Administration

Final exam room announcement
Project forthcoming
Midterm returned hopefully next week

Homework 2
- Minimum Value: 32.00
- Maximum Value: 100.00
- Average: 82.60
- Median: 91.50

Vector Semantics
(Chapter 6)
What do words mean?

First thought: look in a dictionary

Words, Lemmas, Senses, Definitions

**pepper, n.**

**Forms:** Brit. / pəpər /, U.S. / pəpər /

**Pronunciation:** pepper (n.)

**Frequency:** (in current use)

**Etymology:** A borrowing from Latin, Classical Latin: piper, a meat or a kind of pea (as in toilet; pipe; pipo), compare Sans.

1. **The spice, or the plant.**
   
   1. A hot pungent spice derived from the prepared fruits (peppercorns) of the pepper plant, Piper nigrum (see sense 2a). Also, from the same species: black pepper, produced from black peppercorns; and the same (spice) as the pepper, is produced from white peppercorns—see sense 2b. Also, sweet pepper, jalepeño, etc.; used as a noun and as an adjective.

2. The plant Piper nigrum (family Piperaceae), a climbing shrub indigenous to South Asia and also cultivated elsewhere in the tropics, which has alternate stalked entire leaves, with pendulous spikes of small green flowers opposite the leaves, succeeded by small berries turning red when ripe; the source of black pepper. Also, more widely, any plant of the genus Piper or the family Piperaceae.

3. With distinguishing word: any of numerous plants of other families having hot pungent fruits or leaves which resemble pepper in taste and in some cases are used as a substitute for it.

- **U.S.** The California pepper tree, Schinus molle. Cf. PEPPER TREE n. 6.

- A variety of various forms of Capsicum, esp. Capsicum annuum var. annuum. Originally (chiefly with distinguishing word): any variety of the Capsicum annuum group, with elongated fruits having a hot, pungent taste, the source of cayenne, chilli powder, paprika, etc., or of the perennial C. frutescens, the source of Tabasco sauce. Now frequently (more fully sweet pepper): any variety of the C. annuum Grossum group, with large, bell-shaped or apple-shaped, mild-flavoured fruits, usually ripening to red, orange, or yellow and eaten raw in salads or cooked as a vegetable. Also: the fruit of any of these capricums.

- Sweet peppers are often used in their green immature state (more fully green pepper), but some new varieties remain green when ripe.
Lemma pepper

Sense 1: spice from pepper plant
Sense 2: the pepper plant itself
Sense 3: another similar plant (Jamaican pepper)
Sense 4: another plant with peppercorns (California pepper)
Sense 5: *capsicum* (i.e. chili, paprika, bell pepper, etc)

A sense or “concept” is the meaning component of a word
There are relations between senses

Relation: Synonymity

Synonyms have the same meaning in some or all contexts.

- filbert / hazelnut
- couch / sofa
- big / large
- automobile / car
- vomit / throw up
- Water / H₂O
Relation: Synonymity

Note that there are probably no examples of perfect synonymy.
- Even if many aspects of meaning are identical
- Still may not preserve the acceptability based on notions of politeness, slang, register, genre, etc.

The Linguistic Principle of Contrast:
- Difference in form -> difference in meaning

Relation: Synonymity?

Water/H₂O
Big/large
Brave/courageous
Relation: Antonymy

Senses that are opposites with respect to one feature of meaning

Otherwise, they are very similar!
- dark/light
- short/long
- fast/slow
- rise/fall
- hot/cold
- up/down
- in/out

More formally: antonyms can
- define a binary opposition
- or be at opposite ends of a scale
- long/short, fast/slow
- Be reversives:
- rise/fall, up/down

