Today's topics

We're going to switch topics next time...
Last lecture on the WSJ Penn treebank
• Short remarks on Homework 7
• Robustness and Sensitivity
Bikel-Collins Parsing of Section 23

- split –l 100 section 23 into parts {a-w} and manually load-balance over 4 cores:
  - Time in seconds from bash $SECONDS on a Dell XPS 13" laptop (2017) Core i7-7500U
  - 2300 seconds \(\approx\) 38-39 minutes

<table>
<thead>
<tr>
<th>Run1</th>
<th>Total time</th>
<th>Run2</th>
<th>Total time</th>
<th>Run3</th>
<th>Total time</th>
<th>Run4</th>
<th>Total time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h</td>
<td>867</td>
<td>o</td>
<td>863</td>
<td>j</td>
<td>617</td>
<td>n</td>
<td>551</td>
</tr>
<tr>
<td>f</td>
<td>481</td>
<td>m</td>
<td>509</td>
<td>l</td>
<td>527</td>
<td>k</td>
<td>537</td>
</tr>
<tr>
<td>d</td>
<td>431</td>
<td>i</td>
<td>427</td>
<td>b</td>
<td>447</td>
<td>c</td>
<td>475</td>
</tr>
<tr>
<td>s</td>
<td>334</td>
<td>v</td>
<td>220</td>
<td>q</td>
<td>354</td>
<td>g</td>
<td>384</td>
</tr>
<tr>
<td>e</td>
<td>200</td>
<td>a</td>
<td>217</td>
<td>u</td>
<td>232</td>
<td>t</td>
<td>226</td>
</tr>
<tr>
<td></td>
<td></td>
<td>w</td>
<td>76</td>
<td>r</td>
<td>116</td>
<td>p</td>
<td>129</td>
</tr>
</tbody>
</table>
Bikel-Collins Parsing of Section 23

4 terminal screens

• Actual wall times:
  • 1:18:20
  • 1:18:27
  • 1:18:48
  • 1:19:27

Note: these aren’t the single core times because the Dell has only 2 real full cores (4 including hyperthreading).

Note: latest 2018 Dell XPS 13 has 4 real cores (8 incl. hyperthreading) – i7-8550U
https://ark.intel.com/products/122589/Intel-Core-i7-8550U-Processor-8M-Cache-up-to-4_00-GHz
Bikel-Collins Parsing of Section 23

Optimistic view?

- gold tags supplied
- sentence length \( \leq 40 \) only
Bikel-Collins Parsing of Section 23

nevalb data:

