CS 378 Lecture 14

Today
- Shift-reduce parsing
- Midterm review

I ate and drank

Recap Dependencies
- Each word has one parent

Root

I ate the cake with a fork

- Verbs are heads of sentences
- Verbs have nouns, prepositions as children most frequently

Advantages:
- Some attachments make more sense than in constituency
- Easier to adapt to a wide range of languages
Announcements

- A3 due
- No lecture Thursday
- Midterm: Weds 9am - Fri 5pm
  - Ask Qs via email or in private Piazza posts
  - Open book, NOT collaborative!
- Extra OHs today
  1:30-2:30
  No OHs Weds - Fri:

Shift-reduce parsing

- Move through a sentence word-by-word
  + make decisions as we go
Stack: partial parse trees
Buffer: rest of the sentence
Initial state: $\downarrow$ I ate some spaghetti bolognese

Stack: [ROOT]
Buffer: [I ate some spaghetti bolognese]

Three ops: 
- $\text{Shift}$ first word from buffer to stack
- $\text{Reduce}$: left-arc, right-arc
- $\text{Add}$ an arc to combine items on the stack

$S$ [ROOT] $B$ [I ate some spaghetti bolognese]

$S$ [ROOT I] $B$ [ate some spaghetti bolognese]

$S$ [ROOT I ate] $B$ [some spaghetti bolognese]
Garden path: The horse raced past the barn that fell.

2. Left-arc: takes top two els of stack, makes 2nd-to-last a child of the last one, adds to stack

```
[ shift ]  [ shift ]
```

```
S [ ROOT ate ]  B [ some spa bo ]
```

```
S [ ROOT ate some sp ]  B [ bo ]
```

```
S [ ROOT ate sp ]  B [ bo ]
```

DO NOT R-A!
3. Right-arc: takes top two elts from stack makes last a child of 2nd-to-last, adds

Right-arc

S [ ROOT ate sp ] B [ ]

\[
\begin{array}{c}
\text{some} \\
\text{bo}
\end{array}
\]

Right-arc

S [ ROOT ate ]

len 2

I

sp

\[
\begin{array}{c}
\text{some} \\
\text{bo}
\end{array}
\]

R-A

S [ ]

\[
\begin{array}{c}
\text{ROOT} \\
\text{ate} \\
\text{sp} \\
\text{some} \\
\text{bo}
\end{array}
\]
Building Shift-reduce parsers

Our parser is a classifier

Maps from state (Stack, buffer) to one of three actions

Features \( f(S, B) \) \[
\text{"Different weights"} \quad W_{sh} \quad W_{ra} \quad W_{la} \] 

\text{dot product}

Features are complex!

\[ S \left[ \text{ate} \right] \quad B \left[ \text{the ...} \right] \]

Indicator \( \text{last word on stack is a verb \& first Buf word is } \text{the} \)
Midterm review

Start symbol: [VP]

VP → V VP\* 0.5
VP → V NNS 0.5
VP* → NNS PP 1.0
PP → P NNS 1.0

Rest follows the sheet, ignore Q1

-1 from rule

$P(\text{rule|VP})$ must normalize

$sells$  $books$
Scores: \( \log P(T|x) \)

- \( v: -1 \) NNS: -1
- \( v: -1 \)
- \( v: -2 \)
- \( v: -1 \) NNS: -1
- \( p: 0 \)
- \( v: -1 \) NNS: -1

Sells books + books

Splits here

Delexicalize VP

VP

V

sells

VP

NNS
books

PP

sells books p NNS

to books

P

NNS
to books
Logistic regression: loss when $y^{(i)} = +1$

$$- \log P(y = +1 | x)$$

$$= - \log \frac{e^{\mathbf{w}^T f(x)}}{1 + e^{\mathbf{w}^T f(x)}}$$

$$= - \mathbf{w}^T f(x) + \log \left( 1 + e^{\mathbf{w}^T f(x)} \right)$$

$$= - z + \log (1 + e^z)$$

Perception: penalize by how wrong it is correct $= 0$ loss