The Neural Basis of Thought and Language

Final Review Session
Administrivia

• Final in class next Tuesday, May 8th

• Be there on time!

• Format:
  - closed books, closed notes
  - short answers, no blue books

• Final paper due on bSpace on Friday, May 11
Resources

• Textbook!
• Class slides
• Section slides
• Joe Makin's class notes from last year
  - on notes page
Overview

- Bailey Model
  - feature structures
  - Bayesian model merging
  - recruitment learning
- KARMA
  - X-schema, frames
  - aspect
  - event-structure metaphor
  - inference
- Grammar Learning
  - parsing
  - construction grammar
  - learning algorithm
- SHRUTI
- FrameNet
- Bayesian Model of Human Sentence Processing
Important topics

- Regier's model of spatial relation learning
- Bailey's model of verb learning
- KARMA model of metaphor
- Binding and inference
  - SHRUTI, short signatures
- Grammars and learning
- ECG
  - Learning ECG
- Bayes nets
- Model merging, MDL
- Petri nets
- Language
  - Metaphor
  - Aspect
  - Grammars
  - Schemas
  - Frames
  - SemSpec
Q & A
Bayes Nets

• Bayes' Rule / Product Rule
  - $P(A|B) = P(A,B) / P(B)$
  - $P(A,B) = P(A|B) P(B)$
  - $P(B|A) = P(A|B) P(B) / P(A)$
  - All the same!

• Variables have distributions
• Variables depend on other variables
Regier's model

- Learn spatial relation terms
  - e.g. in, on, above
- Neural network + hand-designed “vision” parts
Bailey's model

- Verb learning
- Learn parameters matched to words
  - word senses
  - can connect to simulator
- Model merging!
<table>
<thead>
<tr>
<th>schema</th>
<th>elbow jnt</th>
<th>posture</th>
<th>accel</th>
</tr>
</thead>
<tbody>
<tr>
<td>slide 0.9</td>
<td>extend 0.9</td>
<td>palm 0.9</td>
<td>[6]- 8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>grasp 0.3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>schema</th>
<th>elbow jnt</th>
<th>posture</th>
<th>accel</th>
</tr>
</thead>
<tbody>
<tr>
<td>depress 0.9</td>
<td>fixed 0.9</td>
<td>index 0.9</td>
<td>[2]</td>
</tr>
</tbody>
</table>

### Data #1

<table>
<thead>
<tr>
<th>schema</th>
<th>elbow jnt</th>
<th>posture</th>
<th>accel</th>
</tr>
</thead>
<tbody>
<tr>
<td>slide</td>
<td>extend</td>
<td>palm</td>
<td>6</td>
</tr>
</tbody>
</table>

### Data #2

<table>
<thead>
<tr>
<th>schema</th>
<th>elbow jnt</th>
<th>posture</th>
<th>accel</th>
</tr>
</thead>
<tbody>
<tr>
<td>slide</td>
<td>extend</td>
<td>palm</td>
<td>8</td>
</tr>
</tbody>
</table>

### Data #3

<table>
<thead>
<tr>
<th>schema</th>
<th>elbow jnt</th>
<th>posture</th>
<th>accel</th>
</tr>
</thead>
<tbody>
<tr>
<td>depress</td>
<td>fixed</td>
<td>index</td>
<td>2</td>
</tr>
</tbody>
</table>

### Data #4

<table>
<thead>
<tr>
<th>schema</th>
<th>elbow jnt</th>
<th>posture</th>
<th>accel</th>
</tr>
</thead>
<tbody>
<tr>
<td>slide</td>
<td>extend</td>
<td>grasp</td>
<td>2</td>
</tr>
</tbody>
</table>
Computational Details

- complexity of model + ability to explain data
- maximum a posteriori (MAP) hypothesis

\[
\text{argmax}_m P(m | D) = \text{argmax}_m P(D | m) P(m) \quad \text{by Bayes' rule}
\]

wants the best model given data

how likely is the data given this model?

penalize complex models - those with too many word senses
Model merging

- Start with a simple model
- Merge to refine it
  - “Greedy” merges: reduce cost without thought for future
- Cost metric
  - prefer simple representation
  - prefer to explain data well
Metaphor

