

Longish NLG!

Juhwan Lee
juhwan@pdx.edu

Graham Seamans
gseamans@pdx.edu

Tareq Jallad
tajallad@pdx.edu

Abstract

We're planning to use transfer learning and other techniques to try our hand at long format text generation. We're specifically interested in extending already existing stories.

1 An Existing Research Paper Summary

1.1 Bibliographical info

The research paper chosen by our group is *Pretrained Language Models for Text Generation: A Survey* (<https://arxiv.org/abs/2105.10311>). This paper was published in 2021 and is authored by Junyi Li, Tianyi Tang, Wayne Xin Zhao, and Ji-Rong Wen.

1.2 Background

Text generation has evolved into one of the most important, yet difficult, tasks in natural language processing (NLP). The resurgence of deep learning by neural generation models, particularly the paradigm of pretrained language models, has greatly advanced this field (PLMs). The paper takes on the task of being the first work presenting a comprehensive overview of major advances achieved in the topic of PLMs for text generation, and to provide text generation researchers a synthesis and pointers to related research.

1.3 Summary of contributions

The paper presents general task definitions and outlines the mainstream architecture of PLMs for text generation. The paper explains how different types of inputs, such as random noises, discrete attributes, structured data, will lend itself to different outputs like: unconditioned generation, topic to text gen, and image captioning. Each of these inputs and outputs require that the different PLM's are used. We believe that our project will likely

use unstructured data inputs like a sentence, paragraph, or document. The DiscoBERT architecture seems particularly relevant to our project, but we'll need to look into it more later.

As for architectures, the paper briefly discusses the decoder and encoder-decoder variants of transformers. The former is used by the models GPT and CTRL, and latter used by models MASS, T5 and BART. The paper shows how to adapt an existing PLM to model a variety of input data and meet special properties of the generated text, such as relevance, faithfulness, and order-preservation and then covers fine-tuning strategies. One fine-tuning strategy that this paper presents is to thoroughly train the parameters of PLMs using task-specific data, so that PLMs can capture the semantic characteristics that are unique to the generation task.

1.4 Limitations and discussion

One limitation of this paper, which it states itself, is that most PLMs are pretrained on English text. This makes the scope of the paper limited to English PLMs, rather than PLMs of all languages.

1.5 Why this paper?

Our group chose this paper because it is extremely relevant to our project and because it was published very recently.

1.6 Wider research context

This paper argues for more research into ethical concerns regarding PLM. PLM is currently pretrained on a large corpus crawled from the web without fine-grained filtering, which could raise ethical concerns such as generating private content for users. As a result, researchers must do everything possible to prevent the misuse of PLM. Furthermore, texts generated by PLM can be biased, which is consistent with bias in educational data across gender, race, and religion dimensions. As a result, intervention in PLM is required to prevent

this bias. Despite extensive research on the general approach, PLM is still in its early stages.

2 Project Description

2.1 Main Goal of the Project

The main goal of this project is to become somewhat competent with using NLG / NLP tools, and to hopefully create some readable stories in the style of famous authors. The long term goal and idea of this project is infinite story continuation. To make a never ending story.

2.2 What NLP task will you address

We want to try our hands at long format natural language generation (NLG), or natural language continuation. None of us have any experience with NLG, so we're going to have to spend a significant amount of time getting up to speed with basics, but we're confident that we'll be able to get there. More specifically, we want to try style transfer, for example Shakespeare, and in terms of evaluation we are interested with creating a cohesive narrative in a long generated text. For example, we plan to create fake Shakespeare plays using real Shakespeare plays as our training data.

2.3 What data we plan to use

Because we aim to copy a particular author's writing style, e.g. Shakespeare, we are going to use the corpora of authors as our data set. Since our training data is going to be literally single authors who we are trying to implement, we are trying to create biased models. We are biased in the authors we choose, but don't see this as an ethical concern.

2.4 What method are you planning to use

Our overall game plan is to first implement a simple NLG using hugging face and then to try transfer learning (hopefully still using hugging face). Beyond this we don't have much of a game plan. We're looking into reading papers like: Longformers, and DiscoBERT to understand how to work with longer inputs. One thing we think would be really cool would be to improve the performance of the model by using GAN. However, while GANs have been very successful in image processing, they are difficult to apply in natural language generation.

Another idea we find particularly exciting is to use a model similar to [Hierarchical Neural Story Generation](#) along with a summarizing engine to

feed into our NLG network. This would allow us to leverage the work done in text summarizing and NLG from prompts. Whether each forward pass would make a summary therefore allowing training of the summarizing engine and generation together, or whether they would be two different stacked networks is something we haven't worked out yet, although having one network does sound much cooler.

2.5 What baseline(s) will you use

We haven't been able to find any papers that we can use as a baseline that have tried to take on a similar task. We will definitely continue to look around, but it seems that most papers in this area that we have been able to find focus on creating a story from a prompt. Although this is a way that we may approach our problem, it is not the problem that we want to look into solving.

2.6 How will you evaluate your results?

We plan to evaluate our results with some automated metrics, but it seems that we will need to have some humans involved as well.

One idea that we thought may be useful for machine learned evaluation would be to use the writingprompts subreddit. If we can train a NN to guess the number of up votes from a short story (likely with an encoder only transformer NN), then maybe we can use that to evaluate our model. Alternatively we could use Goodreads, but we don't have access to the books reviewed there. Either way, if we take this approach we will be injecting bias from whatever we get our reviews (labels) and corpus.

As for human evaluation, We will probably like to try to enlist our classmates help. Because we are looking to generate long form continuation of story's our samples will be long. Because of this we'll need to spend a fair bit of time figuring out which examples to send out for evaluation.

We're reading through this monster of a paper, [Evaluation of Text Generation: A Survey](#). It has sections for human evaluation, untrained machine evaluation, and trained machine evaluation.