<table>
<thead>
<tr>
<th>Number of sentence</th>
<th>Number of Error sentence</th>
<th>Number of Skip sentence</th>
<th>Number of Valid sentence</th>
<th>Bracketing Recall</th>
<th>Bracketing Precision</th>
<th>Bracketing FMeasure</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>3</td>
<td>3</td>
<td>2238</td>
<td>58.98</td>
<td>64.55</td>
<td>61.64</td>
</tr>
<tr>
<td>1205</td>
<td>2241</td>
<td>2241</td>
<td>2241</td>
<td>73.71</td>
<td>78.13</td>
<td>78.34</td>
</tr>
<tr>
<td>120520</td>
<td>65.30</td>
<td>65.30</td>
<td>67.48</td>
<td>76.33</td>
<td>79.42</td>
<td>77.42</td>
</tr>
<tr>
<td>1205021</td>
<td>2241</td>
<td>2241</td>
<td>2241</td>
<td>77.02</td>
<td>79.31</td>
<td>78.14</td>
</tr>
<tr>
<td>12050210101</td>
<td>2241</td>
<td>2241</td>
<td>2241</td>
<td>77.42</td>
<td>79.42</td>
<td>78.14</td>
</tr>
<tr>
<td>1205021010109</td>
<td>2241</td>
<td>2241</td>
<td>2241</td>
<td>77.02</td>
<td>79.31</td>
<td>78.14</td>
</tr>
<tr>
<td>120502101010917</td>
<td>2241</td>
<td>2241</td>
<td>2241</td>
<td>77.42</td>
<td>79.42</td>
<td>78.14</td>
</tr>
<tr>
<td>1205021010109176</td>
<td>2241</td>
<td>2241</td>
<td>2241</td>
<td>77.02</td>
<td>79.31</td>
<td>78.14</td>
</tr>
<tr>
<td>120502101010917618</td>
<td>2241</td>
<td>2241</td>
<td>2241</td>
<td>77.42</td>
<td>79.42</td>
<td>78.14</td>
</tr>
<tr>
<td>12050211118917161080</td>
<td>2241</td>
<td>2241</td>
<td>2241</td>
<td>77.02</td>
<td>79.31</td>
<td>78.14</td>
</tr>
<tr>
<td>12050211118917161080686</td>
<td>2241</td>
<td>2241</td>
<td>2241</td>
<td>77.02</td>
<td>79.31</td>
<td>78.14</td>
</tr>
<tr>
<td>1205021111891716108068083</td>
<td>2241</td>
<td>2241</td>
<td>2241</td>
<td>77.02</td>
<td>79.31</td>
<td>78.14</td>
</tr>
<tr>
<td>1205021111891716108068083087</td>
<td>2241</td>
<td>2241</td>
<td>2241</td>
<td>77.02</td>
<td>79.31</td>
<td>78.14</td>
</tr>
<tr>
<td>120502111189171610806808308714</td>
<td>2241</td>
<td>2241</td>
<td>2241</td>
<td>77.02</td>
<td>79.31</td>
<td>78.14</td>
</tr>
<tr>
<td>1205021111891716108068083087140215084</td>
<td>2241</td>
<td>2241</td>
<td>2241</td>
<td>77.02</td>
<td>79.31</td>
<td>78.14</td>
</tr>
<tr>
<td>120502111189171610806808308714021508404</td>
<td>2241</td>
<td>2241</td>
<td>2241</td>
<td>77.02</td>
<td>79.31</td>
<td>78.14</td>
</tr>
<tr>
<td>12050211118917161080680830871402150840405</td>
<td>2241</td>
<td>2241</td>
<td>2241</td>
<td>77.02</td>
<td>79.31</td>
<td>78.14</td>
</tr>
<tr>
<td>1205021111891716108068083087140215084040506</td>
<td>2241</td>
<td>2241</td>
<td>2241</td>
<td>77.02</td>
<td>79.31</td>
<td>78.14</td>
</tr>
<tr>
<td>120502111189171610806808308714021508404050607</td>
<td>2241</td>
<td>2241</td>
<td>2241</td>
<td>77.02</td>
<td>79.31</td>
<td>78.14</td>
</tr>
<tr>
<td>12050211118917161080680830871402150840405060708</td>
<td>2241</td>
<td>2241</td>
<td>2241</td>
<td>77.02</td>
<td>79.31</td>
<td>78.14</td>
</tr>
<tr>
<td>1205021111891716108068083087140215084040506070809</td>
<td>2241</td>
<td>2241</td>
<td>2241</td>
<td>77.02</td>
<td>79.31</td>
<td>78.14</td>
</tr>
<tr>
<td>120502111189171610806808308714021508404050607080910</td>
<td>2241</td>
<td>2241</td>
<td>2241</td>
<td>77.02</td>
<td>79.31</td>
<td>78.14</td>
</tr>
<tr>
<td>12050211118917161080680830871402150840405060708091011</td>
<td>2241</td>
<td>2241</td>
<td>2241</td>
<td>77.02</td>
<td>79.31</td>
<td>78.14</td>
</tr>
</tbody>
</table>
Other sections ...
Robustness and Sensitivity

it’s often assumed that statistical models are less brittle than symbolic models
1. can get parses for ungrammatical data
2. are they sensitive to noise or small perturbations?
Robustness and Sensitivity

Examples

1. Herman mixed the water with the milk
2. Herman mixed the milk with the water
3. Herman drank the water with the milk
4. Herman drank the milk with the water

\[ \log p(\text{water}) = 117, \log p(\text{milk}) = 21 \]
Robustness and Sensitivity

Examples
1. Herman mixed the water with the milk
2. Herman mixed the milk with the water
3. Herman drank the water with the milk
4. Herman drank the milk with the water

\[
\begin{align*}
\logprob &= -50.4 \\
\logprob &= -47.2
\end{align*}
\]

*different PP attachment choices*
Robustness and Sensitivity

First thoughts...

