Lexical Semantics, Distributions, Predicate-Argument Structure, and Frame Semantic Parsing

11-711 Algorithms for NLP
13 November 2018
(With thanks to Noah Smith and Lori Levin)

Semantics so far in course
- Previous semantics lectures discussed composing meanings of parts to produce the correct global sentence meaning
  - *The mailman bit my dog.*
- The “atomic units” of meaning have come from the lexical entries for words
- The meanings of words have been overly simplified (as in FOL): atomic objects in a set-theoretic model

Word Sense
- *Instead, a bank can hold the investments in a custodial account in the client’s name.*
- *But as agriculture burgeons on the east bank, the river will shrink even more.*
- *While some banks furnish sperm only to married women, others are much less restrictive.*
- *The bank is near the corner of Forbes and Murray.*

Four Meanings of “Bank”
- **Synonyms:**
  - bank\(^1\) = “financial institution”
  - bank\(^2\) = “sloping mound”
  - bank\(^3\) = “biological repository”
  - bank\(^4\) = “building where a bank\(^1\) does its business”
- The connections between these different senses vary from practically none (homonymy) to related (polysemy).
  - The relationship between the senses bank\(^4\) and bank\(^1\) is called metonymy.

Antonyms
- White/black, tall/short, skinny/American, ...
- But different dimensions possible:
  - White/Black vs. White/Colorful
  - Often culturally determined
- Partly interesting because automatic methods have trouble separating these from synonyms
  - Same semantic field

How Many Senses?
- This is a hard question, due to vagueness.
Ambiguity vs. Vagueness

- **Lexical ambiguity:** My wife has two kids (children or goats?)
- **vs. Vagueness:** 1 sense, but indefinite: horse (mare, colt, filly, stallion, ...) vs. kid:
  - I have two horses and George has three
  - I have two kids and George has three
- Verbs too: I ran last year and George did too
- **vs. Reference:** I, here, the dog not considered ambiguous in the same way

How Many Senses?

- This is a hard question, due to vagueness.
- Considerations:
  - Truth conditions (serve meat / serve time)
  - Syntactic behavior (serve meat / serve as senator)
  - Zeugma test:
    - #Does United serve breakfast and Pittsburgh?
    - ??She poaches elephants and pears.

Related Phenomena

- Homophones (would/wood, two/too/to):
  - Mary, merry, marry in some dialects, not others
- Homographs (bass/bass)

Word Senses and Dictionaries

- For NLP, databases of word senses are typically organized by lexical relations such as hypernym (IS-A) into a DAG
- This has been worked on for quite a while
- Aristotle’s classes (about 330 BC)
  - substance (physical objects)
  - quantity (e.g., numbers)
  - quality (e.g., being red)
  - Others: relation, place, time, position, state, action, affection

Ontologies
Word senses in WordNet3.0

The noun “bass” has 8 senses in WordNet:
1. bass1 (the lowest part of the musical range)
2. bass2 (the lowest part in polyphonic music)
3. bass1, base1 (an adult male singer with the lowest voice)
4. sea bass1, bass4 (the lean flesh of a saltwater fish of the family Serranidae)
5. freshwater bass1, bass3 (any of various North American freshwater fish with lean flesh, especially of the genus Micropterus)
6. bass6, bass voice1, basso7 (the lowest adult male singing voice)
7. bass1 (the member with the lowest range of a family of musical instruments)
8. bass8 (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

The adjective “bass” has 1 sense in WordNet:
1. bass1, deep3 (having or denoting a low vocal or instrumental range)

“deep voice”;
“a bass voice is lower than a baritone voice”;
“a bass clarinet”

“Rough” Synonymy

• Jonathan Safran Foer’s Everything is Illuminated

Noun relations in WordNet3.0

Is a hamburger food?

