

Katz Back-Off Models (Bigrams)

- For a bigram model, define two sets

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

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

- A bigram model

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

where

This is the probability mass taken by discounting

$$\textcircled{4} \quad \alpha(w_{i-1}) = 1 - \sum_{w \in A(w_{i-1})} \frac{\text{Count}^*(w_{i-1}, w)}{\text{Count}(w_{i-1})}$$

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

Eqn $\textcircled{2}$ 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 $q(w_i | w_{i-1})$

Eqn $\textcircled{3}$ says to distribute the probability/mass taken by discounting and distribute it over all the words in B in proportion to their unigram ML.

Example from Columbia NLP week 1, Sec 6.2 quiz:

corpus: the book STOP
 his house STOP

$$A(\text{his}) = \{\text{house}\}$$

$$B(\text{his}) = \{\text{his, the, book, STOP}\}$$

$$C. \quad c^*$$

$$\text{his} \quad 1$$

$$\text{his, house} \quad 1 \quad 0.5$$

$$\text{If } c^*(\text{the, w}) = c(\text{the, w}) - 0.5$$

What is $q_{BO}(\text{book} | \text{his})$ estimated from this corpus?

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

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

Katz Back-Off Models (Trigrams)

- For a trigram model, first define two sets

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

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

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

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

(5)

where

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

Set of tail words for observed trigrams
Set of tail words for unobserved trigrams
Count of observed trigrams that start with (w_{i-2}, w_{i-1})
total bigram count
Redistribution of discounted probability mass over unobserved trigrams in proportion to the bigram tails
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 $q_{BO}(w_i | w_{i-1})$. 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 $q_{BO}(w | w_{i-1})$

for each $w \in B(w_{i-2}, w_{i-1})$

These are the bigrams that are the last 2 words of unobserved trigrams

questions also (in) A 2 word set
writing w1 "and brown set
not previously seen in it been seen
Dancing N brown is a brown

Extend simple bigram example to include trigrams

* unigram table:

<u>X</u>	<u>C</u>	<u>C^*</u> $\gamma=0.5$	<u>C^*</u> $\gamma=0.75$	<u>w</u>	<u>c</u>
{ the book EOS	5	4.5	4.3	the	8 ✓
{ the house EOS	3	2.5	2.3	book	5 ✓
buy the book EOS	4			house	3
Sell the book EOS	1			buy	6
buy the house EOS	2			Sell	1 ✓
paint the house EOS	1			paint	1 ✓
					8 ✓
				End of Sentence = EOS	

* For complete bigram & trigram tables
see Read file

Find q_{BO} (house | sell, the)

32

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

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

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

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

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

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

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

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

KBD

Trigram example (cont.)

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

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

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

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

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

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

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

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

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

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

$$q(\text{buy}|\text{the}) = \frac{1}{8} \left(\frac{6}{32} \right) / \left(\frac{24}{32} \right) = \frac{1}{8} \left(\frac{6}{24} \right) = \frac{5}{192}$$

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

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

$$q(\text{STOP}|\text{the}) = q(\text{the}|\text{the}) = \frac{1}{192}$$

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

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

$$= 0.3125$$

{we now have everything to calc. $q_{FB0}(\text{house}|\text{sell, the})$ }

From next page
③ of notes
Bigram pg 7/11

KBO Trigram example (conclusion)

$$q_{f_{BO}}(\text{house sell, the}) =$$

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

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

$$\text{denominator} = \frac{1}{192} (8+8+1+1+8) + \frac{2.5}{8} \\ = \frac{24}{192} + \frac{2.5}{8} = \frac{24}{192} + \frac{24(2.5)}{192} = \frac{84}{192} = 5$$

$$q_{f_{BO}}(\text{house} | \text{sell, the}) = (0.5) \frac{\left(\frac{60}{192} \right)}{\left(\frac{84}{192} \right)} = 0.5(0.7143) \\ = \underline{\underline{0.3571}}$$

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

$$q_{f_{BO}}(\text{book} | \text{sell, the}) = \frac{C^*(\text{sell, the, book})}{C(\text{sell, the})} = \frac{0.5}{1} = 0.5$$

$$= \frac{q_{f_{BO}}(\text{the} | \text{sell, the})}{q_{f_{BO}}(\text{STOP} | \text{sell, the})} = \alpha(\text{sell, the}) \frac{q_{f_{BO}}(\text{the} | \text{the})}{5} = \frac{(0.5)(8/192)}{(84/192)}$$

$$= \frac{q_{f_{BO}}(\text{STOP} | \text{sell, the})}{q_{f_{BO}}(\text{sell} | \text{sell, the})} = 0.0476$$

$$q_{f_{BO}}(\text{buy} | \text{sell, the}) = \frac{(0.5)(5/192)}{84/192} = 0.0357$$

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

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