LING 581: Advanced Computational Linguistics

Lecture Notes

February 19th
Bikel Collins and EVALB

• How did you guys do with Bikel-Collins on section 23?

<table>
<thead>
<tr>
<th>Model</th>
<th>Performance on Section 00</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LR</td>
</tr>
<tr>
<td>Collins’ Model 2</td>
<td>89.75</td>
</tr>
<tr>
<td>Baseline (Model 2 emulation)</td>
<td>89.89</td>
</tr>
<tr>
<td>Clean-room Model 2</td>
<td>88.85</td>
</tr>
</tbody>
</table>

Figure 15
Overall parsing results using only details found in (Collins, 1997; Collins, 1999). The first two lines show the results of Collins’ parser and those of our parser in its “complete” emulation mode (i.e., including unpublished details). All reported scores are for sentences of length ≤ 40 words. LR/LP are the primary scoring metrics, labeled precision and labeled recall, respectively. CBs is the number of crossing brackets. 0 CBs and ≤ 2 CBs are the percentage of sentences with 0 and ≤ 2 crossing brackets, respectively. F (the “F-measure”) is the evenly-weighted harmonic mean of precision and recall, or $\frac{L_P \cdot L_R}{\frac{1}{2}(L_P + L_R)}$. 
Steps

- Running evalb
- Need:
  - parameter file: COLLINS.prm
  - Parser output: test.txt.parsed
  - Gold file: test.gold
Steps

• (Assume my data files are on my Desktop and EVALB is installed in my research subdirectory)
  – ../research/EVALB/evalb -p ../research/EVALB/COLLINS.prm test.gold test.txt.parsed

Recall 83.8, Precision 89.2
Task from last time

- Run section 23 of the WSJ Treebank on the Bikel Collins parser
  - Extract the sentences from section 23 (perl etc.)
- Then run EVALB on the section to see how the Bikel Collins parser scores.
  - C code: (OSX) comment out the malloc line
- Report back next time.
System Flowchart

• Diagram:

- WSJ treebank 00–24
  - tregex View Search
    - How?
      - tregex View
      - tregex
        - tregex tsurgeon (-s flag)
          - create using cat
          - Sec 23 trees .mrg
          - Sec 23 gold trees .mrg (one tree per line)
      - EVALB
        - recall precision F-measure ≈86%
          - COLLINS .prm
    - tregex
      - Events .obj.gz
        - Parse trees .txt.parsed (one tree per line)
      - Bikel Collins parser Parse
        - create using cat
          - Treebank trees .mrg
          - Treebank sentences .txt
          - Bikel Collins parser Train
            - create using cat
              - Parse Treebank sentences .txt
Getting EVALB running

• Download from NYU
  – written in (plain) C
• Makefile:
  1. all: evalb
  2. evalb: evalb.c
  3. gcc -Wall -g -o evalb evalb.c
• OSX:
  – make
  – gcc -Wall -g -o evalb evalb.c
  – evalb.c:25:20: fatal error: malloc.h: No such file or directory

```c
20
21#include <stdio.h>
22#include <stdlib.h> //### added for exit, atoi decls
23#include <cctype>
24#include <string.h>
25#include <malloc.h>
```
Getting EVALB running

• Google the error and osx
• Use Spotlight
  – (3 copies)
  – /usr/include/malloc/malloc.h
  – /usr/include/sys/malloc.h

It should compile now…
(but with many warnings about type mismatches)

```c
20 #include <stdio.h>
21 #include <stdlib.h> //### added for exit, atoi decls
22 #include <ctype.h>
23 #include <string.h>
24 #include <malloc.h>
25 #include <malloc/malloc.h>
```

or simply remove or comment out line 25
Extracting the sentences

We need to extract the sentences to run the Bikel Collins parser

• Section 23 contains 2416 sentences

• There are many possible ways to do this: let's look at a couple ...
Extracting the sentences