Sense 1
hamburger, beefburger —
(a fried cake of minced beef served on a bun)

⇒ sandwich
⇒ snack food
⇒ dish
⇒ nutriment, nourishment, nutrition...
⇒ food, nutrient
⇒ substance
⇒ matter
⇒ physical entity
⇒ entity
Verb relations in WordNet3.0

<table>
<thead>
<tr>
<th>Relation</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypernym</td>
<td>From events to superordinate events</td>
<td>( f_1 \rightarrow \text{small} )</td>
</tr>
<tr>
<td>Hyponym</td>
<td>From events to subordinate event (often via specific manner)</td>
<td>( \text{wall} \rightarrow \text{small} )</td>
</tr>
<tr>
<td>Synonym</td>
<td>From verbs (events) to the verbs (events) they entail</td>
<td>( \text{more} \rightarrow \text{step} )</td>
</tr>
<tr>
<td>Antonym</td>
<td>Semantic opposition between lemmas</td>
<td>( \text{increase} \rightarrow \text{decrease} )</td>
</tr>
<tr>
<td>Derived Form</td>
<td>Lemmas with same morphological root</td>
<td>( \text{destroy} \rightarrow \text{destructive} )</td>
</tr>
</tbody>
</table>

• Not nearly as much information as nouns

Still no “real” semantics?

• Semantic primitives:
  \[ \text{Kill}(x,y) = \text{CAUSE}(x, \text{BECOME}(\text{NOT}(\text{ALIVE}(y)))) \]
  \[ \text{Open}(x,y) = \text{CAUSE}(x, \text{BECOME}(\text{OPEN}(y))) \]

• Conceptual Dependency: \( PTRANS, ATRANS, \ldots \)
  - The waiter brought Mary the check
  \[ PTRANS(x) \land \text{ACTOR}(x, \text{Waiter}) \land \text{OBJECT}(x, \text{Check}) \land \text{TO}(x, \text{Mary}) \]
  \[ ATRANS(y) \land \text{ACTOR}(y, \text{Waiter}) \land \text{OBJECT}(y, \text{Check}) \land \text{TO}(y, \text{Mary}) \]

Word similarity

• Human language words seem to have real-valued semantic distance (vs. logical objects)

• Two main approaches:
  - Thesaurus-based methods
    • E.g., WordNet-based
  - Distributional methods
    • Distributional “semantics”, vector “semantics”
    • More empirical, but affected by more than semantic similarity (“word relatedness”)

Thesaurus-based Word Similarity

• Simplest approach: path length
Better approach: weighted links

- Use corpus stats to get probabilities of nodes
- Refinement: use info content of LCS:
  \[ 2 \log P(g.f.) / (\log P(hill) + \log P(coast)) = 0.59 \]

<table>
<thead>
<tr>
<th>entity</th>
<th>0.395</th>
</tr>
</thead>
<tbody>
<tr>
<td>intimate-object</td>
<td>0.167</td>
</tr>
<tr>
<td>natural-object</td>
<td>0.0163</td>
</tr>
</tbody>
</table>

geological-formation 0.00176

0.000113 natural-elevation

0.0000189 hill

0.0000036 coast

Distributional Word Similarity

- Determine similarity of words by their distribution in a corpus
  - “You shall know a word by the company it keeps!” (Firth 1957)
- E.g.: 100k dimension vector, “1” if word occurs within “2 lines”:

<table>
<thead>
<tr>
<th></th>
<th>arts</th>
<th>hall</th>
<th>data</th>
<th>function</th>
<th>large</th>
<th>sugar</th>
<th>summarized</th>
<th>water</th>
</tr>
</thead>
<tbody>
<tr>
<td>apricot</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>pineapple</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>information</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

- “Who is my neighbor?” Which functions?

Who is my neighbor?

- Linear window? 1-500 words wide. Or whole document. Remove stop words?
- Use dependency-parse relations? More expensive, but maybe better relatedness.

