MAPLE – MAsking words to generate blackout Poetry using sequence-to-sequence LEarning

Aditeya Baral
PES University, Bangalore, India
aditeya.baral@gmail.com

Himanshu Jain
PES University, Bangalore, India
nhimanshujain@gmail.com

Deeksha D
PES University, Bangalore, India
deekshad132@gmail.com

Mamatha HR
PES University, Bangalore, India
mamathahr@pes.edu

Abstract

Poetry has morphed rapidly over changing times with non-traditional forms stirring the creative minds of people today. One such type of poetry is blackout poetry. Blackout poetry is a form of poetry in which words in a passage are masked, except for a few which when combined together in order to convey some meaning. With the recent developments in Natural Language Processing aiming to simulate human creativity, we propose a novel approach to blackout poetry generation employing deep learning. We explore four different architectures, namely an encoder-decoder with Bidirectional Long Short-Term Memory (LSTM) and Attention, a Bidirectional LSTM Conditional Random Fields (LSTM-CRF) architecture, Bidirectional Encoder Representations from Transformers (BERT) and Robustly Optimized BERT Pre-training Approach (RoBERTa). The first architecture employs abstractive summarization and the remaining employed sequence labelling to generate poetry. The Transformer based architectures prove to be the best working models, and were also able to pass a Turing Test as well.

1 Introduction

Poems are seen as an outlet through which a poet can express their creativity and emotions and deliver strong and vibrant messages to the readers. Some forms of poetry use rhyming schemes of two or more lines while some other poetry forms impress the reader through the beauty of the words selected and their arrangement. The latter type of poems are free-form, and they don't follow any formal structures.

Blackout poetry (Miller, 2017) is the most recent form of poetry in which one picks out words from a passage to generate free-form poems. Such poems may provide a completely different sense as opposed to the meaning of the passage. This art of forming poems from any passage has swiftly gained popularity over the last decade.

Our work aims to generate blackout poetry from any given passage. We use existing state-of-the-art architectures to generate these free-form, blackout poems using techniques like abstractive summarization and sequence labelling. The results have shown that Transformers are very effective in generation of such poems but no single model is capable of producing satisfactory results. Evaluation of our model was done by performing the Turing Test (Wikipedia, b) where we compared the poems generated by humans and machines.

2 Background

Blackout poetry is a recently established form of poetry that rose to popularity in 2005. A passage consisting of words is taken and "blackened" or masked out, except for a few words such that these leftover words when combined together convey some meaning. Often, instead of simply masking out the words, blackout poets tend to draw patterns related to the poem. It is also seen as a way to repurpose old newspapers and magazines. This form of poetry was popularized by Austin Kleon (Kleon) (see Figure 1), who created such poetry from old newspapers. The New York Times also features a digital blackout poetry generator (Times, 2014) that allows visitors to generate blackout poems on their website.

3 Previous Work

The first and only known work of automated blackout poetry generation was observed during the National Novel Generation Month (NaNoGenMo) (Month) 2016. Liza Daly's (Daly) work was able to generate blackout poems from any given passage of text by looking up sequences of words that followed a given set of parts-of-speech grammar

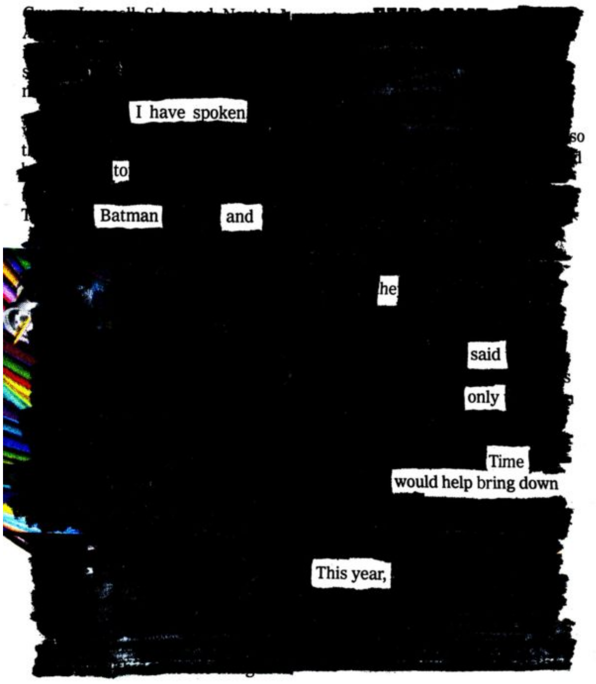

Figure 1: Blackout poetry by Austin Kleon for the New York Times

rules (see Table 1). Although her approach was rule-based, on rare occasions it was able to pick out sequences of words that were able to convey some meaning. However, the approach is very restricted, since it only looks for a given number of grammar-rules and extracts the first sequence of words matching a rule. Additionally, there is no way to verify whether the generated poem is syntactically or semantically correct.

