Imitations of Immortality: learning from human imitative examples in transformer poetry generation

RAY LC
School of Creative Media
City University of Hong Kong
http://www.raylc.org/imitations/

To me the meanest flower that blows can give
Thoughts that do often lie too deep for tears.
- from *Intimations of Immortality*, by William Wordsworth

Abstract

Learning to generate poetry in the style of the poet can make models style experts, but humans who create imitative works take a more general approach that incorporates domain knowledge outside the poet's particular style. Instead of learning from a large corpus of one poet's works, can a machine imitate the deep structure of her style using only one example of her work? To explore generating poetic variations, I wrote 8 poems that imitated the structure and language of 8 diverse poets, and used them to fine tune a transformer model that has seen only one poem by each of the authors. The poetic art works presented show structures borrowing from the human imitation in addition to prompted content of the original work, suggesting the model has learned some aspects of how humans write variations on content by imitating style.

Introduction

Machines can learn to imitate a style by a corpus of examples, but frequently great works of literature have only one examplar. If we want to train a machine to learn to produce Hamlet's soliloquy with the same level of nuance that Shakespeare intended, one approach is to ask humans to produce poetic variations and use them to fine tune the model.

Transformer models like GPT-2 have been found to generate poetry indistinguishable from human-written poetry, given enough training examples [3]. Learning to generate particular styles, however, is more subtle, as models often generate free-form streams that don't adhere to poetic forms. Recent approaches to generating based on poetic forms include imposing strict rules on rhyming [4] and using characters in predefined positions [5]. These approaches don't take into account styles of particular poets, which may impose additional structural limits. Work in this area includes modeling poet style explicitly during line-by-line generation [6], and generating poems based on emotions evoked by particular words associated with particular poets [1]. Having such style-specific models allows us to transfer the style in text to other content [2], creating scenarios such as "how would Virginia Woolf write villanelles about loss?"

Methodology

I chose 8 poems by 8 different authors that represent a cross section of styles ranging from free verse, villanelle, quatrains, haiku, blank verse, rhymed and unrhymed. I composed 8 poems on my own that utilize aspects of each of the originals in terms of form but narrated my own content. For example, here's the Elizabeth Bishop poem and my own side-by-side:

I lost two cities, lovely ones. And, vaster,
some realms I owned, two rivers, a continent.
I miss them, but it wasn't a disaster.

—Even losing you (the joking voice, a gesture
I love) I shan't have lied. It's evident
the art of losing's not too hard to master
though it may look like (Write it!) like disaster.
from *One Art*, by Elizabeth Bishop

I blank out on my age, name, and every syllable
I uttered (Emphatic!) to make my existence remain.
It moved on before me, but even that's forgivable.

But when you (whose voice moves me, face bespectacle)
left me my memories, I give but one refrain:
the science of forgetting is not inscrutable
but being forgotten, that is unforgivable.
from *The Science*, by RAY LC

Note that I imitated not only the villanelle structure, but also the nuanced structures such as using parentheses for a call to action and two evokative images, as well as not simply repeating the ending for each of the last line of the stanzas but varying them in turn, as Bishop did. I believe these are creative inputs that will be difficult for models to pick up using one-shot learning.

Next I fine-tuned the GPT-2 355M and 124M models with each of the 8 original poems only (not putting in my own poems) for 5000 epochs (5480 tokens, average loss 0.01-0.02). Then I prompted these models with the beginning (first stanza or equivalent) of each of the 8 original poems at a range of temperatures from 0.8 to 1.8 to see how the models created new content based on the prefixes. I noticed quite a bit of overfitting, as many runs simply repeated the entire poem verbatim given the first stanza. I failed to find many deviations for the Wallace Steven's poem *Thirteen Ways of Looking at a Blackbird*, possibly because the form there is strictly bound by the roman numerals. I also saw no noticeable difference between 355M and 124M, and decided to work with the 124M model for its smaller number of parameters thereafter. When the models do go off the script (especially at higher temperatures), it was hard to see any consistency in the generated forms, but content was generalized:

I
Among twenty snowy mountains,
The only moving thing
Was the eye of the blackbird.
II
I was of three minds,
Like a tree
In which there are three blackbirds.
from *Thirteen Ways of Looking at a Blackbird*, by Wallace Stevens

I
Among twenty snowy mountains,
The only moving thing
Was the eye of the blackbird.
Q. How did you the blackbird see?
A. I saw behind the blackbird.
Q. But the blackbird cannot fly.
Do you not see the blackbird flying?
from GPT-2 pretrained to 8 original poems by 8 different poets

Thus I decided to pre-train the 124M model to the 8 original poems in addition to each of my poems written in variation with the original (5000 epochs, average loss 0.01-0.02). The idea is to provide an example of how variations could respect the original poem and push the model towards learning the way humans write variational forms.