There are LOTS of metaphors we use

- Power is size
- Knowing is seeing
- Event structure is motion
Event Structure Metaphor

- States are Locations
- Changes are Movements
- Causes are Forces
- Causation is Forced Movement
- Actions are Self-propelled Movements
- Purposes are Destinations
- Means are Paths
- Difficulties are Impediments to Motion
- External Events are Large, Moving Objects
- Long-term, Purposeful Activities are Journeys
Ego Moving versus Time Moving

Fig. 1. (a) Schematic of the ego-moving schema used to organize events in time. (b) Schematic of the time-moving schema used to organize events in time.
## Results

<table>
<thead>
<tr>
<th>PRIME</th>
<th>Meeting is Monday</th>
<th>Meeting is Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ego Moving</td>
<td>26.7%</td>
<td>73.3%</td>
</tr>
<tr>
<td>Object Moving</td>
<td>69.2%</td>
<td>30.8%</td>
</tr>
</tbody>
</table>
KARMA simulator

- Invented by Carson Daly
- Allows metaphor understanding
  - Event structure metaphor
    - Source domain is Petri net
  - Target domain is Bayes net
  - Metaphor maps connect
KARMA

- DBN to represent target domain knowledge
- Metaphor maps link target and source domain
- X-schema to represent source domain knowledge
Temporal synchrony and SHRUTI

- **Binding problem**
  - bind properties to objects
  - don't mix them up!

- **Reflexive reasoning**
  - understand implied information
  - not conscious of this
SHRUTI

• entity, type, and predicate focal clusters

• An “entity” is a phase in the rhythmic activity.

• Bindings are synchronous firings of role and entity cells

• Rules are interconnection patterns mediated by coincidence detector circuits that allow selective propagation of activity

• An episode of reflexive processing is a transient propagation of rhythmic activity
“Harry walked to the café.”

- asserting that \( \text{walk}(\text{Harry}, \text{café}) \)
- Harry fires in phase with agent role
- café fires in phase with goal role
“Harry walked to the café.”

- asserting that walk(Harry, café)
- Harry fires in phase with agent role
- cafe fires in phase with goal role

entity

+Harry

+e
+v
cafe
?e
?v

type

predicates

+ walk
-?
agt
goal
Activation Trace for walk(Harry, café)
Alternative: short signatures

- entity: (+) Harry (?)
- type: (+e) cafe (+v) (+e) cafe (v) (+e)
- predicate: (+) walk (?) (+agt) (+goal)

= 010
= 110
Language

- Grammar
  - Syntax
- Tense
- Aspect
- Semantics
- Metaphor
- Simulation
- Unification
Computer-science style grammar

- Regular grammar
  - X -> a b c Y
- Context-free grammar
  - X -> a Y b Z W
“Harry walked into the café.”
The INTO construction

<table>
<thead>
<tr>
<th>construction INTO</th>
<th>subcase of Spatial-Relation form</th>
</tr>
</thead>
<tbody>
<tr>
<td>self \textsubscript{f} .orth ← “into”</td>
<td>meaning: Trajector-Landmark evokes Container as cont</td>
</tr>
<tr>
<td>evokes Source-Path-Goal as spg</td>
<td>trajector ↔ spg.trajector</td>
</tr>
<tr>
<td>landmark ↔ cont</td>
<td>cont.interior ↔ spg.goal</td>
</tr>
<tr>
<td>cont.exterior ↔ spg.source</td>
<td></td>
</tr>
</tbody>
</table>
Unification

- Basic idea: capture agreement and semantic features in feature structures

I
agreement

number : SG
person : 1st

Went
agreement

VP → Verb NP
VP.agreement ↔ Verb.agreement

Pat
agreement

number : SG
person : 3rd

Shop
agreement

number : person : 1st

- Enforce constraints on these features using unification rules

S → NP VP
NP.agreement ↔ VP.agreement
The Spatial-Phrase construction

construction SPATIAL-PHRASE
constructional constituents
   sr : Spatial-Relation
   lm : Ref-Expr
form
   \( sr_f \) before \( lm_f \)
meaning
   \( sr_m \cdot \text{landmark} \leftrightarrow lm_m \)
The Directed-Motion construction

construction DIRECTED-MOTION
constructional constituents
   a : Ref-Exp
   m: Motion-Verb
   p : Spatial-Phrase
form
   $a_f \ before \ m_f$
   $m_f \ before \ p_f$
meaning 
evokes Directed-Motion as dm
self_{m}.scene $\leftrightarrow$ dm
dm.agent $\leftrightarrow$ a_{m}
dm.motion $\leftrightarrow$ m_{m}
dm.path $\leftrightarrow$ p_{m}

schema Directed-Motion
roles
agent : Entity
motion : Motion
path : SPG
What exactly is simulation?