4 Proposed System

Our ultimate objective is to apply deep learning to generate blackout poems which match or possibly beat Liza Daly's model (which use fixed grammar-rules) in terms of human-nature and readability. We look at two major Natural Language Processing techniques – abstractive summarization (Gupta) and sequence labelling (Wikipedia, a). Abstractive summarization is the process of generating summaries of any given passage, such that the lines in the summary are not chosen from the passage itself, and are thus rewritten by the machine. Sequence labelling , also known as token classification is a method to assign a class or a label to every word or token in a given sequence of text.

Passage	**Generated Blackout Poem**
Loving you feels like the soft touch of velvet rain.Over the hills we walk together again. Violet eyes, ruby lips, a beautiful smile. Every step a together lips smile	-
The skies can't keep their secret! They tell it to the hills The hills just tell the orchards. And they the daffodils! A bird, by chance, that goes that way Soft overheard the whole.	secret hills tell daffodils
The mountain sat upon the plain In his eternal chair, His observation omnifold, His inquest everywhere. The seasons prayed around his knees, Like children round a sire: Grandfather of the days is he, Of dawn the ancestor.	sat his chair
Summer for thee grant I may be When summer days are flown! Thy music still when whippoorwill And oriole are done! For thee to bloom, I 'll skip the tomb And sow my blossoms o'er!	when flown blossoms
Mine enemy is growing old I have at last revenge. The palate of the hate departs; If any would avenge, Let him be quick, the viand flits, It is a faded meat. Anger as soon as fed is dead;	mine enemy makes

Table 1: Poems generated using Liza Daly's rules

5 Workflow

To obtain satisfactory results, the process of data collection as well as pre-processing had to be paid significant attention since data was scarce. We experiment and try four state-of-the-art architectures for the two major approaches to generating blackout poetry.

5.1 Data Collection

Due to the lack of any publicly available dataset containing passages and extracted blackout poems, we resort to using Liza Daly's method to synthetically generate a dataset.

To ensure that our generated data possesses high level of creativity as well as to enforce artistic nature, we use a large collection of traditional poems written and submitted to a public archive (kag). These poems were used as our passages for input. We also restrict the size of the passage to between 8 and 120 words for computational reasons.

The grammar rules employed are different from Liza Daly's work. To obtain a set of systematic grammar rules that would ensure a higher number of sequences being extracted, we choose frequently used grammar rules in poetry. Although it would not completely remove the issue of generating nonsensible data, it did help produce more syntactically accurate data. To do this, a large number of haikus – short, 3 line poems that can be read as a single sentence – were obtained, and the parts-of-speech tags for its constituent words was retrieved. We analysed the repeating nature of the sequence of the parts-of-speech and choose the most frequent rules (see Table 2) (see Table 3).

5.2 Data Pre-Processing

We retain only those poems which are at least 5 words in length. We pre-process our data by first removing any form of bad symbols (characters apart from letters, numbers and punctuation) in the generated poems as well as passages. The choice to convert the data into lowercase is decided based on the model architecture being used (see Table 4). We further remove any duplicate passages and poems by sampling unique pairs from the dataset (see Table 5).

5.3 Evaluation Metric

Since we are attempting to simulate creativity in a machine, a quantitative metric cannot be used to evaluate our model. We resort to using a Turing Test to gauge the quality of our poems. A questionnaire was constructed with 8 human written poems and the best machine generated poems each and were randomly shuffled. We asked the audience to choose among these three options – written by a human, machine or unable to draw a conclusion. The questionnaire was shared with people in the age group 18-22 and observed 120 responses.

Parts-of-Speech Grammar Rule	**Frequency**
PROPN PROPN DET ADJ NOUN ADP DET NOUN	7
PROPN PROPN DET ADJ NOUN ADP DET NOUN NOUN	4
ADJ NOUN DET NOUN VERB ADP DET NOUN	3
VERB PROPN DET NOUN NOUN	2
NOUN ADJ NOUN ADP DET NOUN	2
VERB DET NOUN ADP DET NOUN	2
VERB DET NOUN ADP PROPN NOUN	2
NOUN NOUN DET ADJ NOUN ADP DET NOUN	2
NOUN NOUN NOUN VERB ADP DET NOUN	2

Table 2: Parts-of-Speech Grammar rules

5.4 Models

5.4.1 Abstractive Summarization using Bidirectional LSTM and Attention

We employ abstractive summarization by using an encoder-decoder architecture using a Bidirectional LSTM (Hochreiter and Schmidhuber, 1997) (Rumelhart and McClelland, 1997) and unidirectional LSTM respectively. We also include Bahdanau's Attention (Bahdanau et al., 2014) between the encoder and the decoder to increase performance on large sequences. fastText (Bojanowski et al., 2017) Word embeddings trained on passages are used to initialise the fixed embedding layer. The latent dimensions of the context vector are set to 1024 and the size of the embedding layer is set to 100.

