Katz Back-Off Models (Bigrams)

- For a bigram model, define two sets

$$\mathcal{A}(w_{i-1}) = \{w : \text{Count}(w_{i-1}, w) > 0\}$$

$$\mathcal{B}(w_{i-1}) = \{w : \text{Count}(w_{i-1}, w) = 0\}$$

These are the set of tail words on observed bigrams that start with w_{i-1}

These are the set of tail words on unobserved bigrams that start with w_{i-1}

- A bigram model

$$①\ q_{BO}(w_i \mid w_{i-1}) = \begin{cases} ②\ \frac{\text{Count}^*(w_{i-1}, w_i)}{\text{Count}(w_{i-1})} & \text{If } w_i \in \mathcal{A}(w_{i-1}) \\ ③\ \alpha(w_{i-1}) \boxed{\frac{q_{ML}(w_i)}{\sum_{w \in \mathcal{B}(w_{i-1})} q_{ML}(w)}} & \text{If } w_i \in \mathcal{B}(w_{i-1}) \end{cases}$$

④

where

This is the probability mass taken by discounting

$$\alpha(w_{i-1}) = 1 - \sum_{w \in \mathcal{A}(w_{i-1})} \frac{\text{Count}^*(w_{i-1}, w)}{\text{Count}(w_{i-1})}$$

① Says that if the bigram (w_{i-1}, w_i) is observed, estimate the probability of $q(w_i \mid w_{i-1})$ by equation ②.

If (w_{i-1}, w_i) is not observed, use equation ③ to estimate $q(w_i \mid w_{i-1})$.

Equ ② says to use the ratio of the discounted (w_{i-1}, w_i) bigram count divided by unigram count of w_{i-1} for the estimate of $q(w_i \mid w_{i-1})$

Equ ③ says to distribute the probability mass taken by discounting and distribute it over all the words in B in proportion to their unigram MLE.

Example from Columbia NLP week 1, sec 6.2 quiz:

corpus: the book STOP
his house STOP

discount

If $c^*(v, w) = c(v, w) - 0.5$ what is $q_{BO}(\text{book} \mid \text{his})$ estimated from this corpus?

unigram counts

A(his) = {house} 1

B(his) = {his, the, book, STOP} 1, 1, 1, 2

x	c	c*
his	1	
his, house	1	0.5

Here, $w \in A(\text{his})$ only contains the word "house". In practice, we'll need to sum over many tail words in observed bigrams

$$\alpha(\text{his}) = 1 - \sum_{\text{All } w \in A(\text{his})} \frac{c^*(\text{his}, w)}{c(\text{his})} = 1 - \frac{c^*(\text{his}, \text{house})}{1} = 1 - 0.5 = 0.5$$

Katz Back-Off Models (Trigrams)

- For a trigram model, first define two sets — Set of tail words for observed trigrams

$$\mathcal{A}(w_{i-2}, w_{i-1}) = \{w : \text{Count}(w_{i-2}, w_{i-1}, w) > 0\}$$
$$\mathcal{B}(w_{i-2}, w_{i-1}) = \{w : \text{Count}(w_{i-2}, w_{i-1}, w) = 0\}$$

Set of tail words for unobserved trigrams

- A trigram model is defined in terms of the bigram model:

$$q_{BO}(w_i \mid w_{i-2}, w_{i-1}) = \begin{cases} \frac{\text{Count}^*(w_{i-2}, w_{i-1}, w_i)}{\text{Count}(w_{i-2}, w_{i-1})} & \text{If } w_i \in \mathcal{A}(w_{i-2}, w_{i-1}) \\ \frac{\alpha(w_{i-2}, w_{i-1}) q_{BO}(w_i \mid w_{i-1})}{\sum_{w \in \mathcal{B}(w_{i-2}, w_{i-1})} q_{BO}(w \mid w_{i-1})} & \text{If } w_i \in \mathcal{B}(w_{i-2}, w_{i-1}) \end{cases}$$

discounted count of observed trigrams that start with (w_{i-2}, w_{i-1})

total bigram count

5 redistribution of discounted probability mass over unobserved trigrams in proportion to the bigram tail's

where

discounted probability mass

$$\alpha(w_{i-2}, w_{i-1}) = 1 - \sum_{w \in \mathcal{A}(w_{i-2}, w_{i-1})} \frac{\text{Count}^*(w_{i-2}, w_{i-1}, w)}{\text{Count}(w_{i-2}, w_{i-1})}$$

amount of probability mass taken from discounting observed trigrams

5 Redistribution of probability density - If (w_{i-2}, w_{i-1}, w) has not been observed, the amount of discounted probability density taken from the counts of the observed trigrams is distributed proportionally to the backed-off bigram probability parameter $\boxed{q_{BO}(w_i \mid w_{i-1})}$. This parameter is calculated using the Katz Back-Off Model for Bigrams described on the previous page (note that it's recursive).

