The 30th Annual Conference of the Cognitive Science Society

Computational Modeling of Spoken Language Processing: A hands-on tutorial
Computational Modeling of Spoken Language Processing: A hands-on tutorial

Ted Strauss

New School University

Dan Mirman

Jim Magnuson

University of Connecticut Department of Psychology and Haskins Laboratories
Plan

• Module 1: Introduction, About TRACE
• Module 2: Tour of jTRACE
• Module 3: Classic simulations
• Module 4: Scripting
• Module 5: Linking hypotheses
• Module 6: Lab time, Q&A, one-on-one
An aside

• Why did we develop jTRACE?
  – To facilitate large-scale modeling
  – To promote active testing of TRACE predictions and wider use of modeling
  – To facilitate replication and sharing of simulations

• How did we develop jTRACE?
  – With a budget supplement to an NIDCD R01 and a couple Ted years

• Why are we doing this tutorial?
Module 5: Advanced topics

• How do you decide whether a model has succeeded or failed?
  – Connecting model to human behavior

• Pitfalls: simulations can fail at multiple levels
  – Theory -- most interesting/informative
  – Implementational details/parameters
  – Linking hypotheses -- not a model failure -- equivalent to flawed operational definitions in an experiment!

• Before assuming a failure has theoretical implications, other levels must be excluded
Linking hypotheses

- Informal: does model capture basic trends?
- Formal: linking hypothesis
  - Link **model** to **data** by constructing task constraints for the model analogous to those faced by human subjects
  - **Model**: Activations over time
  - **Data**: Reaction times/accuracy for **specific** decisions or behaviors (lexical decision, eye tracking, ERP)
Simple example 1: Threshold

- Recognition = word unit activation exceeds threshold
- RT ≈ number of processing cycles from word onset
- Activation ~ internal state; what about choice behavior?
Simple example 2: Response probability

- RT \approx \text{number of processing cycles from word onset}
- Additional competition analogous to human choice behavior in many domains
- Formalization of overt choice based on internal states
- When to use: many choice situations, but especially AFC

\[ L_i = \frac{S_i}{\sum S_j} \quad S_i = e^{ka_i} \]
Pitfalls in modeling

1. Material selection
2. Material manipulation
3. Linking hypotheses
4. Logic
Pitfall 1: Material selection

- Frauenfelder & Peeters (1998)
  - Feedback doesn’t help
    - Half their items were recognized more quickly with feedback off
    - Feedback only allows TRACE to account for top-down effects?

- 21 items: 7 phones long, UP at phone 4

- Magnuson, Strauss, & Harris (2005)
  - Tested 900 words
  - With/without feedback
  - Increasing levels of noise
  - 73% of words recognized more quickly w/feedback
Pitfall 2: Material manipulation

- Model materials must be held to same standard as behavioral materials
- Material manipulations should have analogous effects
- Marslen-Wilson & Warren, 1994 (subcategorical mismatch)

Hypothesis:
- If there is lateral inhibition between words,
- Then, if we provide misleading coarticulation consistent with a
  - **Word**: substantial lexical competition
  - **Nonword**: less competition
ne^t\text{t}

net spliced with net

ne^k\text{t}

neck spliced with net

ne^p\text{t}

nep* spliced with net
Pitfall 2: Material manipulation


Lexical decision

\(ne^{tt} 487\)
\(ne^{pt} 609\)
\(ne^{ckt} 610\)
Pitfall 2: Material manipulation

- Original splicing was so late that “neck” was recognized instead of “net”!
- We cross-spliced at the latest slice where “net” was still ultimately recognized

Lexical decision

net^t 487
ne^pt 609
ne^ckt 610
Pitfall 3: Linking hypotheses

- How can we reconcile TRACE activations and human LD RTs?
- Lexical decision need not be based on target
- What if we set threshold so that W2’s activation sometimes passes?

**Lexical decision**

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>E1</th>
<th>E2</th>
<th>E3</th>
</tr>
</thead>
<tbody>
<tr>
<td>net</td>
<td>487</td>
<td></td>
<td></td>
</tr>
<tr>
<td>npt</td>
<td>609</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nct</td>
<td>610</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From TRACE activations

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>E1</th>
<th>E2</th>
<th>E3</th>
</tr>
</thead>
<tbody>
<tr>
<td>net</td>
<td>48</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>nct</td>
<td>53</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Response probability

<table>
<thead>
<tr>
<th>Target</th>
<th>W1</th>
<th>W2</th>
<th>W1</th>
<th>W2</th>
</tr>
</thead>
<tbody>
<tr>
<td>W1</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W2</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W1</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W2</td>
<td>0.30</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Predicted LDT (cycles)

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>net</td>
<td>48</td>
</tr>
<tr>
<td>nct</td>
<td>53</td>
</tr>
</tbody>
</table>

