Fine-Grained Controllable Text Generation Using Non-Residual Prompting

---

15 Nov 2021 (modified: 16 Nov 2021)ACL ARR 2021 November Submission

Authors:*Fredrik Carlsson, Joey Öhman, Severine Verlinden, Fangyu Liu, Magnus Sahlgren, Joakim Nivre*

TL;DR: A new attention schema for controllable text generation

Preprint: yes

Preferred Venue: ACL 2022 Generation

Consent: yes

Consent To Review: yes

Decision by Program Chairs

ACL 2022 Conference Program Chairs

23 Feb 2022ACL 2022 Conference Paper1552 DecisionReaders: Program Chairs, Paper1552

Decision: Accept to main conference

Meta Review of Paper1415 by Area Chair GHwR

ACL ARR 2021 November Paper1415 Area Chair GHwR

08 Jan 2022ACL ARR 2021 November Paper1415 Meta ReviewReaders: Paper1415 Senior Area Chairs, Paper1415 Area Chairs, Paper1415 Authors, Paper1415 Reviewers, Program Chairs

Metareview:

This paper addresses the challenging problem of NLG, i.e., fine-grained controllable text generation, by introducing an encoder-decoder architecture that enables intermediate text prompts at arbitrary time steps. Specifically, the proposed method uses an auxiliary prompt encoder to guide these language models towards certain themes, words, or even phrases by conditioning the language model not just on prior context but additionally on the output of the auxiliary model during generation. The language model does not need to be fine-tuned, and different criteria can be used at different time steps. To evaluate the proposed method, this paper presents a variation of the CommonGen dataset by introducing additional context to each of the sentences in the corpus. Experimental results prove that the introduced method can produce sound and diverse sets of sentences of a pre-specified length with a high probability of using the prompted words and performs well for longer input sentences with contexts.

Summary Of Reasons To Publish:

1. The proposed method is novel for controllable text generation, which can be widely utilized for CLMs.
2. The proposed method is computationally efficient and simple to implement.
3. The modification to the common-gen dataset makes the evaluation of open-ended generation more precise.
4. Comprehensive experiments and strong performance.
5. This paper is well-written.

Summary Of Suggested Revisions:

1. Since the proposed method requires two separate prompt model pretraining phases, the authors should compare its computational complexity with baselines.
2. This paper only tests the method on only one form of control, i.e., the satisfaction of the word inclusion constraint. More explorations about other forms of control are encouraged.
3. More ablation study is suggested to clarify the effect of each element in the proposed method.
4. Typos, missing references, and the confusing presentation should be addressed.

Overall Assessment: 4 = There are minor points that may be revised

Suggested Venues:

Main Conference

Ethical Concernes:

n/a

Official Review of Paper1415 by Reviewer zs4s

ACL ARR 2021 November Paper1415 Reviewer zs4s

28 Dec 2021 (modified: 28 Dec 2021)ACL ARR 2021 November Paper1415 Official ReviewReaders: Program Chairs, Paper1415 Senior Area Chairs, Paper1415 Area Chairs, Paper1415 Reviewers, Paper1415 Authors

Paper Summary:

This paper focuses on enabling prompt based generation with prompts being inserted randomly in the middle of a seqeunce being generated instead of just the start. This paper posits that this approach provides finer grained control than typical prompt based generation approaches for controlled generation. To ensure that the generation is not arbitrarily derailed by perturbation of hidden states at each position--an approach which is common in many controlled generation techniques--this paper proposes "non-residual" prompt based generation. Essentially, two hidden states are associated with each generation step: 1) standard autoregressive hidden state of the base LM, 2) prompt sensitive hidden state. While the prompt sensitive hidden state is used to sample/propose the next token, the hiddden state dynamics are governed by the first stream. To ensure that prompts can be inserted in the middle of sentences, two separate prompt pretraining phases are proposed. For the task of generating sequences with word inclusion constraints, these prompts are pretrained by randomly selecting target words in subsequent generation steps. The sequence invariant pretraining essentially freezes the prompt model and learns additional parameters based on slightly modified pretraining to encourage the prompt's key-value pairs to be positionally invariant. Empirical comparisons are made with relevant prior work baselines on the task of satisfying word inclusion constraints on CommonGen task. Additionally the CommmonGen dataset is extended by having humans provide three sentence context to the instances in the CommonGen dataset in order to test context sensitivity of the proposed approach.

Summary Of Strengths:
-- The method is clean and simple to implement.

-- The experimental design is reasonable with appropriate baselines and informative metrics including human evaluations on "sense" and "context awareness".

-- The method performs reasonably well and outperforms the baselines on several metrics.

