MLP G95

s1781323 Ryotaro Miura
s1412880 Konstantinos Stergiou
s1407470 David Robinson

February 2018

1 Introduction

These days, text generation with neural networks is one of the most popular fields in NLP. Recurrent neural networks (RNN), which have been found to be effective for transaction task, improve the performance[Lop15], and in 2017 the research with Generative Adversarial Network (GAN), which is one one the newest ways to generate contents, is published[WQW17][Xia17]. Our goal is to develop a more natural text generation system based on these concepts.

2 Research Question

Our research goal is to apply GAN and RNN for text generation and to measure how "natural" the generated text is by them objectively. These methods are famous for image processing. Some research applies these methods to NLP tasks. However, especially GAN, this kind of studies does not show the objective performance comparison between various techniques. Our research shows the improvements in this field, knowledge of these new models and the path to making a better system on NLP.

3 Objectives

Our objective is to measure how natural the generated text with several models. We implement Hidden Markov Model as a baseline experiment, and then, RNN and GAN. Also, we look the data with or without speech tags for the n-grams/tokens. We measure the performance with the Bilingual Evaluation Understudy (BLUE) method[Pap+02]. BLUE is widely used to evaluate the quality of text which has been machine-translated from one natural language to another.

4 Data and Task

For this study we will be using the Brown corpus as a dataset. The corpus contains 1,161,192 words from 500 documents across multiple types of sources. The corpus uses a set of 82 POS-tags fro both words and punctuation. For expermiments involving neural networks (but not for baseline experiments using HMMs) these words will be converted into a one-hot vector encoding. This report also aims look at the difference between text generation with and without POS tags so we will also be using a ”two-hot” encoding where we concatenate a standard one-ht encoding with a one-hot encoding using POS-tags rather than words to create a vector in the form shown below.

NN	DET	JJ	'the'	'fat'	'cat'
0	1	0	1	0	0

5 Baseline Experiments

All baseline methods are used to produce 5 sentences which will be reported below. BLEU scores will be added for comparison to our neural network methods outlined later in the study.

There are many existing methods for text generation without using neural networks. For baseline experiments, this report will consider two types of hidden markov model text generators. The first model uses a trigrams with an HMM. The model starts with two dummy start words and randomly chooses words based on the probability of that third word appearing based on the previous two. The model runs until it reaches a sentence ending punctuation mark. The issues with this method include it having no concept of semantics and the fact that stop words often appear twice in a row means that the words generated have no long term dependency. In addition, sentences generated this way can be susceptible to getting stuck in loops.

The second text generation method using HMMs will work like the model before but will use POS tags instead of words. Once each POS tag is generated using the POS trigram, a word will be chosen based on the probabilities of each word appearing from the tag. The results for both HMM methods can be seen in the table below.

Trigram HMM	POS-tag Trigram HMM
Each year these shows have increased steadily over the volumes up to tempt him.	Of which its conservative order conjures the rain put.
A year of over-achievement.	After by the dress of large and premarital historian with the strong reaction bending at such the self-consciousness blood in the glass, take even too, houston, simmel, of the own usage was in all their achieving functions.
High-speed buses on the way in the rise or fall short of forty lashes we finally managed to cover the imputed operating expenses both in the district of small firms have the class discussions were to be a matter of life, liberty and its civilization based on this side of the home folks again, "texture", he said he was only ten days after donation and were washed three times as many as we shall see that the church in revolutionary china, the ruins marked, the sound of miles per year per student would enable her to stop civil rights and obligations must be taken of unusual interest in the tiled kitchen.	Third a united men no., congress went an literary part and the dictionary hideaway.
The formula.	"In three song he forgot partly on them.
He is up for the very smell of the excessive heat which clamped an uneasy, how can you say" – as long as the wife is, like any other desk officer, non-com and trooper was the part of reality.	Drag of the more july and its rehearsals, and all of zhok, anne, and if the old big shows were exposed together.

6 Interim Conclusions

As can be seen, the HMM models are not very effective at generating natural sounding speech. They are reasonably good over short sentences but are limited by their lack of long-term dependency. The model not using POS tags performs better here. This is probably due to its ability to produce more natural short windows of text as every sequence of 3 words shown appears somewhere in the data as apposed to the POS model where words only depend on the tags of the previous 2 words and not on the words themselves. The advantage of the POS-tag model was intended to be its ability to create more grammatically sound sentences, but this is again hampered by its inability to use tags further than 2 words behind. This means their is no noticeable improvement in grammar over the whole sentence. In order to create genuinely natural sounding text it is clear that more sophisticated methods are needed.

7 Future Plan

Our future plans involve writing a Recursive Neural Network(RNN) and a Generative Adversarial Network (GAN) and compare their performance in text generation. Furthermore we will be looking at how using or not Part of Speech (POS) tags for the n-grams/tokens will affect performance. Our baseline as discussed earlier will be the performance in the same task from a Hidden Markov Model (HMM), whose performance is also mentioned in the according section above. We do run the risk of our findings not being conclusive in the case that the performance of the two aforementioned architectures are extremely similar. Furthermore we could be in the scenario that we do not have enough points to analyse/talk about. In the first risk case we will have to explore more cases on top of using or not POS tags or looking at how different language models affect the performance of those two network structures in order to reach a decisive conclusion. In the second risk case we will in a similar fashion as the first risk case, explore more

References

[Pap+02] Kishore Papineni et al. "BLEU: a method for automatic evaluation of machine translation". In: *Proceedings of the 40th annual meeting on association for computational linguistics*. Association for Computational Linguistics. 2002, pp. 311–318.

[Lop15] Konstantin Lopyrev. "Generating news headlines with recurrent neural networks". In: *arXiv preprint arXiv:1512.01712* (2015).

[WQW17] Heng Wang, Zengchang Qin, and Tao Wan. "Text Generation Based on Generative Adversarial Nets with Latent Variable". In: *arXiv preprint arXiv:1712.00170* (2017).

[Xia17] Xuerong Xiao. "Text Generation using Generative Adversarial Training". In: 2017.