Comparing RelGANs, LSTM networks and GPT-2 on text generation of fiction and non-fiction text data using BERTscore

Fien Ockers, *s4618262*
Radboud University, Nijmegen, The Netherlands
f.ockers@student.ru.nl

1 INTRODUCTION

Text generation is the task of generating new text in the same style as some given larger text. In recent years, there have been many developments in this field, leading to some interesting results that sometimes also reached the mainstream media. For example, when the company Botnik published the new Harry Potter chapter their model 'wrote' in 2017 [1]. Since then, more and also better models have been made that are capable of generating text that is almost indistinguishable from a text written by humans. So, it is quite interesting to see which models are currently among the best and also whether they perform well on different kinds of text. For example, there is a clear difference in writing style between fiction and non-fiction text. So besides generating text on the same topic as the given text, a model should also take the writing style into account.

To be able to compare the performance of different models on different kinds of source text, it is necessary to have some sort of evaluation metric. The default metric for this used to be the BLEU score [3]. For text generation, this metric only compared n-grams between the generated text and the source text. This is quite a restricted method and does not give a lot of freedom to the model to generate more diverse text. Hence, researchers have come up with the BERTScore [6]. Here, the contextual BERT word embeddings of the generated text and source text are compared. The difference is that this focuses more on semantic meaning than literal word use as the BLEU score does.

As there is now a presumed better method of evaluation available for text generation, it would be interesting to test some of the best performing models, on different kinds of text, and see which one performs best. The models that will be used in this research are RelGANS, LSTM networks and GPT-2 (the open-source predecessor of the recently published GPT-3).

RelGANs are inspired by the good performance of regular GANs on image generation [2]. Some adaptations are necessary to use a model made for image generation for the task of text generation. The result is that training becomes even more unstable and mode collapse is more common, leading to the model only generating simple and short sentences. Thus, an improvement has been made by adding a relational memory in the RelGAN model.

LSTMs are a kind of Recurrent Neural Networks (RNNs). RNNs are used when any kind of memory is necessary in a prediction task, which is the case for text generation [1]. However, it has been found that modelling long term memory in RNNs is difficult due to the vanishing or exploding gradient. Hence, some adaptations in the form of input, forget and output gates have been made, resulting in Long Short Term Memory (LSTM) networks.

GPT-2 stands for Generative Pretrained Transformer 2 and is a generative model based on the structure of Transformers to be able to generate text [4, 5]. Transformers are relatively complex models consisting of several components, such as multi-headed attention and feed-forward neural networks.

Regarding the source texts that will be used for this comparison research, there will be two main categories. Fiction and non-fiction. For the fiction text, the entire seven books of Harry Potter can be taken from Kaggle [2]. And for the non-fiction text, a dataset of 143k news articles can be taken from Kaggle [3].

2 PLANNING

In Table 1 the planning of the rest of this research can be found. One week extra is planned if some tasks take longer than expected.

Week	Task
48	Finish project plan
49	Write code for LSTM approach
50	Write code for RelGAN approach
51	Write code for GPT-2 approach
1	Write paper
2	Extra week if there is delay
3	Hand in paper

Table 1: Planning for finishing research paper

3 RELATED WORK

4 APPROACH

5 RESULTS AND ANALYSIS

6 DISCUSSION AND OUTLOOK

[1] https://botnik.org/content/harry-potter.html

TxMM'20-'21, Radboud University, Nijmegen, Netherlands

[2] https://www.kaggle.com/alex44jzy/harrypotter
[3] https://www.kaggle.com/snapcrack/all-the-news

REFERENCES

[1] Touseef Iqbal and Shaima Qureshi. 2020. The Survey: Text Generation Models in Deep Learning. *Journal of King Saud University-Computer and Information Sciences* (2020).

[2] Weili Nie, Nina Narodytska, and Ankit Patel. 2018. Relgan: Relational generative adversarial networks for text generation. In *International conference on learning representations*.

[3] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*. 311–318.

[4] Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language Models are Unsupervised Multitask Learners. (2019).

[5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*. 5998–6008.

[6] Tianyi Zhang*, Varsha Kishore*, Felix Wu*, Kilian Q. Weinberger, and Yoav Artzi. 2020. BERTScore: Evaluating Text Generation with BERT. In *International Conference on Learning Representations*. https://openreview.net/forum?id=SkeHuCVFDr