arXiv:1811.05067v1 [cs.AI] 13 Nov 2018

Shall I Compare Thee to a Machine-Written Sonnet? An Approach to Algorithmic Sonnet Generation

John Benhart[1]†, **Tianlin Duan**[2]†, **Peter Hase**[3]†, **Liuyi Zhu**[4]†, **Cynthia Rudin**[5] *
Department of Computer Science[1,4,5]
Department of Statistics[2,3,5]
Duke University, Durham, NC 27708 USA
{john.benhart,tianlin.duan,peter.hase,liuyi.zhu,cynthia.rudin}@duke.edu

Abstract

We provide code that produces beautiful poetry. Our sonnet-generation algorithm includes several novel elements that improve over the state-of-the-art, leading to rhythmic and inspiring poems. The work discussed here is the winner of the 2018 PoetiX Literary Turing Test Award for computer-generated poetry.

1 Introduction

Automated poetry generation is an extremely difficult task; currently no algorithm is able to understand the complexity of language well enough to generate truly meaningful poetry with all the nuances expected from a human poet. This means that designing a poetry generation algorithm is the task of a *centaur*, a human-machine combination that can do better than a machine alone. The human needs to limit the freedom of the algorithm in the right ways, without imposing exceeding restrictions. Too many restrictions lead to boring, news-like poetry, and too much freedom makes the poems appear meaningless. When the correct balance is achieved, then the generated poems are often surprisingly beautiful.

Two poem subsections and a whole poem generated by our algorithm are presented below, where we assigned the poems' names, based on our interpretation of their meaning.

The man whose name I do not know

The earth and of a woman to lament,
I see that you are in the eastern game
To me the odor of his hat and scent,
And if it shall be dear to me his name.
And now I say it is without surprise
It is the odor of his neck and eyes.

The wife of Paul

Its cold for such an hour of protect
To tell them Paul was glad it came his wife
Within a season, or just the subject
It always seems to me for lack of life.
I showed with finger rings among the door
In bits of flight that brought him to the floor.

Whoever you are, at the end of the land

To show that it is any more than all,
And that the land is to ability
Than all that stretches, it begins to crawl
Of all I say it shall be dear to me.
Are not the song of what it is in doubt,
And I will feel the rush of heaven lined
And lessons of the sun and nights throughout
That it cannot be fittest for his mind.
The man that sprang it forth to glorify
I see the secret of the sun and light
I see just as good in the midst of my,
I see the workmen of the day and night.
Whoever you are in the eastern game,
And that it is the odor of his name.

*† denotes equal contribution. Thanks to other members of the Duke Data Science Team.

Preprint. Work in progress.

2 Description of novel poetic elements

There are several previously unstudied aspects of sonnet-generation that we have found to be important, constituting the novel aspects of this work. Our work builds on a long line of prior work on automated-poetry generation [1] [2] [3] [4] [5], where the basic code structures and methods had been developed, including the approach of picking the rhyming words first then generating each line backwards using a language model. Here we focus on aspects which distinguish our approach from other works:

Adding In-line Punctuation: We provide a novel punctuation selection method for adding commas. After all of the words for the poem have been placed, our algorithm randomly picks a number n of commas to add, and the language model is used to evaluate the n positions for commas that result in the highest poem likelihood. Then with a (false) assumption of independence between comma locations, we place commas in these n positions; luckily the mechanism rarely places commas too close together. Here is an example.

> Just make it back to me, and I can feel
> A bit of weight, it must be very, nice
> Him back into the sheath and try to steal,
> Reminding me to help him up a price.

Punctuation appears within the lines of the poems, rather than just at the end. This often adds a nice rhythm to the poem, as in the first line of the quatrain above, though sometimes within our current implementation, the punctuation appears in the wrong place, as in the second line.

Part-of-Speech Restrictions: We observed that computer-generated poems had consistent part-of-speech (PoS) errors that could be easily detected by a human. Using Python NLTK package's tags, we collected many PoS sequences that could not occur, then used this knowledge to ensure that word sequences did not violate these basic rules. For example, a pronoun typically does not directly precede a pronoun in English, meaning "he it" would not occur in our sonnets.

Dynamically Trainable Word Embeddings: In the early stages of training our language model, we fix the word embeddings which serve as the model input, then in later stages we let them be trainable. This allowed the language model to learn some grammar before eventually adjusting its word representations to suit the training corpus.