-- The motivation behind extending the dataset makes sense and would be a useful addition to the datasets for such tasks.

Summary Of Weaknesses:
-- My major concern is that while the paper claims to perform controlled generation, it empirically tests the approach on only one form of control -- satisfaction of the word inclusion constraint. This is very limited, and contrary to the claim in "Discussion & future work", it is not straightforward according to me to extend this approach trivially to other forms of control. Especially, the two pretraining phases in the approach enjoy the simplicity of the training procedure by the virtue of the nature of word inclusion constraint.

-- The approach requires two separate prompt model pretraining phases. A discussion about computational complexity and comparison with baselines would be an informative addition.

-- As described below, the presentation is a little confusing and could be made better.

Comments, Suggestions And Typos:
-- Figure1: The matrix on the right is difficult to interpret. I understand the approach but still don't understand what this matrix is signifying.

-- Position invariant transformation: The text is vague and confusing. Figure 3 makes the approach clearer but its difficult to know what was done just based on the textual description. The transformation and C need to be explained in greater precise detail--current exposition feels a bit handwavy.

-- Tighter integration between the figures, captions, and the text would improve readability.

-- typo on line 238

Overall Assessment: 4 = Strong: This paper is of significant interest (for broad or narrow sub-communities), and warrants acceptance in a top-tier *ACL venue if space allows.
Confidence: 4 = Quite sure. I tried to check the important points carefully. It's unlikely, though conceivable, that I missed something that should affect my ratings.
Best Paper: No
Replicability: 4 = They could mostly reproduce the results, but there may be some variation because of sample variance or minor variations in their interpretation of the protocol or method.
Datasets: 3 = Potentially useful: Someone might find the new datasets useful for their work.
Software: 1 = No usable software released.
Author Identity Guess: 1 = I do not have even an educated guess about author identity.

Official Review of Paper1415 by Reviewer 1Uyw

ACL ARR 2021 November Paper1415 Reviewer 1Uyw

27 Dec 2021ACL ARR 2021 November Paper1415 Official ReviewReaders: Program Chairs, Paper1415 Senior Area Chairs, Paper1415 Area Chairs, Paper1415 Reviewers, Paper1415

Paper Summary:

Controlling open-ended text generation is both challenging and extremely useful for various tasks. In this work, the authors introduce a conceptually simple but effective technique for more granular control over generated text. Essentially, this involves fine-tuning a prompt encoder that takes a prompt containing decoding instructions as input. This is then combined with the CLM through non-residual attention which allows the CLM access across timesteps and helps inform the decoding process.

Summary Of Strengths:

- The paper is clear and well-written.
- The technique proposed is conceptually straightforward and effective. It is also relatively simple to apply to pretty much any CLM.
- Both the qualitative and quantitative results highlight the effectiveness of the technique.

Summary Of Weaknesses:

- It would be interesting to see what other types of controls could be prompted, but I think it's fair to leave that to future work.

Comments, Suggestions And Typos:

- I would consider https://arxiv.org/abs/2107.07150 as relevant related work

Overall Assessment: 4 = Strong: This paper is of significant interest (for broad or narrow sub-communities), and warrants acceptance in a top-tier *ACL venue if space allows.

Confidence: 4 = Quite sure. I tried to check the important points carefully. It's unlikely, though conceivable, that I missed something that should affect my ratings.

Best Paper: No

Replicability: 4 = They could mostly reproduce the results, but there may be some variation because of sample variance or minor variations in their interpretation of the protocol or method.

Datasets: 1 = No usable datasets submitted.

Software: 1 = No usable software released.

Author Identity Guess: 1 = I do not have even an educated guess about author identity.

Official Review of Paper1415 by Reviewer RpRS

ACL ARR 2021 November Paper1415 Reviewer RpRS

20 Dec 2021ACL ARR 2021 November Paper1415 Official ReviewReaders: Program Chairs, Paper1415 Senior Area Chairs, Paper1415 Area Chairs, Paper1415 Reviewers, Paper1415

Paper Summary:

This work addresses the problem of controllable text-generation from large pretrained language models. Their approach uses an auxiliary model to guide these language models towards certain themes, words, or even phrases by conditioning the language model not just on prior context but additionally on the output of the auxiliary model during generation. The language model does not need to be fine-tuned, and different criterion can be used at different time steps. The work also presents a new dataset, which adds to common-gen a set of prompts that make expected generations more specific. This is a nice addition to the currently sparse set of open-ended generation datasets.

Summary Of Strengths:

- The work presents a novel method for controllable text-generation that is widely applicable and computationally efficient.
- The modification to the common-gen dataset should make evaluation of open-ended generation more precise.
- In general, writing is clear and concise.

