

Poetry of the Crowd: A Human Computation Algorithm to Convert Prose into Rhyming Couplets

Quanze Chen and Chenyang Lei and Wei Xu and Ellie Pavlick and Chris Callison-Burch

Computer and Information Science Department

University of Pennsylvania, Philadelphia, PA, USA

{cquanze, chenylei, epavlick}@seas.upenn.edu,

{xwe, ccb}@cis.upenn.edu

Abstract

Poetry composition is a very complex task that requires a poet to satisfy multiple constraints concurrently. We believe that the task can be augmented by combining the creative abilities of humans with computational algorithms that efficiently constrain and permute available choices. We present a hybrid method for generating poetry from prose that combines crowdsourcing with natural language processing (NLP) machinery. We test the ability of crowd workers to accomplish the technically challenging and creative task of composing poems.

Introduction

Automatic poetry generation is a fascinating natural language generation challenge that has been attempted by many researchers and engineers in the past. Most previous work has approached the problem as purely mechanical, either focusing on modern poetry generation from random words or on a specific subproblem like meter analysis (Greene, Bodrumlu, and Knight 2010). Generating poems that are well structured and meaningful remains an unsolved problem. In this research we present a **human computation algorithm** for **generating poetry from prose**. We test the capability of crowdsourcing to accomplish this complex creative activity with the aid of NLP algorithms for expanding lexical choice, identifying rhyming words, and calculating meter.

We contend that for tasks like poetry generation, humans excel in various aspects like picking contextually relevant phrase substitutions and composing grammatically correct sentences from a suggested set of words. Computers excel at other dimensions. Existing algorithms and datasets can suggest potential paraphrases, or calculate the meter of a verse, or list rhyming words, but computers have trouble constructing sensible sentences. We can take advantage of the skills of both humans and computers by splitting our task into a series of subtasks (sketched in Figure 1).

Our Approach

One approach to generating poetry from prose is to consider the problem is two-phased. The first phase is the collection, or brainstorming, of a pool of valid alternatives for each of the expressions in the prose. This creates the set of phrase variations we build our poetry from. The second phase is assembly of the output poetry under the required constraints using choices from the first phase. In the case of rhyming couplets, we will focus on the meter and rhyme constraints. For each task we are able to make effective use of machines and humans.

Lexical Expansion The first step in our process is to expand all the words and phrases in the input prose into meaning-equivalent alternatives. We start by querying a database of English paraphrases to compile sets of possible substitutions for each word or phrase in an input sentence. We use the Paraphrase Database (PPDB), an automatically

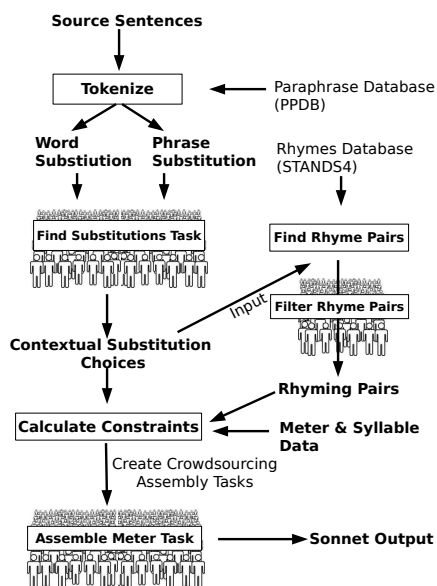


Figure 1: We combine the creative powers and language skills of people with computational algorithms for expanding paraphrase, identifying rhymes and calculating meter

"Wandering through rows of stalls examining **workhorses** and prize hogs may seem to ... have been a strange way for a scientist (especially an elderly one) to spend an afternoon, but there was a certain logic to it."

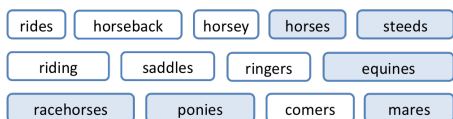


Figure 2: Lexical substitutions filtered by crowdsourcing workers regarding to the context.

constructed collection of 120 million meaning-equivalent re-write rules (Ganitkevitch, VanDurme, and Callison-Burch 2013).