The model uses the Adam (Kingma and Ba, 2014) optimiser and sparse categorical cross-

Passage	Generated Blackout Poem
Loving you feels like the soft touch of velvet rain.Over the hills we walk together again. Violet eyes, ruby lips, a beautiful smile. Every step a blessing walking down the aisle.	violet ruby a beautiful smile
The skies can't keep their secret! They tell it to the hills The hills just tell the orchards. And they the daffodils! A bird, by chance, that goes that way Soft overheard the whole.	secret hills the daffodils by that way
The mountain sat upon the plain In his eternal chair, His observation omnifold, His inquest everywhere. The seasons prayed around his knees, Like children round a sire: Grandfather of the days is he, Of dawn the ancestor.	mountain seasons a sire days of the ancestor
Summer for thee grant I may be When summer days are flown! Thy music still when whippoorwill And oriole are done! For thee to bloom, I 'll skip the tomb And sow my blossoms o'er!	summer days music
Mine enemy is growing old I have at last revenge. The palate of the hate departs; If any would avenge, Let him be quick, the viand flits, It is a faded meat. Anger as soon as fed is dead;	mine enemy is revenge

Table 3: Blackout Poetry generated using statistically obtained rules

Dataset Attribute	Value
Number of Passages	54629
Vocabulary Size (Cased)	114562
Vocabulary Size (Uncased)	100820
Maximum Passage Length	120
Minimum Poem Length	5

Table 4: Dataset attributes after pre-processing

Dataset Attributes	Total	Train	Test
Number of Passages	16903	15222	1681
Vocabulary Size (Cased)	114562		
Vocabulary Size (Uncased)	100820		

Table 5: Dataset Attributes after sampling

entropy as the loss function. The model is trained for 10 epochs. Our training data consists of the passage as the input, and the poem as the expected output (see Table 6).

5.4.2 Bidirectional LSTM-CRF

Our second model uses a Bidirectional LSTM with a Conditional Random Field (Lafferty et al., 2001) layer to perform sequence labelling. A single Bidirectional LSTM layer with 512 units is used with a feed-forward layer with softmax as the activation function. The CRF layer is initialised with 2 classes and is connected to the Bidirectional LSTM stack. The embedding layer is once again initialised with the weights from fastText word embeddings obtained from the passages. The model was trained for 5 epochs with Adam as the optimiser and CRF loss (negative log-likelihood for linear chain CRF) as the loss function. Our training data consisted of passages as input and index based position-labels from the poem as the expected output (see Table 7).

Attribute	Value
embedding_size	100
Bidirectional LSTM units	1024
activation function	softmax
epochs	10
loss	sparse-categorical-crossentropy
optimiser	adam

Table 6: Chosen Parameters for Abstrative Summarization using Bidirectional LSTM with Attention

Attribute	Value
embedding_size	100
Bidirectional LSTM units	512
activation function	softmax
epochs	5
loss	crf_loss
optimiser	adam

Table 7: Chosen Parameters for Bidirectional LSTM-CRF

Attribute	Value
model_id	bert-base-cased, bert-base-uncased
epochs	1

Table 8: Chosen Parameters for BERT

5.4.3 BERT

We apply the bidirectional learning of the Transformer architecture for sequence labelling. Two vanilla BERT (Devlin et al., 2018) pre-trained architectures are chosen, which have been fine-tuned on cased as well as uncased data. The dataset is was converted to lowercase for the cased BERT architectures and is used to fine-tune the model. The base BERT model is used for both mentioned architectures and is fine-tuned for 1 epoch (see Table 8).

5.4.4 RoBERTa

The base RoBERTa (Liu et al., 2019) pre-trained architecture is chosen, which has been pre-trained on cased data. This model was fine-tuned for 1 epoch (see Table 9).

5.5 Post-Processing

Post-processing of the generated output is performed to enhance them and bring back the writing style of the passage. These include steps such as prepending skipped articles before a word, appending punctuation from the original passage after a word and capitalisation of the first letters of words which were converted to lowercase during pre-processing.