- Belief update and/or X-schema execution
“Harry walked into the café.”
“Harry is walking to the café.”
“Harry is walking to the café.”

walker=Harry
goal=cafe

ready
start
ongoing
interrupt
resume
suspended
iterate
finish
done
aborted
cancelled
WALK
“Harry has walked into the wall.”
Perhaps a different sense of INTO?

construction INTO
subcase of spatial-prep form
selfᵢ.orth ← “into”
meaning
evokes Trajector-Landmark as tl
evokes Container as cont
evokes Source-Path-Goal as spg
tl.trajector ↔ spg.trajector
tl.landmark ↔ cont
cont.interior ↔ spg.goal
cont.exterior ↔ spg.source

construction INTO
subcase of spatial-prep form
selfᵢ.orth ← “into”
meaning
evokes Trajector-Landmark as tl
evokes Impact as im
evokes Source-Path-Goal as spg
tl.trajector ↔ spg.trajector
tl.landmark ↔ spg.goal
im.obj1 ↔ tl.trajector
im.obj2 ↔ tl.landmark
“Harry has walked into the wall.”
Map down to timeline

ready → start → ongoing → finish → done

consequence

E

S

R
further questions?
How do you learn…

the meanings of spatial relations,

the meanings of verbs,

the metaphors, and

the constructions?
How do you learn...

the meanings of spatial relations,

the meanings of verbs,

the metaphors, and

the constructions?

That’s the Regier model.
How do you learn...

the meanings of spatial relations,

the meanings of verbs,

the metaphors, and

the constructions?

That's Bailey's model
How do you learn...

the meanings of spatial relations,

the meanings of verbs,

the metaphors, and

the constructions?

conflation hypothesis
(primary metaphors)
How do you learn...

the meanings of spatial relations,

the meanings of verbs,

the metaphors, and

the constructions?
Usage-based Language Learning

(Utterance, Situation) → Analyze → Partial Analysis → Comprehension

Constructions

→ Reorganize

(Comm. Intent, Situation) → Generate → Utterance → Production

Acquisition

Partial Analysis

→ Hypothesize

Hypothesize

→ Acquisition
Main Learning Loop

while <utterance, situation> available and cost > stoppingCriterion
    analysis = analyzeAndResolve(utterance, situation, currentGrammar);
    newCxns = hypothesize(analysis);
    if cost(currentGrammar + newCxns) < cost(currentGrammar)
        addNewCxns(newCxns);
    if (re-organize == true)  // frequency depends on learning parameter
        reorganizeCxns();
Three ways to get new constructions

• Relational mapping
  - throw the ball

• Merging
  - throw the block
  - throwing the ball

• Composing
  - throw the ball
  - ball off
  - you throw the ball off

\{ \text{THROW} < \text{BALL} \}

\{ \text{THROW} < \text{OBJECT} \}

\{ \text{THROW} < \text{BALL} < \text{OFF} \}
Minimum Description Length

• Choose grammar $G$ to minimize $\text{cost}(G | D)$:
  - $\text{cost}(G | D) = \alpha \cdot \text{size}(G) + \beta \cdot \text{complexity}(D | G)$
  - Approximates Bayesian learning;
    $\text{cost}(G | D) \approx \text{posterior probability } P(G | D)$

• **Size of grammar** $= \text{size}(G) \approx 1 / \text{prior } P(G)$
  - favor fewer/smaller constructions/roles; isomorphic mappings

• **Complexity of data given grammar** $\approx 1 / \text{likelihood } P(D | G)$
  - favor simpler analyses
    (fewer, more likely constructions)
  - based on derivation length + score of derivation
further questions?