<table>
<thead>
<tr>
<th>object of direct</th>
<th>object of indirect</th>
<th>of the</th>
<th>of the</th>
<th>of the</th>
<th>of the</th>
<th>of the</th>
<th>of the</th>
<th>of the</th>
<th>of the</th>
</tr>
</thead>
<tbody>
<tr>
<td>cake</td>
<td>cheese</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>tea</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pepsi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>champagne</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>liquid</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>beer</td>
<td>&lt;SOME AMOUNT&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Weights vs. just counting

- Weight the counts by the a priori chance of co-occurrence
- Pointwise Mutual Information (PMI)
- Objects of drink:

<table>
<thead>
<tr>
<th>object</th>
<th>count</th>
<th>PMI</th>
<th>assoc</th>
<th>object</th>
<th>count</th>
<th>PMI</th>
<th>assoc</th>
</tr>
</thead>
<tbody>
<tr>
<td>bunch</td>
<td>2</td>
<td>12.34</td>
<td></td>
<td>wine</td>
<td>2</td>
<td>9.34</td>
<td></td>
</tr>
<tr>
<td>beer</td>
<td>2</td>
<td>11.75</td>
<td></td>
<td>water</td>
<td>7</td>
<td>7.65</td>
<td></td>
</tr>
<tr>
<td>Pepsi</td>
<td>4</td>
<td>11.75</td>
<td></td>
<td>anything</td>
<td>3</td>
<td>5.15</td>
<td></td>
</tr>
<tr>
<td>champagne</td>
<td>4</td>
<td>11.75</td>
<td></td>
<td>much</td>
<td>3</td>
<td>5.15</td>
<td></td>
</tr>
<tr>
<td>liquid</td>
<td>2</td>
<td>10.53</td>
<td></td>
<td>it</td>
<td>3</td>
<td>1.25</td>
<td></td>
</tr>
<tr>
<td>beer</td>
<td>5</td>
<td>10.20</td>
<td></td>
<td>&lt;SOME AMOUNT&gt;</td>
<td>2</td>
<td>1.22</td>
<td></td>
</tr>
</tbody>
</table>

Distance between vectors

- Compare sparse high-dimensional vectors
  - Normalize for vector length
- Just use vector cosine?
- Several other functions come from IR community

Lots of functions to choose from

<table>
<thead>
<tr>
<th>function</th>
<th>formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>assoc prob</td>
<td>[ P(f</td>
</tr>
<tr>
<td>assoc PMI</td>
<td>[ \log P(f</td>
</tr>
<tr>
<td>assoc Lin</td>
<td>[ \log P(f</td>
</tr>
<tr>
<td>assoc test</td>
<td>[ \sqrt{P(f</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>function</th>
<th>formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>sim cosine</td>
<td>[ \frac{&lt;\mathbf{f}, \mathbf{w}&gt;}{</td>
</tr>
<tr>
<td>sim Jaccard</td>
<td>[ \frac{\sum \min(</td>
</tr>
<tr>
<td>sim Dice</td>
<td>[ \frac{2 \sum \min(</td>
</tr>
<tr>
<td>sim PG</td>
<td>[ D(f</td>
</tr>
</tbody>
</table>
Distributionally Similar Words

Rum  vodka  white  read  old  Mathematics  physics  biology
vodka  cognac  brandy  present  medieval  sociology  geology
brandy  whisky  liqueur  call  historic  psychology  anthropology
liqueur  detergent  call  release  famous  anthropology  astronomy
cola  gin  lemonade  sign  original  astronomy  geography
cola  cocoa  chocolate  know  main  geography  theology
lemonade  noodle  issue  accept  indian  theology  geography
tequila  noodle  issue  accept  indian  theology  geography
juice  consider  Japanese  accept  indian  theology  geography

(from an implementation of the method described in e.g. LIN. 1998. Automatic Retrieval and Clustering of Similar Words. COLING-ACL. Trained on newswire text.)

Human-subject Word Associations

<table>
<thead>
<tr>
<th>Stimulus: well</th>
<th>Stimulus: giraffe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of different answers: 20</td>
<td>Number of different answers: 26</td>
</tr>
<tr>
<td>Total count of all answers: 98</td>
<td>Total count of all answers: 98</td>
</tr>
<tr>
<td>brick 16 0.16</td>
<td>NECK 33 0.34</td>
</tr>
<tr>
<td>stone 9 0.09</td>
<td>Animal 9 0.09</td>
</tr>
<tr>
<td>paper 7 0.07</td>
<td>long 7 0.07</td>
</tr>
<tr>
<td>game 5 0.05</td>
<td>tail 7 0.07</td>
</tr>
<tr>
<td>blank 4 0.04</td>
<td>spots 5 0.05</td>
</tr>
<tr>
<td>bridge 4 0.04</td>
<td>long neck 4 0.04</td>
</tr>
<tr>
<td>fence 4 0.04</td>
<td>Africa 3 0.03</td>
</tr>
<tr>
<td>flower 4 0.04</td>
<td>Elephant 2 0.02</td>
</tr>
<tr>
<td>berlin 3 0.03</td>
<td>Hippopotamus 2 0.02</td>
</tr>
<tr>
<td>ceiling 3 0.03</td>
<td>Less 2 0.02</td>
</tr>
<tr>
<td>street 3 0.03</td>
<td></td>
</tr>
</tbody>
</table>

