

Absolute Discounting:

$$p(w_i | w_{i-1}, \dots, w_{i-n+1}) = \frac{\max(c(w_i, w_{i-1}, \dots, w_{i-n+1}) - \delta, 0)}{c(w_{i-1}, \dots, w_{i-n+1})} + \lambda \bar{p}(w_i | w_{i-1}, \dots, w_{i-n+1})$$

$$\lambda = \frac{\delta S}{c(w_{i-1}, \dots, w_{i-n+1})}$$

As we know,  $\sum_{w_i} p(w_i | w_{i-1}, \dots, w_{i-n+1}) = 1$ , ( $w_i$ : seen words + unseen words).

Therefore, we can get:

$$\sum_{w_i} p(w_i | w_{i-1}, \dots, w_{i-n+1}) = \sum_{w_i} \frac{\max(c(w_i, w_{i-1}, \dots, w_{i-n+1}) - \delta, 0)}{c(w_{i-1}, \dots, w_{i-n+1})} + \sum_{w_i} \frac{\delta S}{c(w_{i-1}, \dots, w_{i-n+1})} \bar{p}(w_i | w_{i-1}, \dots, w_{i-n+1})$$

(for seen words in  $w_i$ ,  $\max(c(w_i, w_{i-1}, \dots, w_{i-n+1}) - \delta, 0) = c(w_i, w_{i-1}, \dots, w_{i-n+1}) - \delta$ .)

(for unseen words in  $w_i$ ,  $\max(\dots) = 0$ .)

If we assume there are  $k$  seen words, then we minus  $\delta$   $k$  times.

Rewrite part ① as:

$$\sum_{w_i} \frac{c(w_i, w_{i-1}, \dots, w_{i-n+1})}{c(w_{i-1}, \dots, w_{i-n+1})} + \frac{-k\delta}{c(w_{i-1}, \dots, w_{i-n+1})}$$

Rewrite part ② as:

$$\frac{\delta S}{c(w_{i-1}, \dots, w_{i-n+1})} \cdot \sum_{w_i} \bar{p}(w_i | w_{i-1}, \dots, w_{i-n+1})$$

$$\text{Since } ① + ② = 1, \quad 1 - \frac{k\delta}{c(\cdot)} + \frac{\delta S}{c(\cdot)} = 1$$

$$\therefore k = S$$

Therefore,  $k$  is the words that following  $w_{i-1}$ .

possible number of words.

good exploration!!

f2

## 2.1.2 Top ten words following "good".

Linear interpolation

```
good-eo 0.34524085740673754
good-and 0.05010133909318111
good-food 0.03186822966834152
good-as 0.027009920446559355
good-but 0.026674797534351265
good-for 0.015862224660455344
good-thing 0.014018372379815121
good-place 0.01349218298600822
good-servic 0.013450408309584631
good-too 0.012231227555345938
0.9990103672618639
```

Absolute discount

```
good-eo 0.37149468215972925
good-and 0.05246466115743195
good-food 0.034807923350750905
good-as 0.029610958162301183
good-but 0.02878405470477222
good-for 0.016526845641099585
good-thing 0.01545387504906812
good-servic 0.014674555410421215
good-place 0.014399242189518748
good-too 0.013409217769159092
0.9989004080935491
```