Natural Language Generation in Story Continuation with Transformer-based Models as a Creative Writing Tool

By Maya Epps

Computer Science

Under the mentorship of Dr. Mandy Korpusik

Abstract: Recent work in Natural Language Processing (NLP) has created machine learning models that can generate language very similarly to humans using deep learning models called Transformers. A Transformer-based model called GPT-2 may improve a creative writing tool that generates a continuation of a story for a writer. This proposal explores research to determine whether the application of GPT-2 along with suggestions from prior studies will improve coherence and creativity of suggestions in creative writing tools. To determine the success of this tool, I will collect user feedback to comparatively determine if this implementation improves upon previous systems, and where it still has room to improve.

INTRODUCTION

The Narrative or Story Continuation task asks a machine learning model, given the beginning of an (often human-written) story, to dynamically answer the question “so what happens next?” [7]. My goal is to improve upon identified weak spots within current creative writing support tools that apply this narrative continuation task. Implementing a creative writing support tool by fine-tuning GPT-2 on specific genres of books for narrative continuation should improve the tool’s coherence because GPT-2 improved upon many natural language benchmarks for language generation and context in 2019 [1]. My question is: does applying transformer-based models that are fine-tuned on specific book genres to a narrative continuation task improve coherence and creativity, compared to Recurrent Neural Networks (RNN) with Gated Recurrent Units (GRUs) baseline for a creative writing support tool, as measured by user feedback?

BACKGROUND AND RELATED WORK

The field of Artificial Intelligence (AI) has long explored ways in which computers can write their own stories. Books have been written using AI, as seen in the websites Literai and Booksby.ai. Literai describes itself as “a home for fiction written by computers” where people can post the interesting stories they generate, often using neural networks [5]. Booksby.ai advertises itself as a solution for those who are “tired of books written by authors,” where users can buy novels fully generated by artificial intelligence [2]. Although a lot of these generated stories need improvement, they demonstrate interest in the potential of story generation.

If computers can write stories on their own, can they help people write stories? Current research suggests that writers may be interested in such a tool. For instance, a case study for how writers use a “machine-in-the-loop” system, where a machine plays a supporting role to a human, that generates a suggested next sentence for a story found that participants, including authors, “could envision use cases for future systems.” Other prior research into story continuation, specifically a dissertation by Melissa Roemmele, has tried both a case-based reasoning model that finds the most similar sentence in a separate book to use as a continuation for the given story, and a neural network called a RNN with GRUs to generate unique continuations [7].

Both of these given examples received feedback that the tool had promise, but each provided ways to improve upon their work, including that the generated text often did not make sense. The paper on the case study for how creative writers use a machine-in-the-loop used a model that replaced important words in previously written “skeleton” sentences with words relevant to the story, but they found that “participants wanted more coherent suggestions from the model.” In addition, the authors suggested future work move towards a balance between generating coherent and surprising continuations in an interface with a low level of intrusiveness and a high level of control, meaning the user would decide when to receive suggestions [4]. Roemmele’s work on narrative continuation, which used an RNN with GRUs, found that “the most common piece of feedback was that the suggestions were not coherent” [7]. Both of these studies found that users were looking for more coherent suggestions from the models.

However, a new deep learning model called a Transformer now outperforms the earlier state-of-the-art RNNs using only self-attention, without any recurrence or convolutions. Transformers are deep learning models that allow more parallelization, use a mechanism called attention, and train more efficiently on more data [3]. They have recently sparked huge models trained on massive datasets that have improved the state-of-the-art on many NLP tasks very quickly. Using Transformers and attention mechanisms, recent models such as XLNet, TransformerXL, GPT, and GPT-2 became very good at language generation, even sparking fear that they are so good they would become dangerous for their ability to spark eerily human sounding language very quickly [1][6]. In particular, some are state-of-the-art models in language generation, surpassing RNNs with GRUs such as those used in the aforementioned creative writing tools [1]. In particular, Transformer models are very good at imitating writing style, which may be useful for creative writers [8]. In fact, GPT-2 is already being used as creative inspiration by some authors. Journalist and novelist Sigal Samuel believes GPT-2 can help with writer's block [10]. She adds that GPT-2 can “be an incredible tool for writers precisely because it’s great at defamiliarizing our world. The human-ish language it generates can startle us into seeing things anew,” referencing a literary theory called defamiliarization [10].