Recent events (2013-now)

- RNNs (Recurrent Neural Networks) as another way to get feature vectors
  - Hidden weights accumulate fuzzy info on words in the neighborhood
  - The set of hidden weights is used as the vector!

<table>
<thead>
<tr>
<th>Recycled variables</th>
<th>RNNs</th>
</tr>
</thead>
<tbody>
<tr>
<td>From openi.nlm.nih.gov</td>
<td></td>
</tr>
</tbody>
</table>

Recent events (2013-now)

- RNNs (Recurrent Neural Networks) as another way to get feature vectors
  - Hidden weights accumulate fuzzy info on words in the neighborhood
  - The set of hidden weights is used as the vector!
- Composition by multiplying (etc.)
  - CCG with vectors as NP semantics, matrices as verb semantics(!?)

Semantic Cases/Thematic Roles

- Developed in late 1960’s and 1970’s
- Postulate a limited set of abstract semantic relationships between a verb & its arguments: thematic roles or case roles.
- In some sense, part of the verb’s semantics
Problem: Mismatch between FOPC and linguistic arguments

- John broke the window with a hammer.
  - Broke(j,w,h)
- The hammer broke the window.
  - Broke(h,w)
- The window broke.
  - Broke(w)

- Relationship between 1st argument and the predicate is implicit, inaccessible to the system

Breaking, Eating, Opening

- John broke the window.
- The window broke.
- John is always breaking things.
- We ate dinner.
- We already ate.
- The pies were eaten up quickly.
- Open up!
  - Someone left the door open.
  - John opens the window at night.

Thematic Role example

- John broke the window with the hammer
- John: AGENT role
  window: THEME role
  hammer: INSTRUMENT role

- Extend LF notation to use semantic roles

Thematic Roles

- Is there a precise way to define meaning of AGENT, THEME, etc.?
- By definition:
  - “The AGENT is an instigator of the action described by the sentence.”
- Testing via sentence rewrite:
  - John intentionally broke the window
  - *The hammer intentionally broke the window

Thematic Roles [2]

- THEME
  - Describes the primary object undergoing some change or being acted upon
  - For transitive verb X, “what was Xed?”
  - The gray eagle saw the mouse
    “What was seen?” (A: the mouse)
Can We Generalize?

- **Thematic roles** describe general patterns of participants in generic events.
- This gives us a kind of shallow, partial semantic representation.
- First proposed by Panini, before 400 BC!

### Thematic Roles

<table>
<thead>
<tr>
<th>Role</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent</td>
<td>Volitional causer of the event</td>
<td>The waiter spilled the soup.</td>
</tr>
<tr>
<td>Force</td>
<td>Non-volitional causer of the event</td>
<td>The wind blew the leaves around.</td>
</tr>
<tr>
<td>Experiencer</td>
<td></td>
<td>Mary has a headache.</td>
</tr>
<tr>
<td>Theme</td>
<td>Most directly affected participant</td>
<td>Mary swallowed the pill.</td>
</tr>
<tr>
<td>Result</td>
<td>End-product of an event</td>
<td>We constructed a new building.</td>
</tr>
<tr>
<td>Content</td>
<td>Proposition of a propositional event</td>
<td>Mary knows you hate her.</td>
</tr>
<tr>
<td>Instrument</td>
<td></td>
<td>You shot her with a pistol.</td>
</tr>
<tr>
<td>Beneficiary</td>
<td></td>
<td>I made you a reservation.</td>
</tr>
<tr>
<td>Goal</td>
<td>Destination of a transferred thing</td>
<td>Go to hell</td>
</tr>
</tbody>
</table>

