SNS COLLEGE OF ENGINEERING

Kurumbapalayam(Po), Coimbatore – 641 97
Accredited by NAAC-UGC with 'A' Grade
Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

Department of Artificial Intelligence and Data Science

Course Name – 19AD601 – Natural Language Processing

III Year / VI Semester

Unit 2 – WORD LEVEL ANALYSIS

Topic 3- Smoothing

Smoothing

To keep a language model from assigning zero probability to unseen events, we'll have to shave off a bit of probability mass from some more frequent events and give it to the events we've never seen.

This modification is called smoothing (or discounting).

There are many ways to do smoothing, and some of them are:

- Add-1 smoothing (Laplace Smoothing)
- Add-k smoothing,
- Backoff
- Kneser-Ney smoothing.

Smoothing

Add-1 smoothing (Laplace Smoothing)
The simplest way to do smoothing is to add one to all the n-gram counts, before we normalize them into probabilities.

All the counts that used to be zero will now have a count of 1, the counts of 1 will be 2, and so on. This algorithm is called Laplace smoothing.

Laplace smoothing does not perform well enough to be used in modern n-gram models, but it usefully introduces many of the concepts that we see in other smoothing algorithms, gives a useful baseline, and is also a practical smoothing algorithm for other tasks like text classification.

Smoothing

Laplace smoothing to unigram probabilities

The unsmoothed maximum likelihood estimate of the unigram probability of the word wi is its count ci normalized by the total number of word tokens N

$$P(w_i) = \frac{c_i}{N}$$

Laplace smoothing adds one to each count. Since there are V words in the vocabulary and each one was incremented, we also need to adjust the denominator to take into account the extra V observations.

$$P_{\text{Laplace}}(w_i) = \frac{c_i + 1}{N + V}$$

Smoothing

Add-k smoothing

One alternative to add-one smoothing is to move a bit less of the probability mass from the seen to the unseen events.

Instead of adding 1 to each count, we add a fractional count k (.5? .05? .01?). This algorithm is called add-k smoothing.

$$P^*_{\text{Add-k}}(w_n|w_{n-1}) = \frac{C(w_{n-1}w_n)+k}{C(w_{n-1})+kV}$$

Smoothing

Add-k smoothing

One alternative to add-one smoothing is to move a bit less of the probability mass from the seen to the unseen events.

Instead of adding 1 to each count, we add a fractional count k (.5? .05? .01?). This algorithm is called add-k smoothing.

$$P^*_{\text{Add-k}}(w_n|w_{n-1}) = \frac{C(w_{n-1}w_n)+k}{C(w_{n-1})+kV}$$

THANK YOU