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 2- Evaluating N-grams

Evaluating N-grams

The best way to evaluate the performance of a language model is to embed it in an application and measure how much the application improves.

Extrinsic Evaluation

Extrinsic Evaluation of a N-gram language model is to use it in an application and measure how much the application improves.

To compare two language models A and B:

- Use each of language model in a task such as spelling corrector, MT system.
- Get an accuracy for A and for B
  - How many misspelled words corrected properly
  - How many words translated correctly

Compare accuracy for A and B. The model produces the better accuracy is the better model.

Extrinsic evaluation can be time-consuming.

Evaluating N-grams

Intrinsic Evaluation

An intrinsic evaluation metric is one that measures the quality of a model independent of any application.

When a corpus of text is given and to compare two different n-gram models,

- Divide the data into training and test sets,
- Train the parameters of both models on the training set, and
- Compare how well the two trained models fit the test set.
  - Whichever model assigns a higher probability to the test set

The probabilities of an n-gram model come from the corpus it is trained on, the training set or training corpus. We can then measure the quality of an n-gram model by its performance on some unseen data called the test set or test corpus

In practice, probability-based metric called perplexity is used instead of raw probability as our metric for evaluating language models.

Evaluating N-grams

Perplexity
The perplexity (sometimes called PPL for short) of a language model on a test set is the inverse probability of the test set, normalized by the number of words.

Minimizing perplexity is the same as maximizing probability
• The perplexity PP for a test set W=w1w2...wn is

$$\begin{aligned} \text{perplexity}(W) &= P(w_1 w_2 \ldots w_N)^{-\frac{1}{N}} \\ &= \sqrt[N]{\frac{1}{P(w_1 w_2 \ldots w_N)}} \end{aligned}$$

We can use the chain rule to expand the probability of W:

$$\text{perplexity}(W) = \sqrt[N]{\prod_{i=1}^{N} \frac{1}{P(w_i|w_1 \ldots w_{i-1})}}$$

Evaluating N-grams

Perplexity of Unigram

The perplexity of a test set W depends on which language model we use. Here's the perplexity of W with a unigram language model (just the geometric mean of the unigram probabilities):

$$\text{perplexity}(W) = \sqrt[N]{\prod_{i=1}^{N} \frac{1}{P(w_i)}}$$

Perplexity of Bi-gram

The perplexity of W computed with a bigram language model is still a geometric mean, but now of the bigram probabilities,

$$\text{perplexity}(W) = \sqrt[N]{\prod_{i=1}^{N} \frac{1}{P(w_i|w_{i-1})}}$$

THANK YOU