

Crowdsourced Social Media Poetry

Carolyn Elizabeth Lamb

David R. Cheriton School of Computer Science, University of Waterloo
200 University Avenue West
Waterloo, Ontario, Canada N2L 3G1

Abstract

We describe our program of PhD work in which a computer program creates topical poems out of text found on Twitter. These poems are made using a combination of natural language processing and crowdsourcing and are part of a general research plan involving the creation and evaluation of computer-generated poetry, grounded in domain-specific research on human creativity.

Found Poetry: Background and Motivation

My research is in the area of computational creativity—the discipline of creating computer systems that take on responsibilities which an unbiased observer would deem to be creative (Colton and Wiggins 2012). More specifically, my supervisors and I are doing research in generative poetry. Poetry is an aspect of linguistic creativity, one of the most popular domains in computational creativity research (Loughran and O’Neill 2017). Computer scientists have been creating computer-generated poetry since the 1950s (Lutz 1959). Current generative poetry research efforts range from interactive co-creation systems (eRoGK7 and others 2011 present; Kantosalo et al. 2014) to recurrent neural networks (Goodwin 2016; Wang et al. 2016) to a system that makes goal-directed edits on its own output (Gervás 2016). Despite many advances in the field, it is still very difficult to create generative poetry that makes sense (Funkhouser 2007). Odd, ungrammatical, and incoherent language persists even in systems that use sophisticated techniques (Goodwin 2016; Manurung, Ritchie, and Thompson 2012).

We have chosen for our current research to focus on found poetry: that is, poetry created by juxtaposing phrases from previously existing work. Examples of found poetry created using computers include the New York Times Haiku project (Harris 2013 present), which creates haiku out of sentences found in the New York Times, and DopeLearning (Malmi et al. 2015), which machine learns relationships between lines of rap lyrics. A focus on found poetry allows us to sidestep, for a moment, the issue of creating semantically coherent language from scratch. By training a computer to choose the best lines from an existing corpus, we can instead focus on the more interesting (to us) issue of what makes text poetic.

Copyright © 2017, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Studies of Poetry Evaluation

A key issue in all computational creativity is evaluation (Jordanous 2011). It is important to us that evaluations be falsifiable, evidence-based, and grounded on an interdisciplinary level—drawing both from the psychology of creativity and from domain-specific knowledge about the art, science, or craft being generated by the computer. Computational creativity often suffers from methodological confusion in which evaluations fail to meet these standards—for example, evaluating through a rhetorical argument by the authors (Jordanous 2011; Pearce, Meredith, and Wiggins 2002).

To help remedy this, we performed a thorough interdisciplinary survey which also served as the depth requirement of my PhD. The survey compiles information from psychology and philosophy which is applicable to the measurement of creativity, and discusses its possible applicability, or lack thereof, to computational creativity. In updated and modified form, it is now out on submission to relevant journals, as is an expanded version of a smaller published survey of techniques used in generative poetry (Lamb, Brown, and Clarke 2016b).

We performed two of our own experiments, specific to the domain of poetry. First, by studying non-expert judges’ responses to human-generated poems, we showed that expert and non-expert preferences in this domain can oppose each other, and that popular metrics for evaluating computational creativity fail to remedy this (Lamb, Brown, and Clarke 2015a). This study highlights the need either to use expert judges for creativity, or to separate experts from non-experts and be clear about which group is the intended audience.

Second, by studying expert judges’ responses to computer-generated poems, we developed our own domain-specific criteria for assessing generative poems (Lamb, Brown, and Clarke 2016a). Our criteria are as follows:

- **Reaction:** The poem should evoke feelings of enjoyment and/or interest.
- **Meaning:** The poem should intentionally convey a specific idea. Even if the poem is difficult to understand, its difficulty should enhance the underlying meaning.
- **Novelty:** The poem should be unusual or surprising in some way.

- **Craft:** The poem should make effective use of poetic techniques in service of the other three criteria.

TwitSonnet and Future Work

One can only do so much with evaluation if there is no system to evaluate. Our system, TwitSonnet, is a found poetry system based on Twitter. The idea of TwitSonnet is to construct entertaining poetic summaries of news and other trending events by compiling poems based on rhyming, topical tweets. Twitter found poetry has been previously performed by Pentametron (Bhatnagar 2012) and Mobtwit (Hartlová and Nack 2013), but Pentametron creates poems based only on rhyme/meter and without a topic, while Mobtwit focuses on putting tweets together based on location rather than about a specific topic.

TwitSonnet uses a step-by-step process: harvest tweets based on time, keyword, and number of syllables; detect sets of appropriately rhyming tweets; rate the harvested tweets based on various desired criteria; and put the highest rated tweets together into a sonnet, or, if not enough rhyming tweets are available, a shorter poem.

In 2015, we built a TwitSonnet prototype which crowdsourced the ratings (Lamb, Brown, and Clarke 2015b). Crowdfunder workers rated the tweets based on topicality, sentiment, and imagery. TwitSonnet put the rated tweets together, and we then evaluated the poems. Poems which were strongly topical and had positive sentiment were given a significantly better reception than baseline poems (those for which only rhyme and syllable criteria had been used, and not line ratings).