Language modeling finds the statistical probability of a sequence of words, such as a sentence. GPT-2 achieves state-of-the-art on many language modeling benchmarks and is good at incorporating context into its generations [1]. Therefore, fine-tuning GPT-2 on specific genres of books for narrative continuation may improve a creative writing support tool. As a result, I would like to explore whether applying Transformer-based models that are fine-tuned on specific book genres to a narrative continuation task improves coherence and creativity, compared to a RNN with GRU baseline for a creative writing support tool, as measured by user feedback.

METHODS

This creative writing tool will be implemented first by fine-tuning GPT-2 on a corpus of books. GPT-2 would be used in particular because it is so good at imitating writing style and it performs very well on language benchmarks. The creative writing tool would also be implemented with an RNN with GRUs for fairness of comparison between the two models. For this corpus of books, I plan to use Project Gutenberg, a free collection of 60,000 books that are not copyrighted [9]. Multiple models of GPT-2 will be fine-tuned on Google Colab’s free GPU on a variety of different genres of books, such as historical fiction, fantasy, classics, mystery, romance, science fiction, horror/thriller, action, and adventure because GPT-2 tends to perform better when prompted with topics highly represented in the data [1]. The baseline RNNs with GRUs would be trained on the same genres. Next, the creative writing tool will be created as a web app and published using Heroku. Due to the suggestions from previous studies, the models will have a low level of intrusiveness that only provides suggestions to users when prompted. In addition, to allow the user more control and support non-linear writing, lines of text may be highlighted, and the highlighted text will be the only text used for the narrative continuation instead of everything currently written.

In order to determine if this tool will be helpful for creative writing, Amazon Mechanical Turk (AMT) will be used to crowdsource 50 participants. Before publishing the AMT tasks, this study will require IRB approval as it uses human subjects, and an IRB application will be submitted before AMT tasks are to be published. In order to ensure high quality responses, only master-level writers will be allowed to see the task on AMT. Participants will be taken to the web app and asked to start or continue writing a story in both the GPT-2 and RNN tools and test the tool by generating continuations as they write. When they are finished, they will be asked a series of questions about their experience to evaluate the two implementations, including whether they found the continuations helpful in their writing, which model they preferred and why, what bothered them and how it could be improved, their creative writing experience (i.e. professional, amateur, beginner, etc.), and whether they would use the tool for their own creative writing endeavors.

EXPECTED RESULTS

The expected results are a research paper explaining the findings. This paper will describe whether people found the tool useful, how GPT-2 performed compared to the RNN with GRUs, common feedback, and how the tool could improve in the future. Based on GPT-2’s previous performance, I expect GPT-2 will improve at least the coherence of the tool, and GPT-2 will improve how helpful writers find the system. The creative writing tools (both the GPT-2 and RNN versions) created and tested in the process of this research will also be a secondary result.

The creative writing support tools will generally look like Creative Help by Melissa Roemmele [7]. In addition, it will implement suggestions from previous studies by decreasing intrusiveness, allowing generated text based only on a selected subset of the story, and giving writers control over which of the models (and what genre it was trained on) they use.

CONCLUSION

In conclusion, I believe including Transformer-based models like GPT-2 as well as implementing feedback from prior user studies will improve a narrative continuation-based creative writing tool. In order to test this, I will create the tool by fine-tuning GPT-2 (and training an RNN with GRUs for comparison) on genre-specific corpuses of books and use Amazon Mechanical Turk to crowdsource feedback, such as the creativity and coherence of the tool, as well as how it can improve in the future. I will then share my findings in a paper aimed at the student research workshop at the Annual Conference of the North American Chapter of the Association for Computational Linguistics.