A naive application of PPDB paraphrases is not reliable. The programatically constructed PPDB contains some level of noise and incorrect paraphrases. Moreover, many of the paraphrases are correct only in particular contexts (McCarthy and Navigli 2007). However, we can enhance the quality of our word expansion by having crowd workers filter for contextually appropriate paraphrases. Workers are shown the sentence context and asked to select context fitting paraphrases. We take a majority vote from 5 workers' judgments about each paraphrase. Paraphrases that receive 3 or more votes stay in the substitution set.

Meter Different poem types have different meter constraints. A sonnet should follow iambic pentameter constraints, implying 5 weak-strong syllable pairs per line. It is often difficult for untrained crowd workers to perform scansion on a large set of choice words. Providing pre-calculated meter greatly simplifies this task. We obtain the stress pattern for each word by combining lookups in the CMU pronunciation dictionary (Bartlett, Kondrak, and Cherry 2009) with analysis of predicted meter probability results from a machine learning algorithm trained on sonnet data (Greene, Bodrumlu, and Knight 2010). If a sufficiently reliable result cannot be found, we approximate the meter by programmatically computing the syllable count. We use this procedure to collect all possible meter patterns for each alternative choice generated in the expansion step.

Rhyme Rhyme information is acquired through the STANDS4 API, an online dictionary service.¹ We collect rhyme information of each word encountered in the original text or expanded word set. The intersection of the rhyming sets across different words is taken to find rhyming pairs of words. Using the expanded word set greatly improves our chances of finding valid matching rhyme pairs. In the preliminary experiment, we use the first 5 paragraphs from the book *The Wisdom of Crowds* by James Surowiecki as the input text. We were able to find 179 rhyming pairs (82 unique words) in the original prose, however, with the expanded word set we are able to increase this number to 911 rhyming pairs (295 unique words).

¹http://www.rhymes.net/rhymes_api.php

Writing Iambic Pentameter We ask crowd workers to write lines of poetry based on the input prose and expanded alternative word choices. The worker is presented with an interface showing the original sentence and available word set along with all meter information. An algorithm dynamically compares the composed sentence and the alternative words set to an iambic pentameter template. The worker can build a sentence with the guarantee that available words always fit the meter constraint. To guarantee rhyme, we constrain the ending word or phrase to be one drawn from a rhyming pair in word sets across different sentences and adjust the algorithm template to omit the ending meter accordingly. We rely on the workers to explore the search space and assemble sensible sentence paths capturing aspects of the input prose.

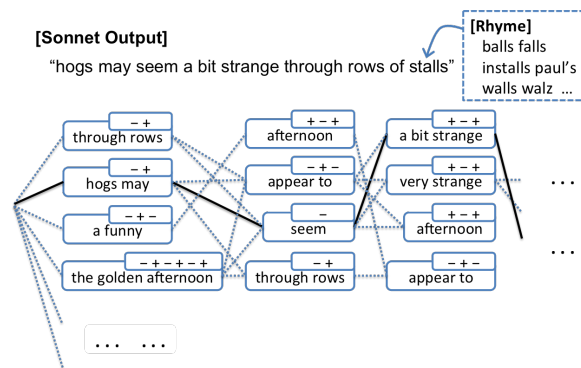


Figure 3: A crowd worker composes a sentence path with the “- + - + - + - +” sonnet stress pattern. The ending word *stalls* rhymes with other words that terminate sentences constructed in another worker’s task.

References

- [Bartlett, Kondrak, and Cherry 2009] Bartlett, S.; Kondrak, G.; and Cherry, C. 2009. On the syllabification of phonemes. In *Proceedings of NAACL*.
- [Ganitkevitch, VanDurme, and Callison-Burch 2013] Ganitkevitch, J.; VanDurme, B.; and Callison-Burch, C. 2013. PPDB: The paraphrase database. In *Proceedings of NAACL*.
- [Greene, Bodrumlu, and Knight 2010] Greene, E.; Bodrumlu, T.; and Knight, K. 2010. Automatic analysis of rhythmic poetry with applications to generation and translation. In *Proceedings of EMNLP*.
- [McCarthy and Navigli 2007] McCarthy, D., and Navigli, R. 2007. Semeval-2007 task 10: English lexical substitution task. In *Proceedings of the 4th International Workshop on Semantic Evaluations*.
- [Surowiecki 2005] Surowiecki, J. 2005. *The Wisdom of Crowds*. Random House LLC.

