Vectoral Models of Meaning

Last time we saw some attractive properties of feature-based models.

▶ sentence ‘is a chicken a meteor?’ produces quick ‘no’

Problem: deciding what counts as a feature is subjective, hard to reproduce.

Solution: use co-occurrences in sentences/texts as features:

▶ ‘violin’ co-occurs with ‘orchestra’ a lot
▶ ‘cello’ co-occurs with ‘orchestra’ a lot
▶ → ‘violin’ and ‘cello’ must be similar! (correct)
▶ holds even if ‘violin’ and ‘cello’ do not ever co-occur!

Can use this to quantify concept similarity and predict confusability!

Can handle cases like synonyms, but not so good with homonyms:

▶ ‘river’ co-occurs with ‘bank’ a lot
▶ ‘investment’ co-occurs with ‘bank’ a lot
▶ → ‘river’ and ‘investment’ must be similar! (nope)
Hyperspace Analogue to Language (HAL)

But there are a lot of co-occurrences!

Kevin Lund & Curt Burgess ’96 (HAL): use matrix computer data struct.

- trained on 2 million words from USENET internet chat forums
- examined 70,000 content words (like ‘deer’, ‘river’,... not ‘the’, ‘of’,...)
- stored counts in matrix with 70,000 by 70,000 (= 5 billion) cells
- store total counts of each pair of words, multiplied by proximity: (10=adjacent, 1=nine words apart)
- matrix is sparse: most counts are zero
- compute difference/similarity using angle of vectors for words (i.e. use proportions of co-occurrences rather than absolute counts)
- specifically, compute as cosine of angle of word vectors (very sensitive at smaller differences, less sensitive at larger)
- correlate with human reaction times in a single-word priming experiment
Latent Semantic Analysis (LSA)

Because it is sparse, HAL has lots of orthogonal (zero overlap) concepts

Thomas Landauer & Susan Dumais ’96 (LSA): use dimensionality reduction

▶ trained on 5 mil. words from Grolier’s Academic American Encyclopedia
▶ examined 60,000 content words in 30,000 articles
▶ stored counts in matrix with 60,000 by 30,000 (= 2 billion) cells
▶ store count of each word in each text
▶ each word still has a useful vector, just of documents rather than words
▶ this matrix is still \textit{sparse}: most counts are zero
▶ use ‘single value decomposition’ to find 300 most predictive dimensions
  ▶ regress a line with minimal sum of square of distance to each points
  (use \textit{square} of distance because otherwise many lines are acceptable)
  ▶ smash data together along this dimension
  ▶ repeat for resulting smashed space
▶ equivalent to factoring matrix into product of three matrices
Latent Semantic Analysis (LSA)

Thomas Landauer & Susan Dumais ’96 (LSA), cont’d:

- result is a ‘suspended hyperplane’ within a hyperspace
- fewer dimensions makes it more dense or ‘smooth’ – no more zeros!
  - pooling counts across many different words
- but, resulting dimensions may not be coherent anymore:
  - e.g.: big/small, nice/mean: smurf, BFG, Stewie, C’thulhu
  - rotating axes gives you smurf/C’thulhu and BFG/Stewie (interpret?)
  - sim. to actual neurons: Halle Berry (Catwoman, or spelled out H,A,...)
- distributed, like neural network
  - multi-use dimensions lets model organize useful axes wherever needed
  - robust to damage/loss of dimension values (cortical columns?)
  - loss of dimensions makes it more easily confused (if not sparse to begin)
- in fact, equivalent to three-layer neural network w. no thresholding
  - weights on connections between layers define factored matrices
- compute difference/similarity using cosine metric, as with HAL
- comparable to human performance on synonym part of TOEFL test!
But what about sensory / kinesthetic meaning?
John Searle (1980): ‘Chinese room’
Paul & Patricia Churchland (1991): ‘luminous room’
Glenberg & Robertson ’00: HAL/LSA don’t incorporate affordances
- stimuli: ‘... fill sweater with leaves/water’
- measure: acceptability
- results: leaves is more acceptable (contra HAL/LSA)

Associations with goals, not just words?

Sensory/kinesth. meaning can be accounted as features (e.g. motor cortex):
- Pulvermüller&al’05: transcranial magnetic stimulation (TMS)
  - stimuli: TMS stimulation to hand/foot are of motor cortex, present word
  - measure: reaction time in lexical decision task
  - results: left hemisphere TMS at hand(foot) facilitates hand/foot words
- Buccino&al’05: transcranial magnetic stimulation (TMS)
  - stimuli: TMS stimulation to hand/foot are of motor cortex, present word
  - measure: motor evoked potential (MEP) at hand/foot neuromusc. jn.
  - results: TMS reduces MEP at hands/feet for hand/foot words
For next time... 

Read:
- Traxler ch 3, pp. 97–116