The denominator in 5 is sum of all $q_{BO}(w \mid w_{i-1})$ for each $w \in \mathcal{B}(w_{i-2}, w_{i-1})$. These are the bigrams that are the last 2 words of unobserved trigrams.

Extend simple bigram example to include trigrams

2 trigrams of interest / the example corpus

x	c	$c^*_{\gamma=0.5}$	$c^*_{\gamma=0.7}$
the book EOS	5	4.5	4.3
the house EOS	3	2.5	2.3
buy the book EOS	4		
sell the book EOS	1		
buy the house EOS	2		
paint the house EOS	1		

*unigram table:

w	c	
the	8	✓
book	5	✓
house	3	
buy	6	
sell	1	✓
paint	1	✓
EOS	8	✓
	32	

End Of Sentence = EOS

* For complete bigram & trigram tables see Rmd File

Find $q_{BO}(\underset{w_i}{\text{house}} \mid \underset{w_{i-2}}{\text{sell}}, \underset{w_{i-1}}{\text{the}})$

$A(w_{i-2}, w_{i-1}) = A(\text{sell}, \text{the}) = \{\text{book}\}$ only 1 observed trigram starting with (w_{i-2}, w_{i-1})

$B(w_{i-2}, w_{i-1}) = B(\text{sell}, \text{the}) = \{\text{the}, \text{house}, \text{buy}, \text{sell}, \text{paint}, \text{EOS}\}$

$$\alpha(w_{i-2}, w_{i-1}) = \alpha(\text{sell}, \text{the}) = 1 - \sum_{w \in A} \frac{c^*(w_{i-2}, w_{i-1}, w)}{c(w_{i-2}, w_{i-1})}$$

$$\alpha(w_{i-2}, w_{i-1})_{\gamma=0.5} = 1 - \frac{c^*(\text{sell}, \text{the}, \text{book})}{c(\text{sell}, \text{the})} = 1 - \left(\frac{1-0.5}{1}\right) = \underline{\underline{0.5}}$$

unobserved trigrams that start with (w_{i-2}, w_{i-1}) are:

(6) $q_{BO}(w_i \mid w_{i-1}) = q_{BO}(\text{house} \mid \text{the})$

(7) $\boxed{\sum_{w \in B} q(w \mid w_{i-1})} = q_{BO}(\text{the} \mid \text{the}) + q_{BO}(\text{house} \mid \text{the}) + q_{BO}(\text{buy} \mid \text{the}) + q_{BO}(\text{sell} \mid \text{the}) + q_{BO}(\text{paint} \mid \text{the}) + q_{BO}(\text{STOP} \mid \text{the})$

denominator of (5)

Sell the the
Sell the house
Sell the buy
Sell the sell
Sell the paint
Sell the EOS

Each of the terms in (6) & (7) are calculated using the KBO Bigrams Model

KBO

Trigram example (cont.)

Now, just need to calc. the bigram q_{BO}'s and put it all together. Start w/ unseen bigrams:

At the bigram level: $A(w_{i-1}) = A(\text{the}) = \{\text{book}, \text{house}\}$
$B(w_{i-1}) = B(\text{the}) = \{\text{the}, \text{buy}, \text{sell}, \text{paint}, \text{EOS}\}$

$$\alpha(w_{i-1}) = \alpha(\text{the}) = 1 - \left[\frac{c^*(\text{the}, \text{book})}{c(\text{the})} + \frac{c^*(\text{the}, \text{house})}{c(\text{the})}\right]$$

$$= 1 - \left(\frac{4.5}{8} + \frac{2.5}{8}\right) = \frac{1}{8} = 0.125$$

$$\sum_{w \in B(w_{i-1})} q_{ML}(w) = \begin{cases} q_{ML}(\text{the}) + q_{ML}(\text{buy}) + q_{ML}(\text{sell}) + \\ q_{ML}(\text{paint}) + q_{ML}(\text{STOP}) \end{cases}$$