Relation: Similarity

Words with similar meanings. Not synonyms, but sharing some element of meaning

- car, bicycle
- cow, horse
Ask humans how similar 2 words are

<table>
<thead>
<tr>
<th>word1</th>
<th>word2</th>
<th>similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>vanish</td>
<td>disappear</td>
<td>9.8</td>
</tr>
<tr>
<td>behave</td>
<td>obey</td>
<td>7.3</td>
</tr>
<tr>
<td>belief</td>
<td>impression</td>
<td>5.95</td>
</tr>
<tr>
<td>muscle</td>
<td>bone</td>
<td>3.65</td>
</tr>
<tr>
<td>modest</td>
<td>flexible</td>
<td>0.98</td>
</tr>
<tr>
<td>hole</td>
<td>agreement</td>
<td>0.3</td>
</tr>
</tbody>
</table>

SimLex-999 dataset (Hill et al., 2015)

Relation: Word relatedness

Also called "word association"

Words be related in any way, perhaps via a semantic frame or field

- car, bicycle: similar
- car, gasoline: related, not similar
Semantic field

Words that
◦ cover a particular semantic domain
◦ bear structured relations with each other.

hospitals
  surgeon, scalpel, nurse, anaesthetic, hospital
restaurants
  waiter, menu, plate, food, menu, chef
houses
  door, roof, kitchen, family, bed

Relation: Superordinate/subordinate

One sense is a **subordinate** of another if the first sense is more specific, denoting a subclass of the other
◦ car is a subordinate of vehicle
◦ mango is a subordinate of fruit

Conversely **superordinate**
◦ vehicle is a superordinate of car
◦ fruit is a subordinate of mango

<table>
<thead>
<tr>
<th>Superordinate</th>
<th>vehicle</th>
<th>fruit</th>
<th>furniture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subordinate</td>
<td>car</td>
<td>mango</td>
<td>chair</td>
</tr>
</tbody>
</table>
These levels are not symmetric

One level of category is distinguished from the others

The "basic level"

Name these items
Cluster of Interactional Properties

Basic level things are “human-sized”

Consider chairs

- We know how to interact with a chair (sitting)
- Not so clear for superordinate categories like furniture
- “Imagine a furniture without thinking of a bed/table/chair/specific basic-level category”
The basic level

Is the level of distinctive actions
Is the level which is learned earliest and at which things are first named
It is the level at which names are shortest and used most frequently

Connotation

Words have **affective** meanings
positive connotations (*happy*)
negative connotations (*sad*)

positive evaluation (*great*, *love*)
negative evaluation (*terrible*, *hate*).
**Concepts** or word senses so far

Have a complex many-to-many association with **words**
- Homonymy
  - Lexemes share a form (phonological, orthographic) but have unrelated meanings
    - Bank (financial institution) versus bank (riverside)
    - Polysemy (multiple senses)
      - A single lexeme with multiple related meanings (bank the building, bank the financial institution)
        - The bank is constructed from red brick
        - I withdrew the money from the bank

Have relations with each other
- Synonymy
- Antonymy
- Similarity
- Relatedness
- Superordinate/subordinate
- Connotation

**But how to define a concept?**
Classical (“Aristotelian”) Theory of Concepts

The meaning of a word:

**a concept defined by necessary and sufficient conditions**

A *necessary* condition for being an X is a condition C that X must satisfy in order for it to be an X.

- If not C, then not X
- “Having four sides” is necessary to be a square.

A *sufficient* condition for being an X is condition such that if something satisfies condition C, then it must be an X.

- If and only if C, then X
- The following necessary conditions, jointly, are sufficient to be a square
  - x has (exactly) four sides
  - each of x’s sides is straight
  - x is a closed figure
  - x lies in a plane
  - each of x’s sides is equal in length to each of the others
  - each of x’s interior angles is equal to the others (right angles)
  - the sides of x are joined at their ends

Example from Norman Swartz, SFU

Problem 1: The features are complex and may be context-dependent

William Labov. 1975

Cup or bowl?
Problem 1: The features are complex and may be context-dependent

William Labov. 1975

What are these?
Cup or bowl?