• does milk forces low attachment?
  (high attachment for other nouns like water, toys, etc.)
  *Is there something special about the lexical item milk?*

• 24 sentences in the WSJ Penn Treebank with milk in it, 21 as a noun
Robustness and Sensitivity

First thoughts... *Is there something special about the lexical item milk?*

- 24 sentences in the WSJ Penn Treebank with *milk* in it, 21 as a noun
- but just one sentence (#5212) with PP attachment for *milk*
Robustness and Sensitivity

• Simple perturbation experiment
  • alter that one sentence and retrain
Robustness and Sensitivity

• Simple perturbation experiment
  • alter that one sentence and retrain

delete the PP with 4% butterfat altogether
Robustness and Sensitivity

• **Simple perturbation experiment**
  • *alter that one sentence and retrain*

or bump it up to the VP level
Robustness and Sensitivity

**Result:**
- High attachment for previous PP adjunct to *milk*

---

**Why such extreme sensitivity to perturbation?**
Logprobs are conditioned on many things; hence, lots of probabilities to estimate:
- **Smoothing**
- *Need every piece of data, even low frequency ones*
Details...

• Two sets of files:

```
dhcp-10-134-211-43:bin sandiway$ diff wsj-02-21-milk.mrg wsj-02-21.mrg
31530,31532c31530,31537
<    (NP (DT a) (NN milk))
<    (PP-LOC (IN in)
<    (NP (DT the) (NNP South)))
---
>    (NP
>    (NP (DT a) (NN milk))
>    (PP (IN with)
>    (NP
>    (ADJP (CD 4) (NN %))
>    (NN butterfat)))
>    (PP-LOC (IN in)
>    (NP (DT the) (NNP South)))
```
Bikel/Collins Parser wrapper

Herman drank the milk with the water.
Herman drank the water with the milk.

MXPOST:

Bikel/Collins: Parse
Train
using Derived Counts: wsj-02-21-milk.obj.gz
using Treebank: wsj-02-21.mrg
using settings: collins.properties

Treebankviewer: Display
Robustness and Sensitivity

• (Bikel 2004):
  • “it may come as a surprise that the [parser] needs to access more than 219 million probabilities during the course of parsing the 1,917 sentences of Section 00 [of the PTB]."
Robustness and Sensitivity

• Trainer has a memory like a phone book:
Robustness and Sensitivity

Frequency 1 observed data for:
\[(\text{NP (NP (DT a)(NN milk)) (PP (IN with)) (NP (ADJP (CD 4)(NN %)) (NN butterfat)))}\]

- \((\text{mod ((with IN) (milk NN) PP (+START+) (+START+) (+START+)) NP-A NPB () false right) 1.0})\)
  - modHeadWord (with IN)
  - headWord (milk NN)
  - modifier PP
  - previousMods (+START+)
  - previousWords ((+START+ +START+))
  - parent NP-A
  - head NPB
  - subcat ()
  - verbIntervening false
  - side right

- \((\text{mod ((+STOP+ +STOP+) (milk NN) +STOP+ (PP) ((with IN)) NP-A NPB () false right) 1.0})\)
  - modHeadWord (+STOP+ +STOP+)
  - headWord (milk NN)
  - modifier +STOP+
  - previousMods (PP)
  - previousWords ((with IN))
  - parent NP-A
  - head NPB
  - subcat ()
  - verbIntervening false
  - side right
Robustness and Sensitivity

76.8% singular events
94.2% 5 or fewer occurrences
Robustness and Sensitivity

• **Full story more complicated than described here...**

• **by picking different combinations of verbs and nouns, you can get a range of behaviors**

<table>
<thead>
<tr>
<th>Verb</th>
<th>Noun</th>
<th>Attachment</th>
</tr>
</thead>
<tbody>
<tr>
<td>drank</td>
<td>water</td>
<td>high high</td>
</tr>
<tr>
<td>mixed</td>
<td>water</td>
<td>low high</td>
</tr>
<tr>
<td>mixed</td>
<td>computer</td>
<td>low low</td>
</tr>
</tbody>
</table>

$f(\text{drank})=0$ might as well have picked *flubbed*