Incorporating Connotation of Meaning into Models of Semantic Representation: An Application in Text Corpus Analysis

Shane T. Mueller, Ph.D.
Klein Associates Division, ARA Inc.
Fairborn, OH
smueller@ara.com
smueller@decisionmaking.com

and

Rich Shiffrin, Ph.D.
Indiana University
Klein Associates Division (KAD)

Background:

- Originally Klein Associates Inc., founded in 1978, acquired by ARA in Sep-05
- 50+ employees
- Customers include DOD and Commercial
Applied Research Associates (ARA)

Company Background

- Founded in 1979, HQ’d in Albuquerque, NM
- Employee owned company
- 17 Divisions across the country
- 1,100+ Employees at locations in US and Canada
- FY06 Sales of ~$170M
- Sustained 27 years of double-digit growth
Background

- Semantic Knowledge forms, and is formed by, meaningful episodes.
- Meaning can be driven by Contextual Semantics:
  - Meaning can be inferred through context.
  - Meaning can depend upon context
REM-II: A Bayesian model of Episodic Memory and Semantic Knowledge

- New developments of a Bayesian feature-based model of episodic memory.
- Incorporate assumptions about how contextual semantic information is learned and used.
- If assumptions are accurate, the model should be able to “read” text corpora and develop meaningful representations.
Traditional Vector-space Approaches (e.g., LSA)

- Meaning for a word is a point in space.
- Geometry is a bad model of semantics (e.g. Tversky, 1975): the dimensions (features) matter.
- This **prototype** approach succeeds at synonymy, but fails to capture polysemy.
Prototypes: The good and the bad

- Allow multiple poor exemplars to accrue information
- Provide an encapsulated representation for fast interpretation of world

BUT....

- Completely misses important aspect of **Contextual Semantics**: connotation of meaning

For example:
- Homonyms/homographs: bank
- Polysemes: mouth
- Connotation or aspect: kitchen
Beyond the prototype

- Prototypes use a single 'average' representation.
- We store feature co-occurrence, rather than feature occurrence.

[ <5 3 | 3 4 | 1 0 0> ]

- Each cell is the co-occurrence of features within experienced exemplars.

<3 6 | 0 1 | 1 3 0>  
<3 0 | 4 3 | 0 0 3>  
<4 1 | 3 5 | 4 2 1>  

- Each row is a conditional prototype, the average representation when some feature was present.

<1 1 | 0 4 | 6 0 1>  
<0 3 | 0 2 | 0 4 1>  
<0 0 | 3 1 | 1 1 4> ]

- Allows contextually distinct representations to be segregated (polysemy, connotation, etc.)
Conditional Prototypes

• Each concept is a set of conditional prototypes (conditioned on the presence of some feature).

• Number of prototypes $=$ number of features
Expanding the Realm of Possibility

Forming Co-occurrence matrix

- Any set of features can produce a co-occurrence matrix.

- 110011 and 001100 form co-occurrence matrix at right

- Knowledge accrues by incorporating co-occurrence matrix from individual episodes into current knowledge structure.
LTM Co-occurrence matrix

- Over time, complex matrix representation will form.
- Word has two primary concepts 1-2-5-6 and 3-4.
- 1256 is stronger than 34
Encoding Episodes from Knowledge Matrix

- Requires sampling features from knowledge matrix
- Generic Encoding: pick row, pick feature, pick new row based on what has been sampled, repeat.
- Biased Encoding: pick row from another trace, pick feature from that row of current matrix, repeat.
  - (Meaning can depend upon context)
Applications

- Laboratory Memory Phenomena (Will not discuss):
  - Frequency Effects in Episodic Memory
  - Forward bias CRP functions in free recall
  - Perceptual learning

- Corpus Analysis: The “Real” World?
  - REM-II can 'read' a text corpus and develop semantic spaces similar to other methods.
  - Multi-language semantic spaces.
  - Techniques for incorporating integral semantics into contextual semantic representations

- Beyond linguistic applications
Operation of Model

- Identify each word in statement
- Encode features for each word
  - (New words start with random feature sets)
- Evaluate likelihood of features versus base rate
- Create semantic composite
- Compute co-occurrence matrix
- Add features into each trace
- Each sentence is used as a context or document (running windows, decay, etc. are also possible)
Demo 1: Toy Problem

- Four probabilistic contexts:
  - $A_1 A_2 P_1$
  - $A_1 A_2 P_2$
  - $B_1 B_2 P_1$
  - $B_1 B_2 P_2$

- 20 Features, 5000 iterations for learning
- Encoded 100 episodes from each prototype (both biased and unbiased)
- Performed MDS in common space for visualization
Expanding the Realm of Possibility

Group A versus Group B episodic encodings

Non co-occurrent words (Words 4 and 8)

Polysemous Word 4 biased toward A or B

Polysemous Word 8 biased toward A or B

A Bias

B Bias
The GAC Corpus

- Product of the MindPixel project
  - Collaborative internet-based project to produce database of millions of human validated true/false statements.
  - Active from 2000 to 2006
  - Similar to OpenMind Common Sense (MIT)
- Released a set of 80,000 validated statements:
  - 660,000 words
  - 29,000 unique tokens
  - Stemming/stop-word removal reduced to 269,000/11,859.
- Rich yet broad source of knowledge.