The category depends on complex features of the object (diameter, etc)

Where does the category "cup" end?
The category depends on the context! (If there is food in it, it’s a bowl)

Boundaries between cups and bowls are context sensitive

Labov’s definition of cup

The term *cup* is used to denote round containers with a ratio of depth to width of 1 ≤ r where \( r = \alpha_1 + \alpha_2 + \ldots + \alpha_n \) and \( \alpha_1 \) is a positive quality when the feature 1 is present and 0 otherwise.

feature
1 = with one handle
2 = made of opaque vitreous material
3 = used for consumption of food
4 = used for the consumption of liquid food
5 = used for consumption of hot liquid food
6 = with a saucer
7 = tapering
8 = circular in cross-section

*Cup* is used variably to denote such containers with ratios width to depth 1 ≤ r where \( r_0 \leq r \leq 1 \) with a probability of \( r_1 - r_0 - r_b \). The quantity \( 1 = r_0 \) expresses the distance from the modal value of width to height.
Ludwig Wittgenstein (1889-1951)

Philosopher of language

In his late years, a proponent of studying “ordinary language”

Wittgenstein (1945)
Philosophical Investigations.
Paragraphs 66, 67

66. Consider for example the proceedings that we call “games”: I mean board-games, card-games, ball-games, Olympic games, and so on. What is common to them all? Don’t say: “There must be something common, or they would not be called “games”” — but look and see whether there is anything common to all. — For if you look at them you will see something that is common to all, but not a relation, nor a whole series of relations, and a whole series of things at that. To repeat don’t think, but look — Look for example at board-games, with their mutability relationships. Now pass to card-games; here you find many correspondences with the first group, but many common features drop out, and others appear. When we pass next to ball-games, much that is common is retained, but much is lost. — Are they all “amusing”? Compare chess withapeut and croquet. Or is there always winning and losing, or competition between players? Think of pistons. In ball-games there is winning and losing but when a child throws his ball at the wall and catches it again, the feature has disappeared. Look at the parts played by skill and luck and at the difference between skill in chess and skill in tennis. Think now of games like ring-a-ring-o’-roses; here is the element of assurance, but how many other characteristic features have disappeared? And we can go through the many, many other groups of games in the same way: we can see similarities crop up and disappear.

And the result of this examination is: we see a complicated network of similarities overlapping and crossing; sometimes overall similarities, sometimes similarities of detail.

67. I use think of no better expression to characterize these similarities than “family resemblance”; for the various likenesses between members of a family build, features of eyes, hair, temperament, etc., etc., overlap and enhance in the same way.

And I shall say: “games” from a family. And for instance the kinds of number form a family in the same way.

Why do we call something a “number”? Well, perhaps because it has a — let us say — relationship with several things that have hitherto been called number; and this can be said to give it an indirect relationship to other things we call the same name. And we catch out concept of number as in spinning a thread we twist fibre on fibre. And the enough of the thread does not make in the fact that some one fibre runs through its whole length, but in the overlapping of many fibres.

But if someone wishes to say: “There is something common to all these concepts — namely the distinction of all their common properties” — I should reply: Now you are only playing with words. One might as well say: “Something runs through the whole thread — namely the continuous pervasion of these fibres.”
What is a game?

Wittgenstein’s thought experiment on "What is a game":

PI #66:
"Don’t say “there must be something common, or they would not be called `games’”—but look and see whether there is anything common to all”

Is it amusing?
Is there competition?
Is there long-term strategy?
Is skill required?
Must luck play a role?
Are there cards?
Is there a ball?
Family Resemblance

<table>
<thead>
<tr>
<th>Game 1</th>
<th>Game 2</th>
<th>Game 3</th>
<th>Game 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC</td>
<td>BCD</td>
<td>ACD</td>
<td>ABD</td>
</tr>
</tbody>
</table>

“each item has at least one, and probably several, elements in common with one or more items, but no, or few, elements are common to all items”  Rosch and Mervis

How about a radically different approach?
Ludwig Wittgenstein

PI #43: "The meaning of a word is its use in the language"

Let's define words by their usages

In particular, words are defined by their environments (the words around them)

Zellig Harris (1954): If A and B have almost identical environments we say that they are synonyms.
What does ong choi mean?

