

Incomprehensible Style Generator

Joshua Garcia-Kimble

William Yang

Anxin Yi

1 Introduction

The goal of this project was to try and create a sentence with a similar style to the sentences that were passed in to train the model. To do this, we passed numerous sentences into the model. We used CKY Parser to parse these sentences based on the full.pcfg file given to us in hw4. We then used these parses to update the weights of the grammar rules in full.pcfg. We then reconstruct a parse tree based on the part of speech of a word passed into a model.

Our code is at the following git repository: [Code Base](#)

2 Experimental Setup

We used ChatGPT to generate two sets of 5 sentences in different styles. The first was in academic style, which gave us the following sentences to train on:

1. The current body of research suggests that there is a positive correlation between regular exercise and improved mental health outcomes.
2. The study's findings reveal a statistically significant difference between the experimental group and the control group in terms of their ability to recall information.
3. The author's analysis of the data demonstrates a clear trend towards greater income inequality in developed economies over the past several decades.
4. The researchers' methodology employed a mixed-methods approach, utilizing both quantitative and qualitative data sources to triangulate their results.
5. The literature review highlights the key theoretical frameworks and debates that have shaped the field of educational psychology over the past century.

The other set of sentences was in the style of a

children's book as follows:

1. Once upon a time, in a magical forest far, far away, there lived a mischievous little squirrel named Nutmeg.
2. The brave little mouse scurried across the meadow, her heart beating fast as she looked for a safe place to hide.
3. The little caterpillar ate and ate, munching on leaves and dreaming of the day when she would spread her wings and fly.
4. The playful puppy wagged his tail and barked with delight as he chased the colorful ball around the yard.
5. The tiny seed was planted in the rich soil, and with sunshine and rain, it grew into a tall, strong tree, providing shade and shelter for all the creatures of the forest.

We used the first set as the training data and generated a sentence from the seed word "Joshua". From here we evaluated the average F1 score between the generated sentence and the sentences in the two above corpora. We compared the average F1 score between the generated sentence and the training set of sentences versus the test set of sentences. Our hypothesis is that if the average F1 score for the training corpora is significantly higher than the average F1 score for the testing corpora, then our model works.

3 Results

Using a seed word "Joshua", our model generated a parsed tree corresponding to the sentence:

[a, United, body, terms, an, a, University, North, National, England, London, World, New, section, France, 2, The, ,, also, was, ,, the, group, of, over, 's, -, the, past, that, found, in, regular, for, there, control, to, their, century, is, between, several,

Training data	0.03
Testing data	0.02

Table 1: Average F1 score results

information, has, County, !, .]

Comparing this tree to the training and testing sentences gives us the F1 scores seen in Table 1.

We see that based on these results, the sentence generated by our model has higher similarity to the sentences that it was trained on compared to other sentences written in a different style. This evaluation provides support that our model is able to learn “style” from some training sentences and generate sentences that tend to be more similar in style to the test sentences.

4 Ethical Considerations

4.1 Impact on Originality and Creativity

If our generator becomes capable of producing coherent text, it could potentially affect the originality and creativity of authors, YouTube creators, blog posters, and others in the creative field. By mimicking their styles, the uniqueness of their work may be undermined.

4.2 Impersonation and Reputation Damage

A more advanced version of our model could be used to impersonate others by generating text in their distinct style. Such impersonation could potentially harm their reputation or lead to other negative consequences.

4.3 Bias in Professional Setting

If this model were to be used in professional settings, we would need to ensure that the PCFG captured a diverse enough lexicon, so the style of various groups’ writings could be used without a bias towards one group. A biased model might inadvertently favor certain styles, undermining the inclusivity and fairness of the generated content.

5 Conclusions

Our model does not produce comprehensible sentences. We think that the massive PCFG that we used is not good at producing comprehensible sentences and in future iterations we may need to find a different model that produces human readable

output while also being able to modify the probabilities in that model to cater towards the training data. Based on our evaluation method, we see that our model does better on the training sentences when compared to the test sentences which means that the model is able to re-weigh the grammar rules in a manner that makes the PCFG more weighted towards the style of the model.

References

A Appendix

Joshua Garcia-Kimble:

Worked about 9 hours on the project. He contributed to making the CKY Parser work for full.pcfg. He also created the algorithm to generate a parse tree from a given part of speech.

William Yang:

Worked about 9 hours on the project. He contributed to updating the grammar rules in the PCFG based on the parses of the training data. He also helped ideate the structure of the code and the different types of data structures needed.

Anxin Yi:

Worked about 9 hours on the project. She created the precision recall evaluation method. This method compared the parse tree generated from the function generateFromSeedWord to the parse trees from our training data and averaged the precision recall from each of these.