- Use the sec 23 gold trees: Perl regex:

```
(S (INTJ (RB No)) (, ,) (NP-SBJ (PRP it)) (VP (VBD was) (RB n't) (NP -PRD (NNP Black) (NNP Monday)))) (. .))
```

```
(S (CC But) (SBAR-ADV (IN while)) (S (NP-SBJ (DT the) (NNP New) (NNP York) (NNP Stock) (NNP Exchange)) (VP (VBD did) (RB n't) (VP (VB fall) (ADVP-CLR (RB apart)) (NP-TMP (NNP Friday)) (SBAR-TMP (IN as) (S (NP-SBJ (DT the) (NNP Dow) (NNP Jones) (NNP Industrial) (NNP Average)) (VP (VBD plunged) (NP-EXT (NP (CD 190.58) (NNS points)) (PRN (: --) (NP (NP (JJ most)) (PP (IN of) (NP (PRP it))) (PP-TMP (IN in) (NP (DT the) (JJ final) (NN hour))) (: --)))))))) (NP-SBJ-2 (PRP it)) (ADVP (RB barely)) (VP (VBD managed) (S (NP-SBJ (-NONE- *-2)) (VP (TO to) (VP (VB stay) (NP-LOC-PRD (NP (DT this) (NN side)) (PP (IN of) (NP (NN chores)))))))))) (. .))
```

```
(S-1 (NP-SBJ-2 (NP (DT Some) (`` ```)) (NN circuit) (NNS breakers) ('' '')) (VP (VBN installed) (NP (-NONE- *))) (PP-TMP (IN after) (NP (DT the) (NNP October) (CD 1987) (NN crash)))))) (VP (VBD failed) (NP (PRP$ their) (JJ first) (NN test)) (PRN (, ,) (S (NP-SBJ (NNS traders)) (VP (VBP say) (SBAR (-NONE- 0) (S (-NONE- *T*-1))))) (, ,)) (S-ADV (NP-SBJ -3 (-NONE- *-2)) (ADJP-PRD (JJ unable) (S (NP-SBJ (-NONE- *-3)) (VP (T 0 to) (VP (VB cool) (NP (NP (DT the) (NN selling) (NN panic)) (PP-LOC (IN in) (NP (DT both) (NNS stocks) (CC and) (NNS futures)))))))))))) (. .))
```

```
(S (NP-SBJ (NP (NP (DT The) (CD 49) (NN stock) (NN specialist) (NNS
```
Extracting the sentences

• Use the POS tagged data in TREEBANK_3:
Extracting the sentences

• Use the POS tagged data in TREEBANK_3:

```
No/RB ,/,
[ it/PRP ]

[ was/VBD n't/RB Black/NNP Monday/NNP ]
./.

But/CC while/IN
[ the/DT New/NNP York/NNP Stock/NNP Exchange/NNP ]
did/VBD n't/RB
[ fall/VB ]
apart/RB
[ Friday/NNP ]
as/IN
[ the/DT Dow/NNP Jones/NNP Industrial/NNP Average/NNP ]
plunged/VBD
[ 190.58/CD points/NNS ]
```
Extracting the sentences

- Use tregex

```plaintext
No, it was n't Black Monday.
But while the New York Stock Exchange didn 't fl
Some "circuit breakers" installed after the O
The 49 stock specialist firms on the Big Board fl
Big investment banks refused to step up to t
Heavy selling of Standard & Poor 's 500-stock in
Seven Big Board stocks -- UAL, AMR, BankAme
The finger-pointing has already begun.
Once again, the specialists were n't
Countered by William Maguire, chairman
When the dollar is in a free-fall "T"-1, even c
Speculators are calling for a degree of liquidity
Many money managers and some traders had a
Then in a lightning plunge, the Dow Jones ind
Final-hour trading accelerated to 108.1 million
At the end of the day, 251.2 million shares we
```
Extracting the sentences

• Use tregex Save Sentences: Perl regex

```
/*T*/-[0-9]+ /*-[0-9]+ /*U*/
\b0\b \b\b
```
Summary

– **WSJ corpus**: sections 00 through 24

– **Training**: normally 02-21 (20 sections)
  
  • concatenate mrg files as input to Bikel Collins trainer
  
  • training is *fast*

– **Evaluation**: on section 23
  
  • use tsurgeon to get data ready for EVALB
  
  • Use either the .pos files or tregex to grab the sentences for parsing
  
  • 2400+ sentences in section 23, parsing is *slow*...
  
