1 Introduction

Large language models such as GPT-3 (Brown et al., 2020), BART, (Lewis et al., 2019) etc., have advanced the state of the art in several natural language generation tasks such as text summarization (Zhang et al., 2020) and machine translation (Liu et al., 2020). However, when it comes to open-ended tasks with a focus on creativity such as generating stories (Fan et al., 2018a), poetry (Ghazvininejad et al., 2016), or various forms of figurative language (Chakrabarty et al., 2021), these state-of-the-art language models are often found to be inadequate.

The principal reason for this is that, in addition to composing grammatical and fluent sentences to articulate the intended content, these tasks usually also require extensive world and commonsense knowledge, as well as discourse modeling, to make sure the outputs maintain long-term coherence while remaining creative. It should also be noted that current approaches to text generation for specialized tasks require lots of training data for supervision. However, most existing corpora for creative forms of text are limited in size. Even if such a corpus existed for creative tasks, learning the distribution of existing data and sampling from it will unlikely lead to truly novel, creative output. Creative composition requires deviating from the norm, whereas standard generation approaches seek to mimic the norm.

This tutorial aims to bring awareness of the important and emerging research area of open-domain creative generation, with a focus on language generation while also touching on multi-modal generation (e.g., image captioning, visual metaphors, and visual story generation). It targets natural language processing (NLP) and artificial intelligence (AI) researchers as well as creative writing practitioners who are interested in building systems that are capable of emulating as well as augmenting human creativity.

In particular, we will review recent studies on creative language generation both at the sentence level as well as longer forms of text. We will provide the audiences with a holistic view of 1) the importance and challenges of building creative language generation systems; 2) methods for different forms of creative language generation such as story (Yang et al., 2022; Yao et al., 2019), poetry (Tian and Peng, 2022), humor (He et al., 2019; Mittal et al., 2022), metaphors (Chakrabarty et al., 2021; Stowe et al., 2021; Chakrabarty et al., 2020b), sarcasm (Chakrabarty et al., 2020a), and hyperbole (Tian et al., 2021) 3) how can models for creativity infer user intention and preferences, allow for fine-grained control, and take (natural language) feedback? In particular, how could the recent advancement of AI shape the future workforce for creativity? We will conclude the tutorial by outlining future research directions in this area.

2 Tutorial Outline

In this tutorial, we will review the history of creative language generation both in shorter and longer forms. Then, we will move to the recent advances in creative language generation that employ transformer-based language models as well as external world knowledge from existing resources. We will also touch upon how much creativity can we elicit from larger models like GPT3 (Brown et al., 2020) and where they are still lacking. Finally, we will discuss the real-world implications of creative language generation and how humans can interact or collaborate with these models to satisfy their specific needs. In particular, we will present recent community efforts in the following topics:

1. Reviewing the history of creative language generation and how neural methods have shown considerable improvements over prior
2. Introductions to contemporary methods for creative text generation along three main axes: content planning for long-form creative text generation, figurative language generation with commonsense knowledge and the surprise or twist factor which we term the XFactor in creative NLG.

3. Discussion on how large-pretrained language models such as GPT-3 can perform creative language generation tasks and what are some of its benefits and where we can still have targeted improvements.

4. Introduction to the challenges in evaluating creative text. What are the possible dangers of relying on crowd workers from Amazon Mechanical Turk (Karpinska et al., 2021; Clark et al., 2021)? What are the tradeoffs of using expert vs crowd worker evaluation of creativity in language generation (Chakrabarty et al., 2023a)?

5. Examining how advances in creative NLG have opened up directions of research in the co-creative domain. How do amateur and skilled writers benefit from these models? How do these models fit into existing creative writing workflows? And how does this technology need to improve to become more impactful and useful to end users?

6. Lessons learned open challenges, and discussion about how to build robust, reliable, and useful systems for creative language.

3 History

Due to the lack of vast research on creative language generation and its importance in training and testing generative models, it is necessary to have a cutting-edge tutorial on an emerging and timely topic. We are unaware of any tutorials on the exact same topic in the past 4 years’ ACL/EMNLP/EACL/NAACL conferences, with the only exception of the ACL 2020 tutorial (Mou and Vechtomova, 2020) on Stylized Text Generation: Approaches and Applications. The tutorial was mainly about style transfer. While there are some overlaps between style transfer and creative language generation, we believe our tutorial will benefit the audiences in terms of learning the vast landscape of creative language generation in the age of pre-trained language models. Finally, our tutorial will also touch upon human-AI collaboration for creativity as well as creativity for vision and language tasks which has not been touched upon in prior tutorials.

