Midterm Review

Natural Language Processing: Jordan Boyd-Graber
University of Colorado Boulder
22. OCTOBER 2014

Roadmap

- Answer your questions
- Go through examples of free response questions

Your Questions

Your Questions

Your Questions

Your Questions

Your Questions

Your Questions

Your Questions

PCFG + LM

Suppose we have the the following language model over the alphabet $\{a, b\}$.

Bigram	Probability
$p(a \mid \texttt{<s>})$	0.45
$p(b \mid \texttt{<s>})$	0.45
$p(\texttt{</s>} \mid \texttt{<s>})$	0.1
$p(a \mid a)$	0.6
$p(b \mid a)$	0.2
$p(\texttt{</s>} \mid a)$	0.2
$p(a \mid b)$	0.2
$p(b \mid b)$	0.4
$p(\texttt{</s>} \mid b)$	0.4

1. Write a PCFG with non-terminals and weights such that it is equivalent to this language model. You should not need more than three non-terminals.
2. Compute the probability of the string `<s> a a b </s>` using the original language model and the corresponding PCFG derivation to show that they're equivalent.

PCFG + LM

PCFG + LM

PCFG + LM

PCFG + LM

FST

For any binary string x, let $w(x)$ denote the the number of 1's in x.

- For any binary string x and any integer i, $0 \leq i < w(x)$, let $f(x, i)$ denote the number of 0's between the i^{th} 1 and the $(i+1)^{\text{st}}$ 1 in the binary string $1x$, where we index the $w(x) + 1$ 1's in $1x$ from left to right starting at zero. Example: If $x = 11000100$, then $w(x) = 3, f(x, 0) = 0, f(x, 1) = 0, f(x, 2) = 3$, and $f(x, i)$ is undefined for $i \geq 3$.
- For any binary string x, let $g(x)$ denote the binary string of length $w(x)$ with i^{th} bit (indexing the bits from left to right starting at zero) equal to the parity of $f(x, i)$ (that is, 0 if even, 1 if odd). Example: If $x = 11000100$, then $g(x) = 001$.

Design a finite state transducer that maps any given input binary string x to the output binary string $g(x)$.

FST

FST

FST

FST

MaxEnt

Take V to be the set of possible words (e.g. "the", "cat", "dog", ...). Take V' to be the set of all words in V **plus** their reverses (e.g. "the", "eht", "cat", "tac", "dog", "god"). You can assume that there are no palindromes in v (e.g. "eye"). Nathan L. Pedant generates $(x, y) : x \in V, y \in V'$ pairs as follows:

- With probability $\frac{1}{2}$ he chooses y to be identical to x
- With probability $\frac{1}{3}$ he chooses y to be the reverse of x
- With probability $\frac{1}{6}$ he chooses y to be some string that is neither x nor the reverse of x

Create a log-linear distribution (i.e. supply features f and weights θ) of the form:

$$p(y|x, \vec{\theta}) = \frac{\exp \sum_i \theta_i f_i(x, y)}{\sum_{y'} \exp \sum_i \theta_i f_i(x, y')} \tag{1}$$

that models Nathan's process perfectly.

MaxEnt

MaxEnt

MaxEnt

MaxEnt