Natural Language Processing

Lecture 19: Lexical Semantics
Word Associations
Word Associations

Rocket

Car

Volkswagen

Beach
Word Associations

automobile 0.85
vehicle 0.8
truck 0.75
carriage 0.7
Volkswagen 0.6
Honda 0.5
race 0.3
road 0.2
library 0.01

*Human labeled list*
Lexical Semantic

“Study of meaning of words and the systematic meaning-relation connections between words.”
Relations between Words/Senses

• Synonymy  Car, Automobile
• Antonymy  Love, Hate; Floor, Ceiling
• Hypeonymy  Honda, Car
• Hypernymy  Car, Honda
• Meronymy  Wheel, Car
• Holonymy  Car, Wheel
Lexical Mini-Ontology

holonym (whole)
meronym (has-a)
meronym (part)

wall.n.1

building.n.1
enclosure.n.1

build.v.1

wall.v.1

surround.v.2

fence.n.1

hyponym

antonym

antonymy

antonym

destroy.v.1

synonym

synonymy

synonym
What’s a fish?

- fish (any of various mostly cold-blooded aquatic vertebrates usually having scales and breathing through gills)
- aquatic vertebrate (animal living wholly or chiefly in or on water)
- vertebrate, craniate (animals having a bony or cartilaginous skeleton with a segmented spinal column and a large brain enclosed in a skull or cranium)
- chordate (any animal of the phylum Chordata having a notochord or spinal column)
- animal, animate being, beast, brute, creature, fauna (a living organism characterized by voluntary movement)
- organism, being (a living thing that has (or can develop) the ability to act or function independently)
- living thing, animate thing (a living (or once living) entity)
- whole, unit (an assemblage of parts that is regarded as a single entity)
- object, physical object (a tangible and visible entity; an entity that can cast a shadow)
- entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Source: WordNet, wordnet.princeton.edu
Word Similarity

How can we measure word similarity?

• Dictionary

• Ontology
Similarity using Dictionary

\[
\text{sim}(\text{whale, lobster}) = \left( \frac{\text{number of shared words}}{\text{number of unique words}} \right)
\]
Similarity Using Ontology

Class Mammalia

Order Artiodactyla
   Genus Giraffidae - giraffe
   Genus Bovidae - gazelle

Order Carnivora
   Genus Caniformia
   Genus Felidae - lion
Information Content

\[ IC(c) = -\log \frac{\text{# words that are equivalent to or are hyponyms of } c}{\text{# words in corpus}} \]

(Adapted from Lin. 1998. An information Theoretic Definition of Similarity. ICML.)
Word Similarity

How can we measure word similarity?

Distributional Information

Firth, 1957: “You shall know a word by the company it keeps”
fertility. Organ meats such as beef and chicken liver, tongue and hear
controlling scours. HOW TO FEED: BEEF AND DAIRY CALVES_ - 0.2 gram Dy
ing process discolors the treated beef and liquid accumulates in prepackag
say. He did say she could get her beef and vegetables in cans this summer
and feed efficiency of fattening beef animals. HOW TO FEED: At the
steaks, chops, chicken and prime beef as well as Tom's favorite dish, stu
ross from him was surmounted by a beef barrel with ends knocked out. In t
counter of boards laid across two beef barrels. There was, of course, no
Because Holstein cattle weren't a beef breed, they were rarely seen on a
2-5 grams of phenothiazine daily; beef calves-.5 to 1.5 grams daily depe
ties of this drug. HOW TO FEED: BEEF CATTLE (FINISHING RATION)_ - To
dairy cows and lesser amounts to beef cattle and poultry. About 90 percen
raises enough poultry, pigs, and beef cattle for most of their needs. Lo
on of liver abscesses in feed-lot beef cattle. Prevention of bacterial pne
pal feed bunk types for dairy and beef cattle: (1) Fence-line bunks- catt
es feed efficiency. HOW TO FEED: BEEF CATTLE_ - 10 milligrams of diet
the rations you are feeding your beef, dairy cattle, and sheep are adequa
itive business more profitable for beef, dairy, and sheep men. The tar
o bear. She was ready to kill the beef, dress it out, and with vegetables
. She had raised a calf, grown it beef-fat. She had, with her own work-wea
with feeding low-moisture corn in beef feeding programs. Several firms ar
he shelf life (at 35 F) of fresh beef from 5 days to 5 or 6 weeks. Howeve
canned pork products. Tests with beef have been largely unsuccessful beca
for eggs, pigs to eat garbage, a beef herd and wastes of all kinds. Separ
their money's worth. A goodmany beef-hungry settlers were accepting the
chicken. Acceptance of radiopasteurization