Discouraging Repetitiveness: We down-weight likelihoods for sequences that contain repeated words, based on the proximity between repeats. During line generation, once a word is placed, future sequences containing the same word are down-weighted. Moreover, a counter is used to track how many words separate the initial occurrence of the word and subsequent occurrences on a given line. The more words between the occurrences, the less the score is down-weighted. A related approach appears in [2], but their paper does not describe the mechanism in detail.

More Appropriate Corpora: Relative to past work, which used a training corpus of contemporary song lyrics [1] [2] [4] [6] , we rely on corpora more suited to sonnet generation, namely *Endymion* by John Keats, collected works of Walt Whitman, and the *Hunger Games* trilogy by Suzanne Collins. We found that, compared to using a combination of texts from a wide variety of authors, using works from a single author provided a more coherent voice to our generated poems. Meanwhile, the texts have their individual fits to the meter and cadence which sonnets require. *Endymion* is written in iambic pentameter, the meter used in sonnets, while Walt Whitman's style is more traditionally poetic than the song lyric corpus.

Narrative Quality: We observe that our poems have a narrative quality that is not matched by other poetry generation methods. Our algorithm frequently utilizes pronouns and other features that add a narrative element to the poem, as is exemplified in *The wife of Paul* above. Our approach sometimes uses the same word multiple times in novel ways. For example, in *The man whose name I do not know*, "odor" is used twice, referring to different things in each situation.

Our algorithm could lift phrases directly from the training corpus, but we never observed entire lines to be lifted. Interestingly, some phrases are frequently generated in poems on a certain topic, even when they never appear exactly within the training text.

Several additional poems are in the supplement to this paper. Our code is available at: https://github.com/peterbhase/poetry-generation

References

[1] Ghazvininejad, M., Shi, X., Choi, Y., & Knight, K. (2016). Generating Topical Poetry. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 1183-1191).

[2] Ghazvininejad, M., Shi, X., Priyadarshi, J., & Knight, K. (2017). Hafez: An Interactive Poetry Generation System. *Proceedings of ACL 2017, System Demonstrations*, 43-48

[3] Tobing, B. C. L., & Manurung, R. (2015). A chart generation system for topical metrical poetry. In *ICCC* (pp. 308-314).

[4] Hopkins, J., & Kiela, D. (2017). Automatically generating rhythmic verse with neural networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers) (Vol. 1, pp. 168-178).

[5] Oliveira, H. G. (2017). A survey on intelligent poetry generation: Languages, features, techniques, reutilisation and evaluation. In *Proceedings of the 10th International Conference on Natural Language Generation* (pp. 11-20).

[6] Wang, Z., He, W., Wu, H., Wu, H., Li, W., Wang, H., & Chen, E. (2016). Chinese poetry generation with planning based neural network. *arXiv preprint arXiv:1610.09889*.

[7] Yan, R., Jiang, H., Lapata, M., Lin, S. D., Lv, X., & Li, X. (2013, August). i, Poet: Automatic Chinese Poetry Composition through a Generative Summarization Framework under Constrained Optimization. In *IJCAI* (pp. 2197-2203).

[8] Yi, X., Li, R., & Sun, M. (2017). Generating Chinese classical poems with rnn encoder-decoder. In *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data* (pp. 211-223). Springer, Cham.

[9] He, J., Zhou, M., & Jiang, L. (2012, July). Generating Chinese Classical Poems with Statistical Machine Translation Models. In *AAAI*.

[10] Yi, X., Sun, M., Li, R., & Yang, Z. (2018). Chinese Poetry Generation with a Working Memory Model. *arXiv preprint arXiv:1809.04306*.

[11] Wang, K., Tian, J., Gao, R., & Yao, C. (2018, May). The machine poetry generator imitating Du Fu's styles. In *2018 International Conference on Artificial Intelligence and Big Data (ICAIBD)* (pp. 261-265). IEEE.

[12] Zhang, X., & Lapata, M. (2014). Chinese poetry generation with recurrent neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 670-680).

[13] Zhou, M., Jiang, L., & He, J. (2009). Generating Chinese couplets and quatrain using a statistical approach. In *Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation, Volume 1* (Vol. 1).