4 Prerequisite Knowledge

Our target audience is general NLP conference attendances; therefore, no specific knowledge is assumed of the audience except basic machine learning and NLP background:

- Familiar with common natural language processing concepts (e.g., word representation, syntax, semantics) as found in an introductory NLP course.
- Familiar with the problems/setups of (open-domain) generation and creative forms of text such as story, poetry, metaphors etc
- Has basic knowledge about machine learning models such as deep neural networks, classifiers, and pre-trained models such as BERT (Devlin et al., 2019), DALLE (Ramesh et al., 2021, 2022), GPT2 (Radford et al., 2019), GPT3 (Brown et al., 2020) BART (Lewis et al., 2020).

5 Tutorial Content

This tutorial presents a systematic overview of the history and the frontier of creative language generation. We will also introduce methods for sentence level and longer forms of creative language generation, and careful consideration in designing the evaluation of model outputs as well as how LLM’s can aid in providing assistance during the process of creative writing. We will then do deep dives. The detailed contents are outlined below.

5.1 Motivation, History, and Challenges [20 mins]

We will first motivate the importance of the problem by looking into works from psychology that examine what it means to be creative (Torrance, 1966) and then demonstrating practical applications of models that can produce creative outputs. Then, we will outline the challenges of building and evaluating creative generation models and systems. We will also include a brief introduction to the history of creative language generation and how
many of the challenges encountered by the community when developing contemporary language models share parallels with those faced by researchers working on these problems prior to the advent of statistical and neural techniques in NLP.

5.2 Recent Methods for Creative Generation [75 mins]

We detail various contemporary methods for creative text generation along three main axes charting progress in each. [VP: This could use a bit more punch]

Content Planning - “Austen’s Plots” [30 mins]

In this section, we will discuss how approaches to control the content of the generated text by sketching a plan (Yao et al., 2019) has enabled pre-trained language models to generate higher quality stories with coherent plot lines (Goldfarb-Tarrant et al., 2020; Rashkin et al., 2020) as well as poetry with form constraints like sonnets (Tian and Peng, 2022). We then discuss the recent phase shift to adapting this style of content planning to large language models such as GPT3 to generate even longer, yet coherent, stories (over 1000 words) via recursive prompting (Yang et al., 2022, 2023).

Figurative Language Generation with Common-sense Knowledge - “The Bard’s Metaphors” [30 minutes]

Pre-trained language models typically excel at understanding the literal meaning of the text and generating responses accordingly. However, when it comes to creative tasks, they often struggle to effectively employ figurative language, which is essential for adding depth and nuance to the text. We will discuss how incorporating commonsense knowledge from external sources (Bosse-lut et al., 2019) enables models to better generate similes and metaphors (Chakrabarty et al., 2020b; Stowe et al., 2021; Chakrabarty et al., 2021) and sarcasm (Chakrabarty et al., 2020a). Finally, we examine how chain-of-thought prompting can elicit better figurative language understanding that was learned during the pre-training of large language models resulting in opportunities to generate higher quality illustrations for the same (Chakrabarty et al., 2022b).

The X-factor - “Dickens’ Twist” [15 minutes]

Finally, there is the ineffable quality of creative writing which grips the reader to keep turning the page. While this element is most challenging to recreate from language models, we discuss works that attempt to do so by learning word-level relationships to generate puns (He et al., 2019) and break down intangible qualities such as humor into their basic principles for modeling (Tian et al., 2022).

5.3 Challenges in Evaluation of Creative NLG outputs [20 mins]

As the community makes progress in improving the various elements of the creative generation process, benchmarking progress becomes more challenging. One of the common practices in evaluating creative output is relying on crowd worker judgments from platforms such as Amazon Mechanical Turk. However, there are multiple challenges in these evaluations (Karpinska et al., 2021; Clark et al., 2021) such as crowd-workers spending limited time on reading and evaluating outputs, underspecified instructions for evaluation, variability in judgments across the same set of workers across different times, constructing proper qualification tests, setting up proper wages for crowd workers. A more promising alternative is to look into how experts might be better suited to evaluating outputs from creative NLG systems because their expectations might differ from amateur crowd workers (Clark and Smith, 2021). We discuss how recent work has delved back into the fundamentals of creativity to design evaluation axes based on the Torrance tests of creative thinking (Torrance, 1966) and measure these using expert judgments (Chakrabarty et al., 2023a).

