Paper: *N-gram Language Models*

Razwan Ahmed Tanvir

Date: 07/1/2022

Quote *"The intuition of the n-gram model is that instead of computing the probability of a word given its entire history, we can approximate the history by just the last few words."*

Overview N- Gram model is a model which predicts the the next word for a sequence of words. An example could make N-Gram clear. From the paper- what will be the probability that the next word for the given sequence- "its water is so transparent that" will be "the"? One way to estimate this probability can be counting the frequency of the sequence with the word "the" in the corpus and calculate the probability. However, the authors mentioned that this approach is not effective as new sentences are always generating in the language, so counting the occurrence is not a good idea for this task. Authors of this paper proposed a better solution for this task where they intend to calculate the probability by looking at the frequency of lat few words rather than considering the whole history of that word in the corpus.

This research has several accomplishments and the outcome of this paper is paper is frequently used in natural language processing tasks. The proposed model is faster and accurate for real life applications. Moreover, the authors proposed an evaluation technique named Perplexity which essentially evaluates the N-Gram models which is currently famous method for model evaluation. The authors tackled the generalisation problem using Laplace Smoothing, Add-k smoothing, Back-off, Interpolation and Kneser-Ney smoothing.

Intellectual Merit The research has proposed a novel approach for predicting the next word which advanced the overall natural language processing techniques. It has used the Markov Assumption to get the probability. The researcher provided valid experiment results and accurate mathematical model to justify their claims. Authors suggested that the model could be evaluated extrinsically but this approach is costly. So, they have used the intrinsic approach which is Perplexity, a novel technique. The authors measured their model using this technique which is the weighted average of the branching factor of a language. The authors have the sufficient experience in this research filed and also have a deeper understanding of the probabilistic models.

Broader Impact This research benefits the natural language processing research and its applications in many ways. The authors contribution of the new approach to N-Gram and the Perplexity as the evaluation method has impacted the processing and analysing of the natural language. Moreover, this research is cited in many natural language processing research and the model has stable implementation in many languages. The author Daniel Jurafsky is a professor of Linguistics and Computer Science at Stanford University and James h. martin is a Professor of Computer Science at the University of Colorado, Boulder.Both of the authors are male and there is no under represented groups as authors.

Keywords Natural Language Processing, Markov Assumption, Chain rule of probability

Discussion Questions

- The authors mentioned about the usage of Neural Network can effectively minimize the problems with N-gram. Discussion on how to use a neural network based model to predict next word is important.
- Is there any way to minimize the parameters for the higher order N-grams?

Table 1: Grade deductions by section

Overview	Intellectual M.	B. Impact	Keywords	Questions	Is Online?