A Kneser-Ney mini-example

January 18, 2008

Imagine we have the following corpus:

$$*\texttt{S}*\ a\ a\ b\ *\texttt{E}*$$

Let us look at what a bigram Kneser-Ney model looks like for this corpus. We use the Chen and Goodman (1998) recursion. We start with $P(a|*\texttt{S}*)$. We have:

$$P^2_{KN}(a|*\texttt{S}*) = \frac{\max\{c(*\texttt{S}*, \texttt{a}) - D, 0\}}{\sum_{w'} c(*\texttt{S}*, \texttt{w}')} + \frac{D_2}{\sum_{w'} c(*\texttt{S}*, \texttt{w}')} N_{1+}(*\texttt{S}*\bullet) P^1_{KN}(a) \tag{1}$$

where

$$P^1_{KN}(a) = \frac{\max\{N_{1+}(\bullet\texttt{a}) - D, 0\}}{\sum_{w'} N_{1+}(\bullet, \texttt{w}')} + \frac{D_1}{\sum_{w'} N_{1+}(\bullet\bullet)} N_{1+}(\bullet) P^0_{KN}(a) \tag{2}$$

and

$$P^0_{KN}(a) = \frac{1}{V} = \frac{1}{3} \tag{3}$$

First, we note that $N_{1+}(\bullet) = V$. This allows us to simplify the first-order backoff model:

$$P^1_{KN}(a) = \frac{\max\{N_{1+}(\bullet\texttt{a}) - D, 0\}}{\sum_{w'} N_{1+}(\bullet, \texttt{w}')} + \frac{D_1}{\sum_{w'} N_{1+}(\bullet\bullet)} \tag{4}$$

Furthermore, regardless of the event outcome in the first order model, it will be observed at least once. Therefore we have that $\max\{N_{1+}(\bullet\texttt{a}) - D_1, 0\} = N_{1+}(\bullet\texttt{a}) - D_1$ and hence

$$P^1_{KN}(a) = \frac{N_{1+}(\bullet\texttt{a})}{\sum_{w'} N_{1+}(\bullet, \texttt{w}')} \tag{5}$$

which is equivalent to equation (23) in Chen and Goodman (1998). In sum this gives us (letting $D = D_2$)

$$P^2_{KN}(a|{*}\mathtt{S}{*}) = \frac{\max\{c({*}\mathtt{S}{*},\mathtt{a}) - D, 0\}}{\sum_{w'} c({*}\mathtt{S}{*},\mathtt{w'})} + \frac{D}{\sum_{w'} c({*}\mathtt{S}{*},\mathtt{w'})} N_{1+}({*}\mathtt{S}{*}\bullet) \frac{N_{1+}(\bullet\mathtt{a})}{\sum_{w'} N_{1+}(\bullet,\mathtt{w'})} \tag{6}$$

$$\tag{7}$$

As an example, let's work out the probabilities with discounts $D = 0.1$. For $P(a|{*}\mathtt{S}{*})$ this gives us:

$$c({*}\mathtt{S}{*}, a) = 1 \tag{8}$$

$$\sum_{w'} c({*}\mathtt{S}{*},\mathtt{w'}) = 4 \tag{9}$$

$$N_{1+}({*}\mathtt{S}{*}\bullet) = 1 \tag{10}$$

$$N_{1+}(\bullet a) = 2 \tag{11}$$

$$\sum_{w'} N_{1+}(\bullet,\mathtt{w'}) = 4 \tag{12}$$

$$\tag{13}$$

giving us

$$P^2_{KN}(a|{*}\mathtt{S}{*}) = \frac{1-0.1}{1} + \frac{0.1}{1} \times 1 \times \frac{2}{4} \tag{14}$$

$$= 0.95 \tag{15}$$

Likewise we get

$$P^2_{KN}(b|{*}\mathtt{S}{*}) = \frac{0}{1} + \frac{0.1}{1} \times 1 \times \frac{1}{4} \tag{16}$$

$$= 0.025 \tag{17}$$

$$P^2_{KN}({*}\mathtt{E}{*}|{*}\mathtt{S}{*}) = 0.025 \tag{18}$$

References

Chen, S. and Goodman, J. (1998). An empirical study of smoothing techniques for language modeling. Technical report, Computer Science Group, Harvard University.