### Thematic Grid or Case Frame

- Example: break
  - The child broke the vase.  
    < agent subj theme obj >
  - The child broke the vase with a hammer.  
    < agent subj theme instr obj PP >
  - The hammer broke the vase.  
    < theme subj instr obj PP >
  - The vase broke.  
    < theme subj >

1. The Thematic Grid or Case Frame shows
   - How many arguments the verb has
   - What roles the arguments have
   - Where to find each argument
   - For example, you can find the agent in the subject position

**Verb Subcategorization**

Verbs have sets of allowed args. Could have many sets of VP rules. Instead, have a SUBCAT feature, marking sets of allowed arguments:

```
+none -- Jack laughed
+nnp -- Jack found a key
+nnp+nnp -- Jack gave Sue the paper
+v+inf -- Jack wants to fly
+v+ing -- Jack keeps hoping for the best
+nnp+vpbase -- Jack watched Sam look at his desk
+nnp+pp:to -- Jack gave the key to the man
```

50-100 possible frames for English; a single verb can have several. *(Notation from James Allen “Natural Language Understanding”)*

### Thematic Grid or Case Frame

- Example: break
  - The child broke the vase.  
    < agent subj theme obj >
  - The child broke the vase with a hammer.  
    < agent subj theme instr obj PP >
  - The hammer broke the vase.  
    < theme subj instr obj PP >
  - The vase broke.  
    < theme subj >

**Diathesis Alternation:**

A change in the number of arguments or the grammatical relations associated with each argument

- Chris gave a book to Dana.  
  < agent subj theme goal >
- A book was given to Dana by Chris.  
  < agent subj theme goal >
- Chris gave Dana a book.  
  < agent subj theme goal >
- Dana was given a book by Chris.  
  < agent subj theme goal >
The Trouble With Thematic Roles

- They are not formally defined.
- They are overly general.
- “agent verb theme with instrument” and “instrument verb theme” ...
  - The cook opened the jar with the new gadget.
  - Susan ate the sliced banana with a fork.

Two Datasets

- Proposition Bank (PropBank): verb-specific thematic roles
- FrameNet: “frame”-specific thematic roles
- These are lexicons containing case frames/thematic grids for each verb.

Proposition Bank (PropBank)

- A set of verb sense-specific “frames” with informal English glosses describing the roles
- Conventions for labeling optional modifier roles
- Penn Treebank is labeled with those verb-sense-specific semantic roles.

“Agree” in PropBank

- arg0: agreeer
- arg1: proposition
- arg2: other entity agreeing

  - The group agreed it wouldn’t make an offer.
  - Usually John agrees with Mary on everything.

“Fall (move downward)” in PropBank

- arg1: logical subject, patient, thing falling
- arg2: extent, amount fallen
- arg3: starting point
- arg4: ending point
- argM-loc: medium
- Sales fell to $251.2 million from $278.8 million.
- The average junk bond fell by 4.2%.
- The meteor fell through the atmosphere, crashing into Cambridge.

FrameNet

- FrameNet is similar, but abstracts from specific verbs, so that semantic frames are first-class citizens.
- For example, there is a single frame called change_position_on_a_scale.
Oil rose in price by 2%. It has increased to having them 1 day a month. Microsoft shares fell to 7 5/8. Colon cancer incidence fell by 50% among men.

Change position on a scale

Many words, not just verbs, share the same frame:

Verbs: advance, climb, decline, decrease, diminish, dip, double, drop, dwindle, edge, explode, fall, fluctuate, gain, grow, increase, jump, move, mushroom, plummet, reach, rise, rocket, shift, skyrocket, slide, soar, swell, swing, triple, tumble

Nouns: decline, decrease, escalation, explosion, fall, fluctuation, gain, growth, hike, increase, rise, shift, tumble