Suppose you see these sentences:
• Ong choi is delicious sautéed with garlic.
• Ong choi is superb over rice
• Ong choi leaves with salty sauces

And you've also seen these:
• ...spinach sautéed with garlic over rice
• Chard stems and leaves are delicious
• Collard greens and other salty leafy greens

Conclusion:
◦ Ongchoi is a leafy green like spinach, chard, or collard greens

Ong choi: *Ipomoea aquatica*
"Water Spinach"
We'll build a new model of meaning focusing on similarity

Each word = a vector
  ◦ Not just "word" or word45.

Similar words are "nearby in space"

We define a word as a vector

Called an "embedding" because it's embedded into a space

The (fairly recent) standard way to represent meaning in NLP

Fine-grained model of meaning for similarity
  ◦ NLP tasks like sentiment analysis
  ◦ With words, requires same word to be in training and test
  ◦ With embeddings: ok if similar words occurred!!!
We'll introduce 2 kinds of embeddings

**Tf-idf**
- A common baseline model
- Sparse vectors
- Words are represented by a simple function of the counts of nearby words

**Word2vec**
- Dense vectors
- Representation is created by training a classifier to distinguish nearby and far-away words

Words, vectors, and co-occurrence matrices
Term-document matrix

Each document is represented by a vector of words

<table>
<thead>
<tr>
<th></th>
<th>As You Like It</th>
<th>Twelfth Night</th>
<th>Julius Caesar</th>
<th>Henry V</th>
</tr>
</thead>
<tbody>
<tr>
<td>battle</td>
<td>1</td>
<td>0</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>good</td>
<td>14</td>
<td>80</td>
<td>62</td>
<td>89</td>
</tr>
<tr>
<td>fool</td>
<td>36</td>
<td>58</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>wit</td>
<td>20</td>
<td>15</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Visualizing document vectors

Henry V [4,13]  
Julius Caesar [1,7]  
As You Like It [36,1]  
Twelfth Night [58,0]
Vectors are the basis of information retrieval

- “Information retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers).”  *Stanford NLP*
- Vectors are similar for the two comedies and different than the history
- Comedies have more fools and wit and fewer battles.

<table>
<thead>
<tr>
<th></th>
<th>As You Like It</th>
<th>Twelfth Night</th>
<th>Julius Caesar</th>
<th>Henry V</th>
</tr>
</thead>
<tbody>
<tr>
<td>battle</td>
<td>1</td>
<td>0</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>good</td>
<td>14</td>
<td>80</td>
<td>62</td>
<td>89</td>
</tr>
<tr>
<td>fool</td>
<td>36</td>
<td>58</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>wit</td>
<td>20</td>
<td>15</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Words can be vectors too

- *battle* is "the kind of word that occurs in Julius Caesar and Henry V"
- *fool* is "the kind of word that occurs in comedies, especially Twelfth Night"
More common: word-word matrix (or "term-context matrix")

Two words are similar in meaning if their context vectors are similar.
Reminders from linear algebra

\[
dot\text{-product}(\vec{v}, \vec{w}) = \vec{v} \cdot \vec{w} = \sum_{i=1}^{N} v_i w_i = v_1 w_1 + v_2 w_2 + \ldots + v_N w_N
\]

vector length \( |\vec{v}| = \sqrt{\sum_{i=1}^{N} v_i^2} \)

Cosine for computing similarity

\[
cosine(\vec{v}, \vec{w}) = \frac{\vec{v} \cdot \vec{w}}{|\vec{v}| |\vec{w}|} = \frac{\sum_{i=1}^{N} v_i w_i}{\sqrt{\sum_{i=1}^{N} v_i^2} \sqrt{\sum_{i=1}^{N} w_i^2}}
\]

\( v_i \) is the count for word \( v \) in context \( i \)
\( w_i \) is the count for word \( w \) in context \( i \).
Cosine as a similarity metric