Outcomes

`http://www.raylc.org/imitations/` compiles the poems generated in the latter manner, in an interactive format. While the results still need to be curated by the writer, I found much more success with poems that have to fit certain forms, perhaps due to the additional pre-training that constrains the model a bit more towards variations that fit the same style as the original. For example:

Good men, the last wave by, crying how bright
Their frail deeds might have danced in a green bay,
Rage, rage against the dying of the light.

Wild men who caught and sang the sun in flight,
And learn, too late, they grieved it on its way,
Do not go gentle into that good night.
from *Do Not Go Gentle into That Good Night*, by Dylan Thomas

Good men, too, were wilt thou find it so?
Lose what is lost, and reap what is lost?
Tread thorns to die and live whence they came.

That awful knowledge it brings or cares,
Opportunities, vain imaginings, but find believe by heart's desire,
Do not go gentle into that good night.
from GPT-2 pretrained *Abscond the Fabled Glorious Primrose Way*, by RAY LC

Note that it incorporated the content in the "thread thorns" line, but still followed the villanelle form mostly. These examples abound in GPT-2-generated content from Allen Ginsberg's "Howl." For example, I used the word "it" as paragraph starters in my variation "Sex", instead of Ginsberg's "who." The "Sex"-pretrained generated poetry similarly managed to switch to "while" in the middle for four lines. It also picked up on the use of dashes to join words that I used in "Sex" on line 20. "Imitations of Immortality" is a creative study on how one-shot machine learning models can begin to learn to produce variations of canonical styles of poetry by pre-training to human-written variants.

Ethical Impact

This work illustrated generation of works from classic examples, but what if the examples are new work by other artists? This creates a dilemma where machines are used to copy the styles of existing emerging artists. When I generated text pretrained from the Allen Ginsberg poem "Howl" and my poem "Sex" and showed it to others, they were not able to distinguish between "Howl" and the generated variant. Due to its length, I myself can be unsure of where the text comes from because there's so much already randomized language in "Howl." This will become even more problematic if artists begin producing works of increasingly random forms in regards to their grammar and style. To constrain these machine practices, standards will be necessary to clearly mark out works generated by machines or by humans so that intellectual property can be preserved. One way would be to attribute the user of the machine generation procedure in our citation.

Supplementary Materials

The original poems and the text generated are at: `http://www.raylc.org/imitations/`
More information about the work and its artist at: `http://www.raylc.org`

References

[1] Brendan Bena and Jugal Kalita. Introducing Aspects of Creativity in Automatic Poetry Generation. *arXiv:2002.02511 [cs]*, February 2020. arXiv: 2002.02511.

[2] Zhenxin Fu, Xiaoye Tan, Nanyun Peng, Dongyan Zhao, and Rui Yan. Style Transfer in Text: Exploration and Evaluation. *arXiv:1711.06861 [cs]*, November 2017. arXiv: 1711.06861.

[3] Nils Kobis and Luca Mossink. Artificial Intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. Paper, arXiv.org, September 2020.

[4] Piji Li, Haisong Zhang, Xiaojiang Liu, and Shuming Shi. Rigid Formats Controlled Text Generation. *arXiv:2004.08022 [cs]*, April 2020. arXiv: 2004.08022.

[5] Yi Liao, Yasheng Wang, Qun Liu, and Xin Jiang. GPT-based Generation for Classical Chinese Poetry. *arXiv:1907.00151 [cs]*, September 2019. arXiv: 1907.00151.

[6] Jia Wei, Qiang Zhou, and Yici Cai. Poet-based Poetry Generation: Controlling Personal Style with Recurrent Neural Networks. In *2018 International Conference on Computing, Networking and Communications (ICNC)*, pages 156–160, March 2018.