  • Can split the sentences up and run on multiple processors
New Homework

• **Question:**
  – How much training data do you need?

• **Homework:**
  – How does Bikel Collins vary in precision and recall on test data if you randomly pick 1..24 out of 25 sections to do the training with?

• **Test section:**
  – I want you to pick a test section that’s not section 23

• **Use EVALB**
  – plot precision and recall graphs
Summary

• **Question:**
  – How much training data do you need?

• **Homework:**
  – How does Bikel Collins vary in precision and recall on test data if you randomly pick 1..24 out of 25 sections to do the training with?

• Test section:
  – I want you to pick a test section that’s not section 23

• Use EVALB
  – plot precision and recall graphs
Perl Code

• Generate training data

```perl
use List::Util 'shuffle';

my $prefix = '/Users/sandiway/research/';
my $dbprefix = "$prefix/dbparser";

my @sections = ("02", "03", "04", "05", "06", "07", "08", "09",
  "20", "21");

# from http://perldoc.perl.org/perlfaq4.html
my @shuffled = shuffle(@sections);

sub trainparser {
    # assume touch wsj-.mrg (i.e. empty file)
    my $filename = "wsj-";

    while (@shuffled) {
        my $section = shift @shuffled;
        system "cat $filename.mrg wsj_$section.mrg > $filename$section.mrg";
        print "Training on $filename$section.mrg\n";
        system "$dbprefix/bin/train 1000 $dbprefix/settings/$collins.properties $filename$section.mrg";
        $filename .= $section;
    }
}

trainparser;
```
## Training Data

<table>
<thead>
<tr>
<th>File Path</th>
<th>Date/Time</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>wsj-120520211011091716180806030714.mrg</td>
<td>Today 5:58 PM</td>
<td>20.1 MB</td>
</tr>
<tr>
<td>wsj-1205202110110917161808060307.observed.gz</td>
<td>Today 5:57 PM</td>
<td>7.4 MB</td>
</tr>
<tr>
<td>wsj-12052021101109171618080603.mrg</td>
<td>Today 5:57 PM</td>
<td>18.6 MB</td>
</tr>
<tr>
<td>wsj-12052021101109171618080603.obj.gz</td>
<td>Today 5:57 PM</td>
<td>14.9 MB</td>
</tr>
<tr>
<td>wsj-12052021101109171618080603.observed.gz</td>
<td>Today 5:57 PM</td>
<td>6.8 MB</td>
</tr>
<tr>
<td>wsj-12052021101109171618080603.mrg</td>
<td>Today 5:56 PM</td>
<td>17.1 MB</td>
</tr>
<tr>
<td>wsj-120520211011091716180806.obj.gz</td>
<td>Today 5:56 PM</td>
<td>14.3 MB</td>
</tr>
<tr>
<td>wsj-120520211011091716180806.observed.gz</td>
<td>Today 5:56 PM</td>
<td>6.5 MB</td>
</tr>
<tr>
<td>wsj-120520211011091716180806.obj.gz</td>
<td>Today 5:56 PM</td>
<td>13.3 MB</td>
</tr>
<tr>
<td>wsj-120520211011091716180806.mrg</td>
<td>Today 5:56 PM</td>
<td>16.1 MB</td>
</tr>
<tr>
<td>wsj-120520211011091716180806.observed.gz</td>
<td>Today 5:55 PM</td>
<td>6 MB</td>
</tr>
<tr>
<td>wsj-1205202110110917161808.mrg</td>
<td>Today 5:55 PM</td>
<td>14.8 MB</td>
</tr>
<tr>
<td>wsj-12052021101109171618.obj.gz</td>
<td>Today 5:55 PM</td>
<td>13.1 MB</td>
</tr>
<tr>
<td>wsj-12052021101109171618.observed.gz</td>
<td>Today 5:55 PM</td>
<td>5.9 MB</td>
</tr>
<tr>
<td>wsj-12052021101109171618.mrg</td>
<td>Today 5:54 PM</td>
<td>14.4 MB</td>
</tr>
<tr>
<td>wsj-120520211011091716.obj.gz</td>
<td>Today 5:54 PM</td>
<td>12 MB</td>
</tr>
<tr>
<td>wsj-120520211011091716.observed.gz</td>
<td>Today 5:54 PM</td>
<td>5.4 MB</td>
</tr>
<tr>
<td>wsj-120520211011091716.mrg</td>
<td>Today 5:54 PM</td>
<td>12.9 MB</td>
</tr>
<tr>
<td>wsj-1205202110110917.obj.gz</td>
<td>Today 5:54 PM</td>
<td>10.5 MB</td>
</tr>
<tr>
<td>wsj-1205202110110917.observed.gz</td>
<td>Today 5:53 PM</td>
<td>4.6 MB</td>
</tr>
<tr>
<td>wsj-12052021101109.obj.gz</td>
<td>Today 5:53 PM</td>
<td>9.7 MB</td>
</tr>
<tr>
<td>wsj-12052021101109.mrg</td>
<td>Today 5:53 PM</td>
<td>11 MB</td>
</tr>
<tr>
<td>wsj-12052021101109.observed.gz</td>
<td>Today 5:53 PM</td>
<td>4.2 MB</td>
</tr>
</tbody>
</table>
Training Data

