Language Models and RNN

Shubham Gupta

January 14, 2020

1 Introduction

- Introduce language modelling(LM)
- LM motivates building RNN

1.1 Language Modelling

- Predict what words come next
- Computes probability distribution of next work x^{t+1} given previous words $x^1, x^2...x^t$ where x is in the vocab V.
- Assign prob to a piece of text

$$\prod_{t=1}^{T} P(x^t|x(t-1), ..., x^1) \tag{1}$$

1.2 N-gram language model

- **n-gram** Chunk of n consecutive words
- Collect stats about how frequent different n-grams are, and use this to predict the next word.
- **Assumption**: x^{t+1} depends only on preceding $n-1$ words
- To get the probabilities, we will count them from a large corpus i.e *statistical approximation*

1.3 Problems

- Throws away too much context
- Sparsity. If the numerator is 0, then the chance of a valid word occuring is not possible, which is incorrect. Just because the n-words were not seen in the dataset does not mean that it is not a valid concept
- Solution? Smoothing. Add a small amount of probability *delta* for every word in the vocabulary
- If denominator is 0, cannot calculate the probability at all.

- Solution? If you cannot find n words in the dataset, backoff and just use the last n-1 or n-2 words instead. This is called **backoff**
- Sparsity problems increase with increase in n
- **Storage**: Size of model increases as the n-grams increase