Input prose

One day in the fall of 1906, the British scientist Francis Galton left his home in the town of Plymouth and headed for a country fair. Galton was eighty-five years old and beginning to feel his age, but he was still brimming with the curiosity that had won him renown and notoriety for his work on statistics and the science of heredity. And on that particular day, what Galton was curious about was livestock. Galton's destination was the annual West of England Fat Stock and Poultry Exhibition,

a regional fair where the local farmers and townspeople gathered to appraise the quality of each other's cattle, sheep, chickens, horses, and pigs. Wandering through rows of stalls examining workhorses and prize hogs may seem to have been a strange way for a scientist (especially an elderly one) to spend an afternoon, but there was a certain logic to it.

Galton was a man obsessed with two things: the measurement of physical and mental qualities, and breeding. And what, after all, is a live-stock show but a big showcase for the effects of good and bad breeding. Breeding mattered to Galton because he believed that only a very few people had the characteristics necessary to keep societies healthy. He had devoted much of his career to measuring those characteristics, in fact, in order to prove that the vast majority of people did not have them.

At the International Exhibition of 1884 in London, for instance, he set up an 'Anthropometric Laboratory', where he used devices of his own making to test exhibition-goers on, among other things, their "Keenness of Sight and of Hearing, Colour Sense, Judgment of Eye, and Reaction Time." His experiments left him with little faith in the intelligence of the average person,

"the stupidity and wrong-headedness of many men and women being so great as to be scarcely credible." Only if power and control stayed in the hands of the select, well-bred few, Galton believed, could a society remain healthy and strong. As he walked through the exhibition that day, Galton came across a weight-judging competition.

A fat ox had been selected and placed on display, and members of a gathering crowd were lining up to place wagers on the weight of the ox. Or rather, they were placing wagers on what the weight of the ox would be after it had been "slaughtered and dressed." For six pence, you could buy a stamped and numbered ticket, where you filled in your name, your address, and your estimate.

The best guesses would receive prizes. Eight hundred people tried their luck. They were a diverse lot. Many of them were butchers and farmers, who were presumably expert at judging the weight of livestock,

but there were also quite a few people who had, as it were, no insider knowledge of cattle. "Many non-experts competed,"

Galton wrote later in the scientific journal *Nature*, "like those clerks and others who have no expert knowledge of horses, but who bet on races, guided by newspapers friends and their own fancies."

The analogy to a democracy, in which people of radically different abilities and interests each get one vote, had suggested itself to Galton immediately. "The average competitor was probably as well fitted for making a just estimate of the dressed weight of the ox, as an average voter is of judging the merits of most political issues on which he votes," he wrote.

Generated poem

for a fair Galton left his home in town
a number of years gave him wide renown
the science of his work had won him that
west to the british live stock and fair fat

the farmers' horses, cattle, poultry, sheep
to Galton - reproduction matters deep
hogs may seem scientifically strange but
all of 'em show off the impact of what

a man obsessed with such things Galton was
few people really mattered just because
the breton was obsessed with things: these two
we had to keep traits pure and measured too

belief that breeding mattered only then
left him with very little faith in men
he implemented tests of sense and sight
the average person was not just nor right

the general public did not demonstrate
both men and women workers to be great
spent much of his career now coming to
that day when a weight-judging he walked through

believe that a well-bred few should remain
the ox had been selected for weight gain
for six pence, a ticket, your name filled in
a wager on the ox once slaughtered been

great many butchers judging? oh quite few
the hundreds tried their luck - good luck to you
the crowds place wagers on the member's fat
that day he came across both the fact that

insider knowledge some had it but few
a farmer expert at weights judge you
most probably there were a great many who
insider knowledge had it but a few

like clerks who have no knowledge in their bet
each person had a single vote to get
but when the judgement was an average rate
the average estimate... its body weight!

the average voter as good as himself
each single interest for one thing itself
as fitted for ox as votes he now wrote
the peoples' interests get put to the vote

Table 1: Example poem generated from the opening chapter of *The Wisdom of Crowds*. Turkers generated 1,767 lines in iambic pentameter from the text. Each line was constrained to end in a potential rhyming word. The authors selected the poem's couplets from among the resulting 42,093 rhyming pairs, and lightly edited the poem for coherence.