-1: vectors point in opposite directions
+1: vectors point in same directions
0: vectors are orthogonal

Frequency is non-negative, so cosine range 0-1

\[
\cos(\mathbf{v}, \mathbf{w}) = \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\| \cdot \|\mathbf{w}\|} = \frac{\sum_{i=1}^{N} v_i w_i}{\sqrt{\sum_{i=1}^{N} v_i^2} \sqrt{\sum_{i=1}^{N} w_i^2}}
\]

Which pair of words is more similar?

\[
\text{cosine(apricot,information)} = \frac{1+0+0}{\sqrt{1+0+0} \sqrt{1+36+1}} = \frac{1}{\sqrt{38}} = .16
\]

\[
\text{cosine(digital,information)} = \frac{0+6+2}{\sqrt{0+1+4} \sqrt{1+36+1}} = \frac{8}{\sqrt{38} \sqrt{5}} = .58
\]

\[
\text{cosine(apricot,digital)} = \frac{0+0+0}{\sqrt{1+0+0} \sqrt{0+1+4}} = 0
\]
Visualizing cosines (well, angles)

But raw frequency is a bad representation

Frequency is clearly useful; if sugar appears a lot near apricot, that's useful information.

But overly frequent words like the, it, or they are not very informative about the context

Need a function that resolves this frequency paradox!
tf-idf: combine two factors

**tf:** term frequency. frequency count (usually log-transformed):

\[
tf_{t,d} = \begin{cases} 
1 + \log_{10} \text{count}(t, d) & \text{if } \text{count}(t, d) > 0 \\
0 & \text{otherwise}
\end{cases}
\]

**Idf:** inverse document frequency: tf-

\[
idf_i = \log \left( \frac{N}{df_i} \right)
\]

- Words like "the" or "good" have very low idf
- Total # of docs in collection
- # of docs that have word i

tf-idf value for word t in document d:

\[
w_{t,d} = tf_{t,d} \times idf_t
\]

**Summary:** tf-idf

Compare two words using tf-idf cosine to see if they are similar
An alternative to tf-idf

Ask whether a context word is particularly informative about the target word.

- **Positive Pointwise Mutual Information (PPMI)**

Pointwise Mutual Information

**Pointwise mutual information:**
Do events \( x \) and \( y \) co-occur more than if they were independent?

\[
\text{PMI}(X, Y) = \log_2 \frac{P(x, y)}{P(x)P(y)}
\]

**PMI between two words:** (Church & Hanks 1989)
Do words \( x \) and \( y \) co-occur more than if they were independent?

\[
\text{PMI}(\text{word}_1, \text{word}_2) = \log_2 \frac{P(\text{word}_1, \text{word}_2)}{P(\text{word}_1)P(\text{word}_2)}
\]
<table>
<thead>
<tr>
<th></th>
<th>computer</th>
<th>data</th>
<th>pinch</th>
<th>result</th>
<th>sugar</th>
</tr>
</thead>
<tbody>
<tr>
<td>apricot</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>pineapple</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>digital</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>information</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>computer</th>
<th>data</th>
<th>pinch</th>
<th>result</th>
<th>sugar</th>
</tr>
</thead>
<tbody>
<tr>
<td>apricot</td>
<td>-</td>
<td>-</td>
<td>2.25</td>
<td>-</td>
<td>2.25</td>
</tr>
<tr>
<td>pineapple</td>
<td>-</td>
<td>-</td>
<td>2.25</td>
<td>-</td>
<td>2.25</td>
</tr>
<tr>
<td>digital</td>
<td>1.66</td>
<td>0.00</td>
<td>-</td>
<td>0.00</td>
<td>-</td>
</tr>
<tr>
<td>information</td>
<td>0.00</td>
<td>0.57</td>
<td>0.47</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Summary

- Survey of Lexical Semantics
- Idea of Embeddings: Represent a word as a function of its distribution with other words
- Tf-idf
- Cosines
- PPMI (barely)