

BACON: Deep-Learning Powered AI for Poetry Generation with Author Linguistic Style Transfer

Alejandro Rodriguez
Campolindo High School
300 Moraga Rd, Moraga, CA 94556
alejandro.rodriguez19@auhsdschools.org

Abstract

This paper describes BACON¹, a basic prototype of an automatic poetry generator that writes original, meaningful poetry in the style of any given author, with a quality and resemblance high enough as to be indistinguishable from the existing works of such author. It implements (a basic version of) what is known as linguistic style transfer.

The application of artificial neural networks to the problem of style transfer has been implemented successfully for paintings by Gatys, Ecker, and Bethge (2015). Linguistic style transfer for prose works has been recently explored for example in Fidler and Goldberg (2017), Xu et al. (2012).

The prototype approaches the problem of automatic poetry generation with linguistic style transfer by splitting the solution in two components: (1) a linguistic style modeler (LSM), which builds a probabilistic model of the style used by any given author, and, (2) a deep-learning powered automatic poem generator (APG) which uses the model generated by the LSM to guide the generation of original, meaningful poetry, with rich aesthetic rules, in the style of such author.

For the APG module BACON builds on the research developed by Hopkins and Kiela (2017). Their approach addresses the problem of automatically creating correct and meaningful poetry as “a constraint satisfaction problem imposed on the output of a generative model, where the constraints to restrict the types of generated poetry can be modified at will.” Their solution combines, in a pipeline, the following two components: (1) a generative language model representing content, which is implemented through a Long Short Term Memory (LSTM) Recurrent Neural Network (RNN), and (2) a discriminative model representing

form, which is implemented through a Weighted Finite State Transducer (WFST).

The LSM module generates a probabilistic model of a given author’s linguistic style by extracting high-entropy n-grams -through Term-Frequency-Inverse Document Frequency (TF-IDF)- and latent topics -through Latent Dirichlet Analysis (LDA)- from the author corpus, which is parsed against two Vector Space Models (VSM): (1) a large set of English poetry texts, consisting of 7.6 million words and 34.4 million characters, taken from 20th century poetry books, and (2) a large set of general English language texts, consisting of a full English Wikipedia dump, consisting of 5.5 million documents and 43.6 million pages.

Linguistic style transfer is achieved by probabilistic conditioning/boosting of the high-entropy n-grams and topic words in the LSM applied to the APG module.

An extrinsic evaluation procedure was performed by conducting an indistinguishability study with a selection of poems written by a human poet and automatically generated poems in the style of that same author - a variant of the so-called Turing test for art works. The results show that participants were unable to tell the difference between human and BACON-generated poems in any statistically significant way.

References

- Fidler, J., Goldberg, Y. (2017). Controlling linguistic style aspects in neural language generation. *ArXiv Preprint ArXiv:1707.02633*.
- Gatys, L. A., Ecker, A. S., Bethge, M. (2015). A Neural Algorithm of Artistic Style. *ArXiv:1508.06576 [Cs,q-Bio]*.
- 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* (Vol. 1, pp. 168–178).
- Xu, W., Ritter, A., Dolan, B., Grishman, R., Cherry, C. (2012). Paraphrasing for style. *Proceedings of COLING 2012*, 2899–2914.

¹ BACON stands for **B**asic **A**I for **C**ollaborative **p**Oetry **w**riti**N**g. The name is coined after Sir Francis Bacon who, according to some, was who actually wrote William Shakespeare’s plays.