A language generation system that can compose creative poetry

February 26 2020, by Ingrid Fadelli

Brendan Bena presenting his work at a conference. Credit: UC Colorado Springs.

Over the past few decades, researchers have developed increasingly advanced artificial intelligence (AI) tools and computational techniques that can be applied in a variety of settings. Among these, techniques that

can generate written or spoken language have attracted considerable attention, particularly with the introduction of new voice assistants, robots and new interactive devices.

Researchers at the University of Colorado (UC)- Colorado Springs and Drury University have recently developed a unique language generation system that can produce creative poetry verses. Their system, presented in a paper [pre-published on arXiv](#), is a fine-tuned adaptation of GPT-2, a pre-trained language model developed by OpenAI.

Jugal Kalita, the professor at UC Colorado Springs supervising the recent study, has been conducting research into natural language generation for the past 30 years, starting from his graduate days at the University of Pennsylvania. His first paper on natural language generation, published in 1988, was aimed at producing paragraphs of text that might appear in a typical journal, following a basic set of rules. More recently, inspired by advancements in [artificial neural networks](#) for natural language processing (NLP), Prof. Kalita and his students started developing deep learning techniques for the generation of short articles, dialogs and creative writing.

"The idea of investigating the topic of automatic poetry generation came about at the beginning of the summer of 2019, when Brendan Bena, a summer research intern at the University of Colorado, Colorado Springs, from Missouri's Drury University, showed interest in automatically generating song lyrics," Prof. Kalita told TechXplore. "He originally wanted to look at creating a system that would attempt to mimic the emotions elicited through song lyrics."

As most song lyrics are protected by copyright, finding large datasets to train deep learning models on lyric generation can be very challenging. Bena and Prof. Kalita thus decided to develop a deep learning tool for poetry generation instead. Yet rather than focusing on features such as

the structure or rhythm of poetry, like most previous poetry generation studies, they explored its more emotional and creative aspects.

Joy Model Poem	Sadness Model Poem
Heard I a song of joy,	A long trail of falling mist
A song of happy sound,	Had made its way here, and now
Fills all the air I breathe,	Aerily it seemed, as if to drown
To him I sing, to him	The discordant thunder clang.
I sing the happy song.	It seemed to drown the music of the rain;
All night long on the steep green grass	In this lost place of sorrow
I ride and sing	Far off

Examples of the poetry evoking emotions produced by the researchers' language generation system. Credit: Bena & Kalita.

"After realizing there was a much larger portion of research, as well as data, in the field of poetry generation, we shifted our focus to this particular topic," Bena told TechXplore. "The work was largely based on the overarching task of text generation that came with much previous research. However, unlike previous efforts, we wished to focus more on the content, emotion and creativity of the text, as opposed to the structure or rhythm found in prior poetry generation studies."

To develop their poetry generation system, Bena and Prof. Kalita first gathered a large corpus of text from the Project Gutenberg and UC-Santa Cruz Dreambank databases. They browsed through the Gutenberg database looking for words included in EmoLex, an emotion-lexicon dataset developed by the National Research Council of Canada.

The researchers then split the resulting dataset into different 'emotion categories," looking at the number of EmoLex words contained in each extract, and used this data to the train a deep neural network. The model they trained is an adaptation of GPT-2, an architecture that learns to generate new fragments of text by modeling the style of language used in the data its trained on.

"We also fed our artificial neural network a combination of dream data and poetry to create what is known as 'dream poetry,'" Bena explained. "In the end, we had five separate emotion models for the emotions of joy, sadness, trust, anger and anticipation, but we also had a dream poetry model. This system, as stated previously, focuses less on the structure found in a lot of poetry generation work and more on a free-verse style of poetry that looks to imitate and reproduce the finesse and creativity of real poets."

The researchers asked human users to evaluate the poems created by their system, while also employing the Coh-Metrix tool to assess the quality of the verses it generated. They found that it produced poems that effectively elicited sadness and joy 87.5% and 85% of the time, respectively. In addition, when trained on both dream data and poetry, their system generated unique 'dreamlike' poetry verses that captured elements of what is known as 'dream poetry' with a score of 3.2 on the Likert scale.

Dream Poem 1

I am in the air,
A little bird in my cradle
Comes to me and says to me,
"Oh, what a beautiful little bird
Is this that I have made,
I am a little bird in my cradle."
I know the bird; I have known it;
I have seen it, when I can't see it;

Dream Poem 2

And as at dawn the sun was sinking down,
I thought of the evening's long-sought bliss.
And with a thrill of pride in my heart,
I looked upon the world once more.
A thousand thousand hearts I felt,
And thoughts of all my past were o'er.

Examples of dream poetry produced by the researchers' language generation system. Credit: Bena & Kalita.

"Our findings suggest that text can, in fact, be generated so that it elicits emotion in readers and that it can resemble the types of creativity that artists look to inject in their work," Bena said. "We believe our research to be a novel work in the field of creative poetry generation and hope that our study will open the door to future work in this area."

Bena and Prof. Kalita are among the first to demonstrate initial glimmers of machine creativity in poetry generation. In their next studies, the researchers plan to improve the quality of the poems composed by their system, while also applying their approach to the creation of poetry in other languages.

"If we curate the training data a bit more, we believe that a neural network architecture could better capture the emotions and dream-like aspects of the poetry we seek to create," Bena said. "In fact, while the EmoLex dictionary is a very useful dataset, its vocabulary does not

account for all the older-style English found in some of the Gutenberg poetry."

In the future, the researchers hope to replicate their experiment focusing on phrase or segment-level lexicons, as this could allow them to capture dependencies in emotion-based text more effectively. Their study could also be repeated using a more sophisticated neural network-based architecture, which may improve the quality of the poetry produced both in terms of grammar and sentence structure.

As Bena and Prof. Kalita have already used their system to produce dream poetry verses, they could eventually apply it to other creative styles as well, such as erasure poetry. Erasure poetry is produced by taking specific or random words from an existing text and then using them to form new verses.

"Finally, we are also working on generating poetry in various different languages using transfer learning," Prof. Kalita said. "For instance, Shaun Tucker, a Masters student at UC-Colorado Springs has been generating poetry in a number of Indo-European languages using OpenAI's pre-trained GPT-2 model. So far, we have generated poems in English, Spanish, Ukrainian, Hindi, Bengali and Assamese and found that the deep learning generative model GPT-2, which has been pre-trained with a large body of English text, can be trained with prose and poems in all these languages to generate poetry."

  More information: Introducing aspects of creativity in automatic poetry generation. arXiv:2002.02511 [cs.CL]. arxiv.org/abs/2002.02511

Citation: A language generation system that can compose creative poetry (2020, February 26)