Adverb: increasingly

Conversely, one word has many frames

Example: rise

- Change-position-on-a-scale: Oil rose in price by two percent.
- Change-posture: a protagonist changes the overall position or posture of a body.
  - Source: starting point of the change of posture.
  - Charles rose from his armchair.
- Get-up: A protagonist leaves the place where they have slept, their Bed, to begin or resume domestic, professional, or other activities. Getting up is distinct from Waking up, which is concerned only with the transition from the sleeping state to a wakeful state.
  - I rose from bed, threw on a pair of camouflage shorts and drove my little Toyota Corolla to a construction clearing a few miles away.
- Motion-directional: In this frame a Theme moves in a certain Direction which is often determined by gravity or other natural, physical forces. The Theme is not necessarily a self-mover.
  - The balloon rose upward.
- Sidereal-appearance: An Astronomical_entity comes into view above the horizon as part of a regular, periodic process of (apparent) motion of the Astronomical_entity across the sky. In the case of the sun, the appearance begins the day.
  - At the time of the new moon, the moon rises at about the same time the sun rises, and it sets at about the same time the sun sets. Each day the sun’s rise offers us a new day.

FrameNet

- Frames are not just for verbs!
- Verbs: advance, climb, decline, decrease, diminish, dip, double, drop, dwindle, edge, explode, fall, fluctuate, gain, grow, increase, jump, move, mushroom, plummet, reach, rise, rocket, shift, skyrocket, slide, soar, swell, swing, triple, tumble
- Nouns: decline, decrease, escalation, explosion, fall, fluctuation, gain, growth, hike, increase, rise, shift, tumble
- Adverb: increasingly

FrameNet

- Includes inheritance and causation relationships among frames.
- Examples included, but little fully-annotated corpus data.

SemLink

- It would be really useful if these different resources were interconnected in a useful way.
- SemLink project is (was?) trying to do that
- Unified Verb Index (UVI) connects
  - PropBank
  - VerbNet
  - FrameNet
  - WordNet/OntoNotes

Semantic Role Labeling

- Input: sentence
- Output: for each predicate*, labeled spans identifying each of its arguments.

Example:

\[[agent The batter] hit [patient the ball] [time yesterday]\]

*Predicates are sometimes identified in the input, sometimes not.
But wait. How is this different from dependency parsing?

- Semantic role labeling
  - [agent The batter] hit [patient the ball] [time yesterday]
- Dependency parsing
  - [subject The batter] hit [object the ball] [modifier yesterday]

Subject vs agent

- Subject is a grammatical relation
- Agent is a semantic role

- In English, a subject has these properties
  - It comes before the verb
  - If it is a pronoun, it is in nominative case
    - I/he/she/we/they hit the ball.
    - *Me/him/her/us/them hit the ball.
  - If the verb is in present tense, it agrees with the subject
    - She/he/it hits the ball.
    - I/we/they hit the ball.
    - *She/he/it hit the ball.
    - *I/we/they hits the ball.
    - I hit the ball.
    - I hit the ball.

Semantic Role Labeling

- Input: sentence
- Output: segmentation into roles, with labels

Example from book:
- [agent The Examiner] issued [patient a special edition] [modifier yesterday]

Semantic Role Labeling: How It Works

- First, parse.
- For each predicate word in the parse:
  - For each node in the parse:
    - Classify the node with respect to the predicate.
Yet Another Classification Problem!

- As before, there are many techniques (e.g., Naïve Bayes)
- Key: what features?

Features for Semantic Role Labeling

- What is the predicate?
- Phrase type of the constituent
- Head word of the constituent, its POS
- Path in the parse tree from the constituent to the predicate
- Active or passive
- Is the phrase before or after the predicate?
- Subcategorization (= grammar rule) of the predicate

Feature example

- Example sentence:
  
  \[
  \text{arg}_0 \text{The Examiner} \text{ issued } \text{arg}_1 \text{ a special edition } \text{arg}_M \text{ yesterday}\]

- Arg0 features:
  
  issued, NP, Examiner, NNP, path, active, before, VP->VBD NP PP

Additional Issues

- Initial filtering of non-arguments
- Using chunking or partial parsing instead of full parsing
- Enforcing consistency (e.g., non-overlap, only one arg0)
- Phrasal verbs, support verbs/light verbs
  