$q_{ML}(\text{the}) = \frac{8}{32}$, $q_{ML}(\text{buy}) = \frac{6}{32}$, $q_{ML}(\text{sell}) = \frac{1}{32}$,

$q_{ML}(\text{paint}) = \frac{1}{32}$, $q_{ML}(\text{STOP}) = \frac{8}{32}$ $\therefore$

$$\sum_{w \in B(w_{i-1})} q_{ML}(w) = \frac{1}{32}(8 + 6 + 1 + 1 + 8) = \frac{24}{32} \quad \therefore$$

from eqn ③ of KBO Bigram Models page 7/11

$$\alpha(w_{i-1})$$

$$q(\text{the} \mid \text{the}) = \frac{1}{8}\left(8/32\right) / \left(24/32\right) = \frac{1}{8}\left(\frac{8}{24}\right) = \frac{8}{192}$$

$$q(\text{buy} \mid \text{the}) = \frac{1}{8}\left(6/32\right) / \left(24/32\right) = \frac{1}{8}\left(\frac{6}{24}\right) = \frac{6}{192}$$

$$q(\text{sell} \mid \text{the}) = \frac{1}{8}\left(1/32\right) / \left(24/32\right) = \frac{1}{8}\left(\frac{1}{24}\right) = \frac{1}{192}$$

$$q(\text{paint} \mid \text{the}) = q(\text{sell} \mid \text{the}) = \frac{1}{192}$$

$$q(\text{STOP} \mid \text{the}) = q(\text{the} \mid \text{the}) = \frac{8}{192}$$

For our observed bigram: $q(\text{house} \mid \text{the}) = \dfrac{c^*(\text{the}, \text{house})}{c(\text{the})}$

$q(\text{house} \mid \text{the}) = \frac{2.5}{8} = 0.3125$

{We now have everything to calc. $q_{BO}(\text{house} \mid \text{sell}, \text{the})$}

KBO Trigram example (conclusion)

$$q_{BO}(\text{house} \mid \text{sell, the}) = \frac{\alpha(\text{sell, the})\, q_{BO}(\text{house} \mid \text{the})}{\left[q_{BO}(\text{the} \mid \text{the}) + q_{BO}(\text{house} \mid \text{the}) + q_{BO}(\text{buy} \mid \text{the}) + q_{BO}(\text{sell} \mid \text{the}) + q_{BO}(\text{paint} \mid \text{the}) + q_{BO}(\text{STOP} \mid \text{the})\right]}$$

$$\text{numerator} = \alpha(\text{sell, the})\, q_{BO}(\text{house} \mid \text{the}) = 0.5\left(\frac{2.5}{8}\right) = 0.5\left(\frac{60}{192}\right)$$

$$\text{denominator} = \frac{1}{192}(8+6+1+1+8) + \frac{2.5}{8} = \frac{24}{192} + \frac{2.5}{8} = \frac{24}{192} + \frac{24(2.5)}{192} = \frac{84}{192} = \delta$$

$$q_{BO}(\text{house} \mid \text{sell, the}) = (0.5)\frac{(60/192)}{(84/192)} = 0.5(0.7143)$$

$$= \underline{\underline{0.3571}}$$

Sanity check: $\sum_{w \in B} q_{BO}(w \mid \text{sell, the}) = 1$?

$$q_{BO}(\text{book} \mid \text{sell, the}) = \frac{c^*(\text{sell, the, book})}{c(\text{sell, the})} = \frac{0.5}{1} = 0.5$$

$$q_{BO}(\text{the} \mid \text{sell, the}) = \alpha(\text{sell, the})\frac{q_{BO}(\text{the} \mid \text{the})}{\delta} = \frac{(0.5)(8/\cancel{192})}{(84/\cancel{192})}$$

$$= q_{BO}(\text{STOP} \mid \text{sell, the}) = 0.0476$$

$$q_{BO}(\text{buy} \mid \text{sell, the}) = \frac{(0.5)(6/192)}{84/192} = 0.0357$$

$$q_{BO}(\text{sell} \mid \text{sell, the}) = q_{BO}(\text{paint} \mid \text{sell, the}) = \frac{(0.5)(1/192)}{84/192} = 0.0060$$

0.0119

$$(0.3571) + (0.5) + 2(0.0476) + 0.0357 + 2(0.0060) = \underline{\underline{1.0000}}$$