torehouse". Glendora dropped a chicken and a flurry of feathers, and went
will specialize in steaks, chops, chicken and prime beef as well as Tom's fa
ard as the one concerned with the chicken and the egg. Which came first? Is
he millions of buffalo and prairie chicken "!
"Come on, there's some cold chicken and we'll see what else". They wen
ves to extend the storage life of chicken at a low cost of about 0.5 cent per
CHICKEN CADILLAC# Use one 6-ounce chicken breast for each guest. Salt and pe
sodium ion juice, to about half cover the chicken breasts. Bake slowly at least one-
d, in butter. Sprinkle over top of chicken breasts. Serve each breast on a th
around, they had a hard time". #CHICKEN CADILLAC# Use one 6-ounce chicken
successful, and the shelf life of chicken can be extended to a month or more
ay from making a cake, building a chicken coop, or producing a book, to found
, they decided, but a deck full of chicken coops and pigpens was hardly suita
im. "Johnny insisted on cooking a chicken dinner in my honor- he's always bee
utes. Kid Ory, the trombonist chicken farmer, is also one of the solid a
y Johnson reaching around the wire chicken fencing, which half covered the tr
yes glittering behind dull silver chicken fencing. "That was Tee-wah I was t
wine in the pot roast or that the chicken had been marinated in brandy, and
yed this same game and called it "Chicken". He could not go through the f
f the Mexicans hiding in a little chicken house had passed through his head,
'll never forget him cleaning the chicken in the tub". A story, no doubt
. Organ meats such as beef and chicken liver, tongue and heart are planne
p. "Miss Sarah, I can't cut up no chicken. Miss Maude say she won't". Aga
pot. "What is it"? he asked. "Chicken", Mose said, and theatrically lick
im"? Adam shook his head. "Chicken", Mose said. She was a child too m
## Context Vectors

<table>
<thead>
<tr>
<th></th>
<th>arts</th>
<th>boil</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>0</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>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>digital</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>information</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

**Figure 20.9** Co-occurrence vectors for four words, computed from the Brown corpus, showing only 8 of the (binary) dimensions (hand-picked for pedagogical purposes to show discrimination). Note that *large* occurs in all the contexts and *arts* occurs in none; a real vector would be extremely sparse.
Hypothetical Counts based on Syntactic Dependencies

<table>
<thead>
<tr>
<th></th>
<th>Modified-by-ferocious(adj)</th>
<th>Subject-of-devour(v)</th>
<th>Object-of-pet(v)</th>
<th>Modified-by-African(adj)</th>
<th>Modified-by-big(adj)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lion</td>
<td>15</td>
<td>5</td>
<td>0</td>
<td>6</td>
<td>15</td>
</tr>
<tr>
<td>Dog</td>
<td>7</td>
<td>3</td>
<td>8</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>Cat</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Elephant</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>…</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Weighting the Counts

Point-wise Mutual Information

$$\text{PMI}(w, f) = \log_2 \frac{p(w, f)}{p(w) \times p(f)} = \log_2 \frac{N \times \text{count}(w, f)}{\text{count}(w) \times \text{count}(f)}$$

Positive Point-wise Mutual Information

$$\text{P-PMI} (w, f) = \text{PMI} (w, f), \text{ if PMI} (w, f) > 0$$
$$= 0, \text{ otherwise}$$
Singular Value Decomposition

\[ A = U D V^T \]
Word Vector Construction

1. Get word co-occurrence counts
2. Choose a weighting scheme
3. Use dimension reduction
Using Word Vectors

• Euclidean Distance

\[ d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2} = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}. \]

• Cosine Similarity

\[ \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}. \]
# Distributionally Similar Words

**Rum**
- vodka
- cognac
- brandy
- whisky
- liquor
- detergent
- cola
- gin
- lemonade
- cocoa
- chocolate
- scotch
- noodle
- tequila
- juice

**Write**
- read
- speak
- present
- receive
- call
- release
- sign
- offer
- know
- accept
- decide
- issue
- prepare
- consider
- publish

**Ancient**
- old
- modern
- traditional
- medieval
- historic
- famous
- original
- entire
- main
- indian
- various
- single
- african
- japanese
- giant

**Mathematics**
- physics
- biology
- geology
- sociology
- psychology
- anthropology
- astronomy
- arithmetic
- geography
- theology
- hebrew
- economics
- chemistry
- scripture
- biotechnology

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

The diagram includes words such as:
- Antonyms in red: foul, hideous, horrid, disgusting, repulsive, beastly
- Synonyms in green: fair, ordinary, poor, excellent, handsome, charming, cute, beautiful, wonderful, pleasing

The positions of the words on the graph suggest their relative meanings and connotations.
Gender