[14] Cheng, W. F., Wu, C. C., Song, R., Fu, J., Xie, X., & Nie, J. Y. (2018). Image Inspired Poetry Generation in XiaoIce. *arXiv preprint arXiv:1808.03090*.

[15] Liu, B., Fu, J., Kato, M. P., & Yoshikawa, M. (2018). Beyond Narrative Description: Generating Poetry from Images by Multi-Adversarial Training. *arXiv preprint arXiv:1804.08473*.

[16] Loller-Andersen, M., & Gambäck, B. Deep Learning-based Poetry Generation Given Visual Input.

[17] Díaz-Agudo, B., Gervás, P., & González-Calero, P. A. (2002, September). Poetry generation in COLIBRI. In *European Conference on Case-Based Reasoning* (pp. 73-87). Springer, Berlin, Heidelberg.

[18] Gervás, P. (2001). An expert system for the composition of formal spanish poetry. In *Applications and Innovations in Intelligent Systems VIII* (pp. 19-32). Springer, London.

[19] Greene, E., Bodrumlu, T., & Knight, K. (2010, October). Automatic analysis of rhythmic poetry with applications to generation and translation. In *Proceedings of the 2010 conference on empirical methods in natural language processing* (pp. 524-533). Association for Computational Linguistics.

[20] Manurung, H. (2004). An evolutionary algorithm approach to poetry generation.

[21] Oliveira, H. G. (2012). PoeTryMe: a versatile platform for poetry generation. *Computational Creativity, Concept Invention, and General Intelligence, 1,* 21.

[22] Wong, M. T., Chun, A. H. W., Li, Q., Chen, S. Y., & Xu, A. (2008, April). Automatic haiku generation using VSM. In *WSEAS International Conference. Proceedings. Mathematics and Computers in Science and Engineering* (No. 7). World Scientific and Engineering Academy and Society.

Supplement to “Shall I Compare Thee to a Machine-Written Sonnet? An Approach to Algorithmic Sonnet Generation"

Love and pain

The neighbor of the sun and stars above
And countless fishes swimming in the rain
Is rumble lightly on the breasts of love
With you, and I will make the song of pain.

The telescope

And sing the telescope for what, it seems
The praise of speech and day is it correct
That moves in darkness as it seems to teams
With which it only needs that they reflect.
Is not a stretcher, but I shall be strand
To answer he does it is to contain
Has no control of what to understand
I had the praise of verse, and our twain.
I had the passion of a good divine
A knoll against the gate for sort of sight
Began to blossom at the time of wine,
And, looking through a knoll against the night.
The woods and weed that brought them on this day
I was in my proposal is it play.

The pass through the mountains

To step away, and lands in front of me
The tunnel, and I slow into a smile,
And when, I see that possibility
A moment of support it takes a while.
A sort of lyric is the little hand
To try if I can let it understand.

Should I stay or should I go?

Been ready for support it takes a few
I know it’s time for me to spend the night
If he was so extreme it might be true,
It will be harsh because his voice is right.

Apollo

Of each, and now, it seems to me that feels
Whoever you are ever clarified
Him in his lasso on his neck and heels,
It shall be full of gold and terrified.
Are not the tale of ashes and, consort,
And that, it seems to me, and understand,
Untied and dews it shall be any sort
The carriage of his neck and face to hand.

The very, very heavy boulder

The upper boulders in the woods, and flee
Just as it stands between the woods, and bay,
And in a stir and quiver like the sea
Alone to bear a slope, and only gay.
Began to stop it rests with me and freeze
Was fond of my proposal is it cried
In sheets, and sweeping round it with the seas,
And when it stands beside the open side.
The way, to hold it over, and ahead
About the air in great and lack of mind
A morning from the open door with red
I had the good of them without, and find.
The kitchen to his knee, and joined the place
The huckleberry rolled him on his face.

Reflections in the water

To the reflection of the western sea
Of other men and women yearning, for,
I see the splendor of the day, and three,
Or in a bush to crawl along the shore.
And if it shall be dear to me the bees,
And vainly try to hold it at command
The secret of the earth, and of the seas
The forests of the east and through the sand.
To me, I see the common day and night
A dreamer that it shall be in its way
And from the breast of feathers on his right
In you whoever you are soon to stay.
Of each and all it seems to me the same,
And that it is the glory of his name.