Later we expanded TwitSonnet into a version which could perform its own tweet rating, and thus avoid the need to hire crowd workers for each poem, while also giving more creative responsibility (and therefore, perhaps, more creativity) to the system itself (Lamb, Brown, and Clarke 2017). This time the tweet ratings were based on our set of four poetic criteria (Lamb, Brown, and Clarke 2016a). We posted the results on <http://twitsonnet.tumblr.com>. Unfortunately, this time the poems which were created using automatic line rating did not do better in the evaluation than the baseline poems. In other words, the automated version of the tweet ratings was not adding any poetic value, and was not useful in its current form.

A possible future avenue of work would be to increase the sophistication of TwitSonnet's line rating techniques. However, we are planning to go the opposite route and to find new applications for the crowdsourced version of the system. There is an established tradition of making found poetry out of more substantive parts of news items, such as Hart Seely's "The Poetry of D.H. Rumsfeld" (Seely 2003)—as opposed to Twitter commentary and users' arguments with each other. We feel that in the present political climate, a system like this, with a text source other than Twitter, would provide more focused and relevant news poems, and we are interested in using the crowdsourced version of TwitSonnet to mine especially poignant or relevant lines from substantive public documents regarding current events.

References

- Bhatnagar, R. 2012. Pentametron. <http://pentametron.com>, accessed December 9, 2015.
- Colton, S., and Wiggins, G. A. 2012. Computational creativity: The final frontier? In *Proceedings of the 20th European conference on artificial intelligence*, 21–26. IOS Press.
- eRoGK7, et al. 2011-present. Gnoetry daily. <https://gnoetrydaily.wordpress.com>, accessed December 9, 2015.
- Funkhouser, C. T. 2007. *Prehistoric digital poetry: an archaeology of forms, 1959-1995*. University Alabama Press.
- Gervás, P. 2016. Constrained creation of poetic forms during theme-driven exploration of a domain defined by an n-gram model. *Connection Science* 28(2):111–130.
- Goodwin, R. 2016. Adventures in narrated reality. <https://medium.com/artists-and-machine-intelligence/adventures-in-narrated-reality-6516ff395ba3>. Retrieved November 29, 2016.
- Harris, J. 2013-present. Times haiku: serendipitous poetry from the New York Times. <http://haiku.nytimes.com/>, accessed December 9, 2015.
- Hartlová, E., and Nack, F. 2013. Mobile social poetry with tweets. *Bachelor thesis, University of Amsterdam*.
- Jordanous, A. 2011. Evaluating evaluation: Assessing progress in computational creativity research. In *Proceedings of the Second International Conference on Computational Creativity*, 102–107.
- Kantosalo, A.; Toivanen, J. M.; Xiao, P.; and Toivonen, H. 2014. From isolation to involvement: Adapting machine creativity software to support human-computer co-creation. In *Proceedings of the Fifth International Conference on Computational Creativity*, 1–7.
- Lamb, C.; Brown, D. G.; and Clarke, C. L. 2015a. Human competence in creativity evaluation. In *Proceedings of the Sixth International Conference on Computational Creativity*, 102–109.
- Lamb, C. E.; Brown, D. G.; and Clarke, C. L. 2015b. Can human assistance improve a computational poet? In *Proceedings of BRIDGES*, 37–44.
- Lamb, C.; Brown, D. G.; and Clarke, C. L. 2016a. Evaluating digital poetry: Insights from the CAT. In *Proceedings of the Seventh International Conference on Computational Creativity*.
- Lamb, C.; Brown, D. G.; and Clarke, C. 2016b. A taxonomy of generative poetry techniques. *Proceedings of BRIDGES*.
- Lamb, C.; Brown, D. G.; and Clarke, C. L. 2017. Incorporating novelty, meaning, reaction and craft into computational poetry: a negative experimental result. In *Proceedings of the Eighth International Conference on Computational Creativity*. Association for Computational Creativity.
- Loughran, R., and O'Neill, M. 2017. Application domains considered in computational creativity. In *Proceedings of the Eighth International Conference on Computational Creativity*, 197–204. Association for Computational Creativity.
- Lutz, T. 1959. Stochastische texte. *Augenblick* 4(1):3–9. qtd.

in (Roque, 2011); translated to English by Helen MacCormack, 2005.

Malmi, E.; Takala, P.; Toivonen, H.; Raiko, T.; and Gionis, A. 2015. DopeLearning: a computational approach to rap lyrics generation. *arXiv preprint arXiv:1505.04771*.

Manurung, R.; Ritchie, G.; and Thompson, H. 2012. Using genetic algorithms to create meaningful poetic text. *Journal of Experimental and Theoretical Artificial Intelligence* 24(1):43–64.

Pearce, M.; Meredith, D.; and Wiggins, G. 2002. Motivations and methodologies for automation of the compositional process. *Musicae Scientiae* 6(2):119–147.

Seely, H. 2003. The poetry of Donald Rumsfeld: Recent work by the secretary of defense. http://www.slate.com/articles/news_and_politics/low_concept/2003/04/the_poetry_of_dh_rumsfeld.html, accessed June 27, 2017.

Wang, Z.; He, W.; Wu, H.; Wu, H.; Li, W.; Wang, H.; and Chen, E. 2016. Chinese poetry generation with planning based neural network. *arXiv preprint arXiv:1610.09889*.