REFERENCES

[1] A. Radford, J. Wu, D. Amodei, D. Amodei, J. Clark, M. Brundage, and I. Sutskever, “Better

Language Models and Their Implications,” *OpenAI*, 14-Feb-2019. [Online]. Available: <https://openai.com/blog/better-language-models/>. [Accessed: 04-Nov-2020].

[2] A. Refsgaard and M. Thybo Loose, “Tired of books written by authors? Try books written by

AI,” *Booksby.ai*. [Online]. Available: <https://booksby.ai/>. [Accessed: 04-Nov-2020].

[3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser, and I.

Polosukhin, “Attention is all you need,” in *Advances in Neural Information Processing Systems (NIPS)*, 2017, pp. 5998–6008.

[4] E. Clark, A. S. Ross, C. Tan, Y. Ji, and N. A. Smith, “Creative writing with a machine in the

loop: Case studies on slogans and stories,” in *Proceedings of IUI,* 2018.

[5] “Fiction Written by Computers,” *Literai*. [Online]. Available: <https://www.literai.com/>.

[Accessed: 04-Nov-2020].

[6] J. Howard, “Some thoughts on zero-day threats in AI, and OpenAI's GPT-2,” *Some thoughts*

*on zero-day threats in AI, and OpenAI's GPT-2 ·*, 15-Feb-2019. [Online]. Available:

<https://www.fast.ai/2019/02/15/openai-gp2/>. [Accessed: 21-Oct-2020].

[7] M. Roemmele, "NEURAL NETWORKS FOR NARRATIVE CONTINUATION," Ph.D

dissertation, University of Southern California, Los Angeles, 2018. Accessed on: November 4, 2020. [Online]. Available: <https://roemmele.github.io/publications/dissertation.pdf>

[8] N. Yakovenko, “The Best Deep Natural Language Papers You Should Read - BERT, GPT-2

and looking forward,” *Medium*, 04-Dec-2019. [Online]. Available: <https://medium.com/@Moscow25/the-best-deep-natural-language-papers-you-should-read-bert-gpt-2-and-looking-forward-1647f4438797>. [Accessed: 18-Oct-2020].

[9] *Project Gutenberg*. [Online]. Available: <https://www.gutenberg.org/>. [Accessed:

04-Nov-2020].

[10] S. Samuel, “How I'm using AI to write my next novel,” *Vox*, 30-Aug-2019. [Online].

Available:<https://www.vox.com/future-perfect/2019/8/30/20840194/ai-art-fiction-writing-language-gpt-2>. [Accessed: 11-Oct-2020].

BUDGET

I expect this work will take one semester to complete. If a semester is 16 weeks long, I estimate I will need to spend 6 hours per week for a total of 16 x 6 = 96 hours on this project. First, the IRB application will be submitted, hopefully for approval the next month. During this time, I can begin to train the GPT-2 and RNN with GRU models. I will then spend around 4 weeks creating the creative writing support tool with these models. Once the IRB application is accepted, I will be able to publish the AMT tasks (at least 10 at a time, to keep AMT fees down.) Therefore, this part of the study should take around 5 weeks to get all 50 participants (but can be compressed in time if needed). This hopefully leaves about two weeks for analysis of the results. Other materials such as the training data from Project Gutenberg are free, and Heroku services should be free for a trial period of 550 hours of use, which should be sufficient for this project. A rough estimate of what this project would cost is $300. Each of the 50 AMT participants will be paid $5 per AMT HIT. AMT also requires at least 20% per HIT in fees, which is $1 per HIT. Therefore, the Amazon Mechanical Turk portion will require 50 x (5 + 1) = $300.