Kneser-Ney Bigram Example

Unigram Model

counts	big	cats	chase	dogs	i	like	walks
	5	2	2	4	1	2	1

i like big dogs
big dogs like walks
big dogs chase big cats
cats chase big dogs

Unigram Model with OOV Handling

Consider all words that occur only once in the training data as OOV, and replace them with the special token <UNK>. The OOV words in our sample corpus are *i* and *walks*. The updated unigram count table is now:

	UNK	big	cats	chase	dogs	like
counts	2	5	2	2	4	2
probs	$\frac{2}{17}$	$\frac{5}{17}$	$\frac{2}{17}$	$\frac{2}{17}$	$\frac{4}{17}$	$\frac{2}{17}$

Bigram Model with OOV Handling

Replace the OOV words in the training data with the UNK token, and add sentence markers. The UNK token is counted like any other word in the count table:

counts	UNK	big	cats	chase	dogs	like	</s>
UNK						1	1
big			1		4		
cats				1			1
chase		2					
dogs				1		1	2
like	1	1					
<s>	1	2	1				

<s> UNK like big dogs </s>
<s> big dogs like UNK </s>
<s> big dogs chase big cats </s>
<s> cats chase big dogs </s>

Bigram Model with OOV Handling and Laplace (Add-1) Smoothing

Add 1 to all counts and compute probabilities:

probs	UNK	big	cats	chase	dogs	like	</s>
UNK	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{2}{9}$
big	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{2}{12}$	$\frac{1}{12}$	$\frac{5}{12}$	$\frac{1}{12}$	$\frac{1}{12}$
cats	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{2}{9}$
chase	$\frac{1}{9}$	$\frac{3}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
dogs	$\frac{1}{11}$	$\frac{1}{11}$	$\frac{1}{11}$	$\frac{2}{11}$	$\frac{1}{11}$	$\frac{2}{11}$	$\frac{3}{11}$
like	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
<s>	$\frac{2}{11}$	$\frac{3}{11}$	$\frac{2}{11}$	$\frac{1}{11}$	$\frac{1}{11}$	$\frac{1}{11}$	$\frac{1}{11}$

<s> UNK like big dogs </s>
<s> big dogs like UNK </s>
<s> big dogs chase big cats </s>
<s> cats chase big dogs </s>

Kneser-Ney Bigram Example

Kneser-Ney Smoothing

	<UNK>	big	cats	chase	dogs	like	</s>	$C(w_1)$	$C(w_1\bullet)$
<UNK>						1	1	2	2
big			1		4			5	2
cats				1			1	2	2
chase		2						2	1
dogs				1		1	2	4	3
like	1	1						2	2
<s>	1	2	1					4	3
$C(\bullet w_2)$	2	3	2	2	1	2	3		15

<s> <UNK> like big dogs </s>
<s> big dogs like <UNK> </s>
<s> big dogs chase big cats </s>
<s> cats chase big dogs </s>

BT : bigram type
$C(w_1)$: unigram count of w_1
$C(w_1\bullet)$: BTs starting with w_1
$C(\bullet w_2)$: BTs ending with w_2

$$P_{KN}(w_2|w_1) = \frac{max(C(w_1 w_2) - d, 0)}{C(w_1)} + \lambda(w_1) \times P_{CONT}(w_2)$$

$$\lambda(w_1) = \frac{d}{C(w_1)} \times C(w_1\bullet)$$

$$P_{CONT}(w_2) = \frac{C(\bullet w_2)}{C(\text{non-zero bigram types})}$$

$$\begin{aligned}
P_{KN}(dogs|big) &= \frac{max(C(\text{big dogs}) - d, 0)}{C(\text{big})} + \lambda(big) \times P_{CONT}(dogs) \\
&= \frac{max(C(\text{big dogs}) - d, 0)}{C(\text{big})} + \frac{d}{C(big)} \times C(big\bullet) \times \frac{C(\bullet dogs)}{C(\text{non-zero bigram types})} \\
&= \frac{max(4 - .75, 0)}{5} + \frac{.75}{5} \times 2 \times \frac{1}{15} \\
&= .67
\end{aligned}$$

KN *big* row	<UNK>	big	cats	chase	dogs	like	</s>
big (T_1)	0	0	.05	0	.65	0	0
$\lambda(big)$	.3	.3	.3	.3	.3	.3	.3
$P_{CONT}(w_2)$	2/15	3/15	2/15	2/15	1/15	2/15	3/15
$P_{\text{KN}} =$ $T_1 + (\lambda * P_{CONT})$	.04	.06	.09	.04	.67	.04	.06

Add1 *big* row	<UNK>	big	cats	chase	dogs	like	</s>
$P_{\text{Add1}}(w_2|big)$	.083	.083	.167	.083	.417	.083	.083

KN *dogs* row	<UNK>	big	cats	chase	dogs	like	</s>
dogs (T_1)	0	0	0	.0625	0	.0625	.3125
$\lambda(dogs)$	.5625	.5625	.5625	.5625	.5625	.5625	.5625
$P_{CONT}(w_2)$	2/15	3/15	2/15	2/15	1/15	2/15	3/15
$P_{\text{KN}} =$ $T_1 + (\lambda * P_{CONT})$	.075	.1125	.075	.1375	.0375	.1375	.425

PP(big cats like big dogs) Unigram: ≈ 5.13, Bigram (Add-1): ≈ 4.46, Bigram (KN): ≈ 3.87

Exercise: Compute $P_{KN}(chase|cats)$