Questions for Flipped Classroom Session of COMS 4705 Week 13, Fall 2014. (Michael Collins)

Question 1 Consider an application of global linear models to dependency parsing. In this scenario each input x is a sentence. GEN(x) returns the set of all dependency parses for x. The feature vector $f(x, y)$ for any sentence x paired with a dependency parse tree y is defined as

$$f(x, y) = \sum_{(h,m)\in y} g(x, h, m)$$

where g is a function that maps a dependency (h, m) together with the sentence x to a local feature vector. Here h is the index of the head-word of the dependency, and m is the index of the modifier word.

We'd like $f(x, y)$ to be a 2-dimensional feature vector, with the following values for its three components:

$f_1(x, y)$ = Num of times a dependency with head *car*, and modifier *the* is seen in (x, y)

$f_2(x, y)$ = Num of times a dependency with head part-of-speech *NN*, modifier part-of-speech *DT*, no verb between the DT and NN is seen in (x, y)

Give a definition of the function g that leads to this definition of $f(x, y)$. You can assume that $POS(i)$ for $i \in \{1 \ldots n\}$ returns the part-of-speech of word i in the sentence.

Question 2 In this question we develop a dynamic programming approach to finding the highest scoring dependency parse for a sentence. Each dependency parse y is a set of (h, m) pairs, where h is the index of the head word in the dependency, m is the index of the modifier word. The global feature vector for a dependency parse is

$$f(x, y) = \sum_{(h,m) \in y} g(x, h, m)$$

and the score for a dependency parse is

$$\theta \cdot f(x, y) = \sum_{(h,m) \in y} \theta \cdot g(x, h, m)$$

Consider an input sentence $x_1 \ldots x_n$ that we wish to parse. We will construct a special context-free grammar for the sentence such that there is a one-to-one mapping between parse trees in the context-free grammar, and dependency structures. Each rule in the context-free grammar has an associated score; the score for the entire parse tree is the sum of scores for the rules that it contains; this score is equal to the score for the dependency parse corresponding to the parse tree.

The grammar we construct will be used to parse the input

$$0.2 \;\; 1.1 \;\; 1.2 \;\; 2.1 \;\; 2.2 \;\; \ldots \;\; n.1 \;\; n.2$$

The context-free grammar for a sentence $x_1 \ldots x_n$ is the following:

For $i = 1 \ldots n$, introduce the rule

$$C[i,i,l,1] \rightarrow i.1 \text{ with Score = 0}$$

For $i = 0 \ldots n$, introduce the rule

$$C[i,i,r,1] \rightarrow i.2 \text{ with Score = 0}$$

For all i, j, k such that $0 \leq i \leq k < j \leq n$, generate the following rules:

$$\begin{array}{lcl}
C[i,j,l,0] & \rightarrow & C[i,k,r,1] \;\; C[k+1,j,l,1] \text{ with score } \theta \cdot g(x, j, i) \\
C[i,j,r,0] & \rightarrow & C[i,k,r,1] \;\; C[k+1,j,l,1] \text{ with score } \theta \cdot g(x, i, j) \\
C[i,j,l,1] & \rightarrow & C[i,k,l,1] \;\; C[k,j,l,0] \text{ with score 0} \\
C[i,j,r,1] & \rightarrow & C[i,k+1,r,0] \;\; C[k+1,j,r,1] \text{ with score 0}
\end{array}$$

The root symbol in the context-free grammar is

$$C[0,n,r,1]$$

Question 2a How say we parse the sentence $x_1 \ldots x_n =$ *John saw Mary*. Show the parse tree corresponding to the dependency structure where *saw* is the head word for the entire sentence, and *John* and *Mary* are both modifiers to *saw*.

Question 3 Recall that a Brown clustering model consists of:

- A vocabulary $\mathcal{V}$
- A function $C : \mathcal{V} \rightarrow \{1, 2, \ldots k\}$ defining a *partition* of the vocabulary into k classes
- A parameter $e(v|c)$ for every $v \in \mathcal{V}$, $c \in \{1 \ldots k\}$
- A parameter $q(c'|c)$ for every $c', c \in \{1 \ldots k\}$

Recall also that given a corpus consisting of a sequence of words $w_1 \ldots w_n$, the quality of a Brown clustering model defined by C, e and q, is

$$\text{Quality}(C, e, q) = \sum_{i=1}^{n} \log e(w_i|C(w_i))q(C(w_i)|C(w_{i-1}))$$

Question 3a Now say our corpus is the sentence

the dog the dog the dog the dog

and we have $k = 2$. What are the optimal values of C, e and q?