- Used 40 features, read corpus multiple times, randomized order of statements.
Some results: “FLY”
Comparison to LSA
### Top ten most similar to probes

<table>
<thead>
<tr>
<th>color</th>
<th>food</th>
<th>europe</th>
<th>man</th>
<th>fly</th>
<th>fire</th>
<th>car</th>
<th>earth</th>
<th>animal</th>
</tr>
</thead>
<tbody>
<tr>
<td>color</td>
<td>food</td>
<td>europe</td>
<td>man</td>
<td>fly</td>
<td>fire</td>
<td>car</td>
<td>earth</td>
<td>animal</td>
</tr>
<tr>
<td>blue</td>
<td>eat</td>
<td>italy</td>
<td>woman</td>
<td>flap</td>
<td>match</td>
<td>drive</td>
<td>around</td>
<td>breath</td>
</tr>
<tr>
<td>violet</td>
<td>cereal</td>
<td>locate</td>
<td>physically</td>
<td>flame</td>
<td>wise</td>
<td>motorcycle</td>
<td>worm</td>
<td></td>
</tr>
<tr>
<td>combination</td>
<td>restaurant</td>
<td>portugal</td>
<td>attractive</td>
<td>touch</td>
<td>term</td>
<td>riding</td>
<td>predator</td>
<td>usually</td>
</tr>
<tr>
<td>red</td>
<td>rice</td>
<td>germany</td>
<td>virgin</td>
<td>airplane</td>
<td>wise</td>
<td>driving</td>
<td>usually</td>
<td>human</td>
</tr>
<tr>
<td>orange</td>
<td>regularly</td>
<td>belgium</td>
<td>average</td>
<td>balloon</td>
<td>off</td>
<td>automobile</td>
<td>intelligent</td>
<td>eat</td>
</tr>
<tr>
<td>primary</td>
<td>restaurant</td>
<td>music</td>
<td>naked</td>
<td>lighter</td>
<td>hot</td>
<td>form</td>
<td>sun</td>
<td>cockroach</td>
</tr>
<tr>
<td>yellow</td>
<td>mouth</td>
<td>japan</td>
<td>crave</td>
<td>african</td>
<td>lightbulb</td>
<td>move</td>
<td>spherical</td>
<td></td>
</tr>
<tr>
<td>combine</td>
<td>order</td>
<td>spain</td>
<td>reach</td>
<td>pop</td>
<td>term</td>
<td>vehicle</td>
<td>rotate</td>
<td></td>
</tr>
<tr>
<td>purple</td>
<td>usually</td>
<td>japan</td>
<td></td>
<td>fan</td>
<td>off</td>
<td>consume</td>
<td>moon</td>
<td></td>
</tr>
</tbody>
</table>

---

*ARA - Expanding the Realm of Possibility*
Demonstration of Connotation

Orange

ORANGE
COLOR
JUICE
RED
BROWN
SHADE
PURPLE
FAVORITE
YELLOW
COMBINATION
DANGER
BLUE
COMBINE
PRIMARY
FRUIT
ASSOCIATE
RAGE
APPLE
MIX
PINK

Orange as Color
complementary ↑
favorite ↑
orange ↓
red =
color ↓
juice ↓
brown ↓
shade ↓
marmalade ↑
hue ↑
mix ↑
yellow ↓
purple ↓
blue ↓
icon ↑
rage ↑
bluebird ↑
primary ↓
pigment ↑
violet ↑
combine ↓

Orange as Fruit
orange =
color =
juice =
favorite ↑
marmalade ↑
yellow ↑
red ↓
ripe ↑
pomegranate ↑
combination =
purple ↓
brown ↓
either ↑
toot ↑
associate ↑
rose ↑
plum ↑
blue ↓
fruit ↓
shade ↓
lime ↑
Summary

- At its core, a model of semantic representation.
- Extends a **process model** of episodic memory model to produce representations akin to statistical language models.
- Contextual semantics: meaning is inferred through and depends upon context.
- We are looking at ways to incorporate **non-context-based features** (part-of-speech, language-of-origin, etc.)
- Such a system can be useful beyond linguistic applications