Summary Of Weaknesses:

My biggest problem with the paper is lack of specific applications and a more theoretical motivation of design choices. A few of these points are discussed below:

- The idea of non-residual attention is novel and interesting, but it's unclear if it's really necessary. It makes intuitive sense that using the altered hidden states could have a negative impact on generation since the language model is not fine-tuned for this case, but we see that at least using a single altered hidden state is not bad for "fluency" of the text generated by the model. Since this method requires a second pass of the input through the model, it would be nice to know whether its actually needed. The authors motivate it rather informally, but do not provide experimental evidence of its necessity.
- It is mentioned during the section on pretraining that sentence length is added as part of the objective but the precise mechanism used is not explained.
- I don't quite understand the need for relative positional encodings. The authors write in the corresponding section: "Overcoming this requires a significant amount of training of the prompt model (See Appendix C.2)" but do not explain why this is the case in either the appendix or the following paragraphs.

Further, there are moments where the authors use vague terminology, e.g., "This non-residual property of each prompt assures that the hidden state of the CLM does not deteriorate over time" in line 212 or "this detrimentally impacts the pre-trained CLM's textual abilities" in line 926. Lastly, the authors do not discuss talk additional memory/time constraints required by their method, which is an important consideration

Comments, Suggestions And Typos:
Most of my questions are mentioned in the “weaknesses” section.

- Calling perplexity a fluency metric is a bit of a stretch... Most degenerate repetitions have quite low perplexity.
- I believe the method is called Keyword2Text, not Key2Text
- Line 899: vi -> via

Overall Assessment: 4 = Strong: This paper is of significant interest (for broad or narrow sub-communities), and warrants acceptance in a top-tier *ACL venue if space allows.
Confidence: 4 = Quite sure. I tried to check the important points carefully. It's unlikely, though conceivable, that I missed something that should affect my ratings.
Best Paper: No
Replicability: 5 = They could easily reproduce the results.
Datasets: 4 = Useful: I would recommend the new datasets to other researchers or developers for their ongoing work.
Software: 1 = No usable software released.
Author Identity Guess: 1 = I do not have even an educated guess about author identity.

Official Review of Paper1415 by Reviewer NoHr

ACL ARR 2021 November Paper1415 Reviewer NoHr
16 Dec 2021ACL ARR 2021 November Paper1415 Official ReviewReaders: Program Chairs, Paper1415 Senior Area Chairs, Paper1415 Area Chairs, Paper1415 Reviewers, Paper1415
Paper Summary:
The paper presents a novel approach that allows to control text generation task using prompted words. This is done by using "non-residual attention" during generation process that allows more explicitly to attend to prompted words during the decoding phase, thus making the method more suitable for longer inputs. In order to test the method the authors also generate the extension of the CommonGen dataset by introducing additional context to each of the sentences in the corpus. Finally, the authors propose a modified training schema of the model in order to make the training faster. The experiment results show that the method produces sound and diverse set of sentences of a pre-specified length with high probability of using the prompted words, and performs well for longer input sentences with contexts.

Summary Of Strengths:

- the paper is well written and easy to follow
- the method is sound and well motivated
- authors use a variety of metrics to check the quality of the generated texts

Summary Of Weaknesses:

- it would be very useful to make the C2Gen dataset that was constructed for this paper publicly available
- how well does the model generalizes to CLM models other than GPT-2 Large? Would the non-residual attention still work well for smaller models (of course in expense of the performance)?

- would be useful to get more insights on how the elements of the proposed training schema would affect the performance and training time comparing to the conventional fine-tuning

Comments, Suggestions And Typos:
line 327 - should be a continuation of the previous sentence lien 360-361 - no commas

It would be useful to include in the paper body specific mentions of the extra results, explanation, and experiments that are included in the Appendix. Not all of them are mentioned (e.g. details on human evaluation, Residual vs Non-Residual Prompts, etc.) and I almost missed some of the important ablations and details that were described in the Appendix.

Overall Assessment: 3.5
Confidence: 3 = Pretty sure, but there's a chance I missed something. Although I have a good feel for this area in general, I did not carefully check the paper's details, e.g., the math or experimental design.
Best Paper: No
Replicability: 3 = They could reproduce the results with some difficulty. The settings of parameters are underspecified or subjectively determined, and/or the training/evaluation data are not widely available.
Datasets: 3 = Potentially useful: Someone might find the new datasets useful for their work.
Software: 1 = No usable software released.
Author Identity Guess: 1 = I do not have even an educated guess about author identity.