CORK INSTITUTE OF TECHNOLOGY
INSTITIÚID TEICNEOLAÍOCHTA CHORCAÍ

http://www.cit.ie

Natural Language Processing Lab

Week4: Language Model

Praveen Joshi

12/10/2020

Slides credit: Niranjan Balasubramanian

What is Language modeling?

- Task of building a predictive model of language
- A language model is used to predict two types of quantities.

1. Probability of observing a sequence of words from a language.

   e.g., Pr(Colorless green ideas sleep furiously) = ?

2. Probability of observing a word having observed a sequence.

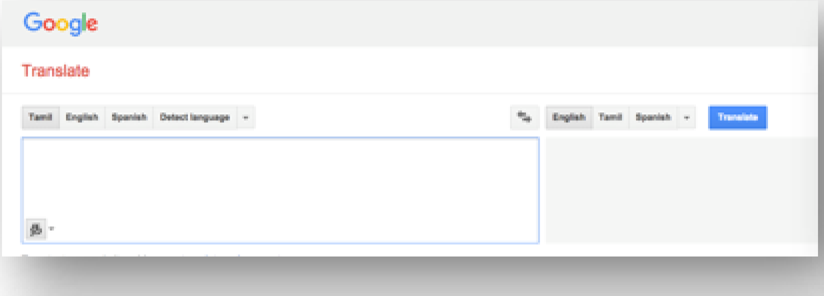
   e.g. Pr(furiously | Colorless green ideas) = ?

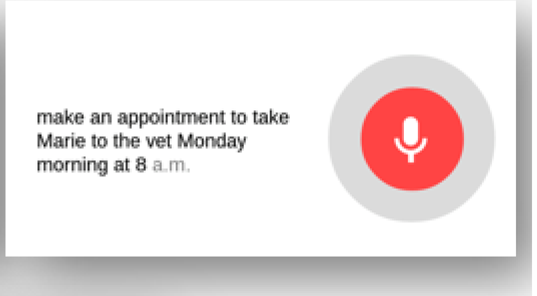
Why model Language?

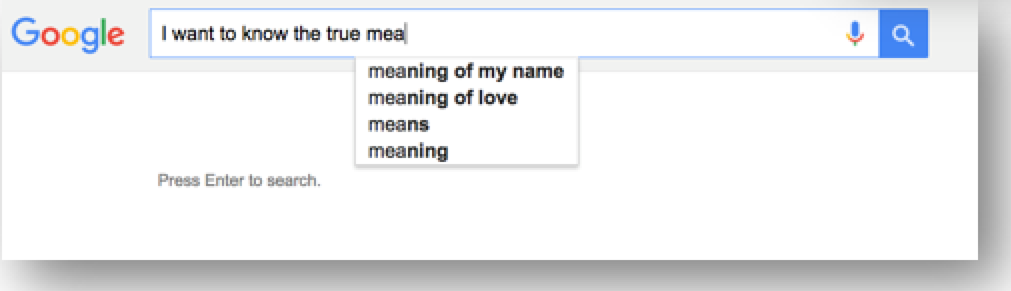
- The probability of observing a sequence is a measure of *goodness*.

- If a system outputs some piece of text, I can assess its goodness.
  - Many NLP applications output text.

- Example Applications
  - Speech recognition
  - OCR
  - Spelling correction
  - Machine translation
  - Authorship detection




make an appointment to take Marie to the vet Monday morning at 8 a.m.



I want to know the true mea

meaning of my name
meaning of love
means
meaning

Press Enter to search.

How to model language?

- Count! (and normalize).
  - Need some source text – corpora.

- Main Issues

Issue 1: We can generate infinitely many new sequences.

e.g., Colorless green ideas sleep furiously is not a frequent sequence.

Issue 2: We generate new words all the time.

e.g., Truthiness, #letalonethehashtags

Markov Assumption

- Next event in a sequence depends only on its immediate past (context).

$$Pr(w_{k+1} | \underbrace{w_{i-k}, w_{i-k+1}, \cdots, w_k}_{\text{Context}})$$

- *n*-grams

Unigrams	$\Pr(w_{k+1})$
Bigrams	$\Pr(w_{k+1} \mid w_k)$
Trigrams	$\Pr(w_{k+1} \mid w_{k-1}, w_k)$
4-grams	$\Pr(w_{k+1} \mid w_{k-2}, w_{k-1}, w_k)$

- Note:
  - Other contexts are possible and in many cases preferable.
  - Models tend to be more complex.

Reliable Estimation vs. Generalization

- We can estimate unigrams quite reliably but they are often not a good model.

- Higher order n-gram require large amounts of data but are better models.
  - However, they have a tendency to *overfit* the data.

- Example sentences generated from Shakespeare language models:

Unigram	Every enter now severally so, let.
Bigram	then all sorts, he is trim, captain.
Trigram	Indeed the duke; and had a very good friend.
4-gram	It cannot be but so.

CORK INSTITUTE OF TECHNOLOGY

INSTITIÚID TEICNEOLAÍOCHTA CHORCAÍ

http://www.cit.ie

Thank you

Praveen Joshi

12/10/2020