Training data size

Number of sentences vs. Number of sections

0 5000 10000 15000 20000 25000 30000 35000 40000
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Training Data

• What does the graph for the F-score (Labeled Recall/Labeled Precision) look like?

F-score

# of sentences used for training

(report your results next time...
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

\[
f(\text{water}) = 117, \quad f(\text{milk}) = 21
\]
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

Different PP attachment choices

logprob = -50.4

logprob = -47.2
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

*Could just one sentence out of 39,832 training examples affect the attachment options?*
Robustness and Sensitivity

- **Simple perturbation experiment**
  - *alter that one sentence and retrain*
Robustness and Sensitivity

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

```
NP-SBJ-1  ADVP
NNP Borden  RB even
  VBD tested  NP
    DT a  NN milk in
    NP PP-LOC
      CC but
      VBD decided
      SBAR
        -NONE-  S
          NP-SBJ
            DT the
            NNP South
          VP
            ,
            VP
              ,
              CC
                but
              VBD decided
              SBAR
                -NONE-  S
                  NP-SBJ
                    DT the
                    NNP South
                  VP
                    ,
                    VP
                      ,
                      CC
                        but
                      VBD decided
                      SBAR
                        -NONE-  S
                          NP-SBJ
                            DT the
                            NNP South
                          VP
                            ,
                            VP
                              ,
                              CC
                                but
                              VBD decided
                              SBAR
                                -NONE-  S
                                  NP-SBJ
                                    DT the
                                    NNP South
                                  VP
                                    ,
                                    VP
                                      ,
                                      CC
                                        but
                                      VBD decided
                                      SBAR
                                        -NONE-  S
                                          NP-SBJ
                                            DT the
                                            NNP South
                                          VP
                                            ,
                                            VP
                                              ,
                                              CC
                                                but
                                              VBD decided
                                              SBAR
                                                -NONE-  S
                                                  NP-SBJ
                                                    DT the
                                                    NNP South
                                                  VP
                                                    ,
                                                    VP
                                                      ,
                                                      CC
                                                        but
                                                      VBD decided
                                                      SBAR
                                                        -NONE-  S
                                                          NP-SBJ
                                                            DT the
                                                            NNP South
                                                          VP
                                                            ,
                                                            VP
                                                              ,
                                                              CC
                                                                but
                                                              VBD decided
                                                              SBAR
                                                                -NONE-  S
                                                                  NP-SBJ
                                                                    DT the
                                                                    NNP South
                                                                  VP
                                                                    ,
                                                                    VP
                                                                      ,
                                                                      CC
                                                                        but
                                                                      VBD decided
                                                                      SBAR
                                                                        -NONE-  S
                                                                          NP-SBJ
                                                                            DT the
                                                                            NNP South
                                                                          VP
                                                                            ,
                                                                            VP
                                                                              ,
                                                                              CC
                                                                                but
                                                                              VBD decided
                                                                              SBAR
                                                                                -NONE-  S
                                                                                  NP-SBJ
                                                                                    DT the
                                                                                    NNP South
