etc.)
- Linguistic expressions that realize this frame (*buy*, *buyer*, *purchase*, etc.)
- Relations between frames
  - Buying inherits from a Getting frame
  - Renting is inherited by a Buying frame (NOTE: NOT an ISA relationship)