5.4 Human AI Collaboration for Creativity [30 mins]

Recent developments in natural language generation (NLG) using large language models have brought us closer than ever to the goal of building AI-powered creative writing tools. In this section, we will discuss the potential of NLG to have a significant impact in the creative writing domain–especially with respect to brainstorming, generation of story details, and research assistance (Chakrabarty et al., 2023b). We will focus on different interaction interfaces for AI-assisted creativity, the extent to which they understand user intent, and finally, whether the human-AI collaboration improves the final creative output. We will end this section with the positives as well as limitations of current models as identified by expert andprofessional writers.
5.5 Conclusion, Future Directions, and Discussion [25 min]

We will conclude the tutorial by discussing future directions to build impactful, reliable and useful systems for creative language generation.

6 Tutorial Coverage and Suggested Reading List

While the tutorial will include our own work (Yao et al., 2019; He et al., 2019; Mittal et al., 2022; Goldfarb-Tarrant et al., 2020; Chakrabarty et al., 2020b, 2021; Akoury et al., 2020; Stowe et al., 2021; Tian et al., 2021; Tian and Peng, 2022; Padmakumar and He, 2022; Chakrabarty et al., 2022a; Yang et al., 2022), we anticipate that roughly 40% of the tutorial content will be pulled from work by other researchers in NLP and machine learning communities include but not limited to (Ghazvininejad et al., 2016; Fan et al., 2018b, 2019; Van de Cruys, 2020; Riedl and Young, 2010; Lin and Riedl, 2021; Brahman and Chaturvedi, 2020; Mirowski et al., 2023; Clark et al., 2021). A more comprehensive list of related papers will be provided before the tutorial.

7 Tutorial Instructors

Our instructors consist of experts who have conducted research in different aspects related to the tutorial topic.

Nanyun (Violet) Peng  Nanyun (Violet) Peng is an Assistant Professor in the Department of Computer Science at the University of California Los Angeles. She received her Ph.D. in Computer Science from Johns Hopkins University. Her research focuses on the generalizability of NLP technologies, with applications to creative language generation, low-resource information extraction, and zero-shot cross-lingual transfer. Her works have won the Outstanding Paper Award at NAACL 2022, the Best Paper Award at AAAI 2022 Deep Learning on Graphs workshop, and have been featured an IJCAI 2022 early career spotlight. She has given a tutorial at NAACL 2018 on information extraction.

Tuhin Chakrabarty  Tuhin Chakrabarty is a Ph.D. candidate in Computer Science at Columbia University and a part of the Natural Language Processing group, where he is advised by Smaranda Muresan. His research is supported by the Columbia Center of Artificial Intelligence & Technology (CAIT) and Amazon Science Ph.D. Fellowship. He was also a fellow at The New York Times R&D team working on Natural Language Generation. His overarching research question centers around how we can use large language models for creativity. He has published several papers in various NLP conferences and journals including ACL, NAACL, TACL and EMNLP.

He He  He is an Assistant Professor of Computer Science and the Center for Data Science at New York University. She is affiliated with the CILVR Lab, the Machine Learning for Language Group, and the Alignment Research Group. Her research focuses on building intelligent systems that can communicate with humans effectively and enable individuals to achieve their goals. Today’s systems are often opaque, brittle, and difficult to control, which limits their usefulness in human-centered applications. To make them our trustworthy collaborators, her research aims to (i) understand the computational foundation of generalization in novel scenarios, and (ii) build interactive systems that align with users’ goals. She has given a tutorial at EMNLP 2021 on robustness and adversarial examples in NLP.

Vishakh Padmakumar  Vishakh Padmakumar is a Ph.D. student in Data Science at New York University advised by He He. His research is broadly in the field of natural language processing and human-AI collaboration with a focus on collaborative text generation for creative writing tasks and other interactive settings. Prior to this, he was a Graduate Research Associate at the NYU Center for Social Media and Politics working on political stance classification and multimodal content sharing in online disinformation campaigns. He has published papers at several NLP and machine learning venues including ACL, EMNLP, and ICML and was the chair of the ACL 2023 Student Research Workshop.

References


Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. 2019. COMET: Commonsense transformers for automatic knowledge graph construction. In Proceedings of the 57th Annual Meeting of the Association for...