                                                                                  VP
                                                                                    ,
                                                                                    VP
                                                                                      ,
                                                                                      CC
                                                                                                    but
                                                                                      VBD decided
                                                                                      SBAR
                                                                                                    -NONE-  S
                                                                                                      NP-SBJ
                                                                                                        DT the
                                                                                                        NNP South
                                                                                                      VP
                                                                                                        ,
                                                                                                        VP
                                                                                                          ,
                                                                                                          CC
                                                                                                            but
                                                                                                          VBD decided
                                                                                                          SBAR
                                                                                                            -NONE-  S
                                                                                                              NP-SBJ
                                                                                                                DT the
                                                                                                                NNP South
                                                                                                              VP
                                                                                                                ,
                                                                                                                VP
                                                                                                                  ,
                                                                                                                  CC
                                                                                                                    but
                                                                                                                  VBD decided
                                                                                                                  SBAR
                                                                                                                    -NONE-  S
                                                                                                                      NP-SBJ
                                                                                                                        DT the
                                                                                                                        NNP South
                                                                                                                      VP
                                                                                                                        ,
                                                                                                                        VP
                                                                                                                          ,
                                                                                                                          CC
                                                                                                                                            but
                                                                                                                          VBD decided
                                                                                                                          SBAR
                                                                            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*

Could just one sentence out of 39,832 training examples affect the attachment options? 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
Two sets of files:

```bash
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:

(((Herman (NNP)) (drank (VB)) (the (DT)) (milk (NN)) (with (IN)) (the (DT)) (water (NN)) (. .))))
(((Herman (NNP)) (drank (VB)) (the (DT)) (water (NN)) (with (IN)) (the (DT)) (milk (NN)) (. .))))

Bikel/Collins:

Parse
Train

using Derived Counts: wsj-02-21-milk.obj.gz
using Treebank: wsj-02-21.mrg
using settings: collins.properties

Treebankviewer:

Total time reading ModelCollection object: 00:11.220.
-41.251 (S (NP-A (NPB (NNP Herman))) (VP (VB +unknown+) (NP-A (NPB (DT the) (NN milk))) (PP (IN with) (NP-A (NPB (DT the) (NN water))))))
-42.445 (S (NP-A (NPB (NNP Herman))) (VP (VB +unknown+) (NP-A (NPB (DT the) (NN water))) (PP (IN with) (NP-A (NPB (DT the) (NN milk))))))
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:

\((\text{NP} (\text{NP} (\text{DT} a)(\text{NN} \text{ milk})) (\text{PP} (\text{IN} \text{ with}) (\text{NP} (\text{ADJP} (\text{CD} 4)(\text{NN} \%))(\text{NN} \text{ butterfat}))))\)

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

- \((\text{mod} ((+\text{STOP}+ +\text{STOP}+) (\text{milk} \text{ NN}) +\text{STOP}+ (\text{PP}) ((\text{with} \text{ IN})) \text{ NP-A} \text{ NPB} () \text{ false} \text{ right}) 1.0)\)
  - 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

B-Collins: Observed Data

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</td>
</tr>
<tr>
<td>mixed</td>
<td>water</td>
<td>low</td>
</tr>
<tr>
<td>mixed</td>
<td>computer</td>
<td>low</td>
</tr>
</tbody>
</table>

\[ f(\text{drank}) = 0 \]

might as well have picked \textit{flubbed}