

N-Grams Overview

N-grams are fundamental and give you a foundation that will allow you to understand more complicated models in the specialization. These models allow you to calculate probabilities of certain words happening in a specific sequence. Using that, you can build an auto-correct or even a search suggestion tool.

Other applications of N-gram language modeling include:

Speech recognition



$P(\text{I saw a van}) > P(\text{eyes awe of an})$

Spelling correction



“He entered the **ship** to buy some groceries” - “ship” a dictionary word

- $P(\text{entered the shop to buy}) > P(\text{entered the ship to buy})$

Augmentative communication



Predict most likely word from menu for people unable to physically talk or sign.
(Newell et al., 1998)

This week you are going to learn to:

- Process a text corpus to N-gram language model
- Handle out of vocabulary words
- Implement smoothing for previously unseen N-grams
- Language model evaluation