Threshold

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Predicted LDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>40</td>
</tr>
<tr>
<td>0.15</td>
<td>50</td>
</tr>
<tr>
<td>0.20</td>
<td>60</td>
</tr>
<tr>
<td>0.25</td>
<td>70</td>
</tr>
<tr>
<td>0.30</td>
<td></td>
</tr>
</tbody>
</table>
The “visual world” paradigm

Eye tracking

Example trials

Averaged data

Fixation proportion

Fixation proportion

Time

Target

Competitor

Unrelated

200 - 400 - 700 ms

Target

Unrelated

Competitor

Unrelated

Click on the candle
Linking hypothesis

- Link fixation data to TRACE word activations
- Luce Choice rule: Activations and strengths based on entire lexicon; **choice** only includes the 4 onscreen items -- analogous to choice faced by subjects

\[ S_i = e^{ka_i}, \quad L_i = \frac{S_i}{\Sigma S_j} \]

\[ L_{target} = \frac{S_{target}}{(S_{target} + S_{comp} + S_{d1} + S_{d2})} \]

- Provides clear and testable predictions about number and nature of the items in the display
- See Dahan et al. (2001) for examples where this simple linking hypothesis accurately predicts changes in fixation proportions depending on display
Cohort present

Cohort absent

Dahan, Magnuson, Hogan & Tanenhaus, 2001

Fixation data

Target (w1, net) fixations
- W1 | w1w1 - ne(t)t
- W1 | n3w1 - ne(p)t
- W1 | w2w1 - ne(ck)t

Cohort (w2, neck) fixations
- W2 | w1w1 - ne(t)t
- W2 | n3w1 - ne(p)t
- W2 | w2w1 - ne(ck)t

Model predictions

Target (w1, net)
- w1w1 - ne(t)t
- n3w1 - ne(p)t
- w2w1 - ne(ck)t

Cohort (w2, neck)
- w1w1 - ne(t)t
- n3w1 - ne(p)t
- w2w1 - ne(ck)t

Average vowel offset

Average word offset

Reponse probability
Advantages of linking hypotheses

• Formalizes model of both internal state \textit{and} overt behavior for specific tasks
• Also links cognitive systems -- e.g., visual attention and lexical activation
• Facilitates the interpretation of behavioral data from complex experimental paradigms
• Without careful, explicit modeling of task constraints via linking hypotheses, simple model activations can be misleading
Pitfall 4: Logic

• Sometimes, the most reasonable predictions turn out to be **wrong when you test them by simulation**

• Case in point: word frequency

• Assumptions about modeling: resting activation $\approx$ bottom-up connection strength

• Assumptions about empirical results:
  – Absence of frequency effects in early responses
  – Therefore, frequency is a late, top-down bias
Proposed loci of frequency

- HF words recognized more quickly than LF words
- Early sampling (e.g., fast reactions) sometimes fails to detect frequency effects (Connine, Titone, & Wang, 1993)
- Conclusion: frequency is a late/2nd stage bias?

Constant bias (resting levels)  Late bias (???)  Bottom-up dependent (connection strengths)

Bed  Bell  Bed  Bell  Bed  Bell

Late, external

b  d  L  E  b  d  L  E  b  d  L  E
Time course implications

- Is a late frequency effect evidence for 2\textsuperscript{nd} stage?
- No; consistent with bottom-up dependence
Cohort frequency

Sample critical display

• LF target (bell)
• LF cohort (bench)
• HF cohort (bed)
Model predictions

\[ S_i = e^{ka_i} \]

\[ R' = R + s(\log[f + 1]) \]

\[ a'_pi = a_{pi} + a_{pi}[s(\log[f + 1])] \]

\[ S_i' = S_i(\log[f + s]) \]
Model-data comparisons

Data

Model (weight version)

Fixation proportion

Msecs since target onset

Time since target onset (msecs)

RMS

Model (weight version)

Th {\textsuperscript{2}} L D

r^2 .97 .96 .94 .80

RMS .09 .03 .03 .06
Gauging success and failure

• Given an apparent mismatch between model and data, you must avoid the pitfalls -- you must exclude:
  – Poor analogs to materials (representativeness, manipulations)
  – Insufficient linking hypothesis
  – Worst case: mismatch between expectations and data -- jTRACE was created to encourage more testing of logical expectations

• Then, if you still have a mismatch between model and data, you must determine the level of the failure
  – Parameter value?
  – Specific aspects of model mechanisms
  – Theoretical assumptions
Gauging success and failure

• Model flexibility, model fit
  – Fit measures: $r^2$, RMS error
  – Debates over whether fit is sufficiently constraining
  – Parameter space partitioning (Pitt et al., 2005)

• Comparing two models
  – Occam’s razor

• More important?
  – Breadth
Using jTRACE

• Now what?
  – Work through the examples in the gallery (see documentation in ‘help’)
  – Make a plan for doing your own simulations
  – Email us if you need help (or find bugs, or have feature requests*)
  – Save your simulations! This is a great way to save yourself work and facilitate replication

• Right now: ‘lab time’