  - *take a nap*: verb *take* is syntactic head of VP, but predicate is *napping*, not *taking*

Example

Two datasets, two systems

- Example from book uses PropBank
- Locally-developed system SEMAFOR works on SemEval problem, based on FrameNet
Shallow approaches to deep problems

- For many problems:
  - Shallow approaches much easier to develop
    - As in, possible at all for unlimited vocabularies
  - Not wonderful performance yet
    - Sometimes claimed to help a particular system, but often doesn’t seem to help
    - Definitions are not crisp
      - There clearly is something there, but the granularity of the distinctions very problematic
- Deep Learning will fix everything?

Questions?

Similarities to WSD

- Pick correct choice from N ambiguous possibilities
- Definitions are not crisp
- Need to pick a labelling scheme, corpus
  - Choices have big effect on performance, usefulness

Shallow approaches to deep problems

- For both WSD and SRL:
  - Shallow approaches much easier to develop
    - As in, possible at all for unlimited vocabularies
  - Not wonderful performance yet
    - Sometimes claimed to help a particular system, but often doesn’t seem to help
    - Definitions are not crisp
      - There clearly is something there, but the granularity of the distinctions very problematic
- Deep Learning will fix everything?
**SEMAFOR**

- A FrameNet-based semantic role labeling system developed within Noah’s research group
  - It uses a dependency parser (the MST Parser) for preprocessing
  - Identifies and disambiguates predicates; then identifies and disambiguates each predicate’s arguments

**Noun compounds**

- A very flexible (productive) syntactic structure in English
  - The noun noun pattern is easily applied to name new concepts (Web browser) and to disambiguate known concepts (fire truck)
  - Can also combine two NPs: incumbent protection plan, [undergraduate [computer science] [lecture course]]
  - Sometimes creates ambiguity, esp. in writing where there is no phonological stress: Spanish teacher
  - People are creative about interpreting even nonsensical compounds
  - Also present in many other languages, sometimes with special morphology
  - German is infamous for loving to merge words into compounds. e.g. Fremdsprachkenntnisse, ‘knowledge of foreign languages’

**Noun compounds**

- SemEval 2007 task: Classification of Semantic Relations between Nominals
  - 7 predefined relation types
    1. Cause-Effect: flu virus
    2. Instrument-User: laser printer
    3. Product-Producer: honeybee
    4. Origin-Entity: rye whiskey
    5. Purpose-Tool: soup pot
    6. Part-Whole: car wheel
    7. Content-Container: apple basket
  - [http://nlp.cs.swarthmore.edu/semeval/tasks/task04/description.shtml](http://nlp.cs.swarthmore.edu/semeval/tasks/task04/description.shtml)

**Noun compounds**

- SemEval 2010 task: Noun compound interpretation using paraphrasing verbs
  - A dataset was compiled in which subjects were presented with a noun compound and asked to provide a verb describing the relationship
  - nut bread elicited: contain(21); include(10); be made with(9); have(8); be made from(5); use(3); be made using(3); feature(2); be filled with(2); taste like(2); be made of(2); come from(2); consist of(2); be composed of(1); be blended with(1); be created out of(1); encapsulate(1); be flavored with(1)

**Thesaurus/dictionary-based similarity measures**

\[
\begin{align*}
\text{sim}_\text{path}(c_1, c_2) &= -\log \text{pathlen}(c_1, c_2) \\
\text{sim}_{\text{Resnik}}(c_1, c_2) &= -\log P(\text{LCS}(c_1, c_2)) \\
\text{sim}_{\text{Lee}}(c_1, c_2) &= 2 \times \log P(\text{LCS}(c_1, c_2)) / \log P(c_1) + \log P(c_2) \\
\text{sim}_{\text{Jc}}(c_1, c_2) &= 2 \times \log P(\text{LCS}(c_1, c_2)) / (\log P(c_1) + \log P(c_2)) \\
\text{sim}_{\text{Lesk}}(c_1, c_2) &= \sum_{c \in \text{REL5}} \text{overlap}(\text{gloss}(r(c_1)), \text{gloss}(q(c_2)))
\end{align*}
\]