Attribute	Value
model_id	roberta-base
epochs	1

Table 9: Chosen Parameters for RoBERTa

6 Results

Our results show a significant improvement in quality over the poems generated by Liza Daly. The GPT-2 language model was used to measure the perplexity of the generated poems. Our model was able to obtain an average perplexity score of 5758.87, while Liza Daly's poems obtained an average perplexity score of 6511.533. Since a lower perplexity score indicates a more probable sequence, we can conclude that our model was capable of generating better and more probable poems.

Out of 56k randomly generated poems, only 0.1% of Liza Daly's poems formed valid grammatical sequences while 3% of our generated poems were valid. Although it is to be noted that this comparison is baseless since blackout poems are often grammatically incorrect.

However, we do observe that all our models are highly inconsistent, and no model is able to consistently generate good quality results. This is due to the nature of the training dataset being used, which is synthetically generated using a set of grammar rules and hence contains quite a few bad examples of poetry. We observe that the Transformer based architectures (see Table 10) (see Table 11) (see Table 12) perform nearly the same (but with BERT being more consistent) and both outperform the sequence model based architectures used by a significant margin. The Bidirectional LSTM-CRF model performs the worst, with the model not generating any kind of output.

7 Turing Test Analysis

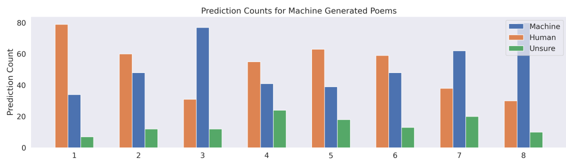

Figure 2: Prediction counts of machine generated poems

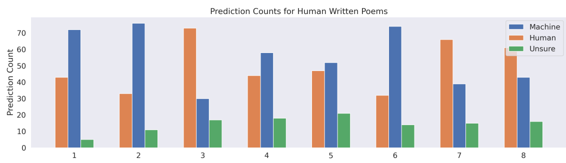

Figure 3: Prediction counts of human written poems

We observe that for a few machine generated poems, people were easily able to make the right

Passage	Generated Blackout Poem
You want someone to hear you run your mouth you talk loud now my aggravations got the best of me you tell your stories i try to wrap my brain I've lost count I pick the best and write your fiction it makes for a laugh or something to fill a void shut your mouth	mouth aggravations the best stories about a laugh
So the steering wheel showed a ship in my dad's coupe from years ago cars in boys' mind brakes just won't slip so the steering wheel showed a ship the fresh minted smell brewed air sip glow flown style blur torn roam wild show so the steering wheel showed a ship in my dad's coupe from years ago	forest grass the thorn flowers like a shore
I am loosed, I am free I have no responsibility no longer to be found the shackles and chains that had me bound a new life waits ahead it is the unknown what I most dread though I'm free as a dove I'd surrender if bound by your love let not go is my plea without you I am lost hold onto me no matter the cost & responsibility shackles that new life	responsibility shackles that new life as a dove
I wonder if the river gets tired, it runs and runs but never stops, foaming swirling round the rocks, it mustn't be tired because if it were it surely would rest, it only runs past to be admired, so rivers never do get tired, though lazy sometimes, yes when rain isn't giving her best but when the rain is feeling well, the river rushes madder still to get to the ocean blue	foaming swirling the tired rivers to the ocean
Every spring after the rain, new life comes again, when seeds are sprouting, and even the smallest petals of grain	spring seeds the smallest flower

Table 10: Poems generated using Abstractive Summarization

Passage	Generated Blackout Poem
Ghosts are many in the stories But quite rare in realities, Yet children are afraid of those Cry at night dreams sometimes	Ghosts are many But rare in realities, sometimes
A pocketful of sympathy Is really rather wonderful. To stop a scratch from stinging. Or a bruise from black and bluing. A pocketful of sympathy - Can stop a heart from hurting, Or catch a tear that's falling Like a raindrop down a cheek. A pocketful of sympathy Costs absolutely nothing, It's the cheapest kind of plaster That you'll ever ever find. And a pocketful of sympathy Is like Lindsay's Magic Pudding 'Cos the more of it you give away The more you leave behind.	pocketful of sympathy Is the cheapest of Magic
Fire never dies, just smoulders, like love you need to fan it, to keep the flame alive. Like the smouldering embers, that needs only a little attention, to become a flame again. So the parting lovers, need only to kiss to ignite the flame, and start the passion again.	Fire smoulders, the embers, a little flame
In the city of sorrows, Is where we see so much Racism, Against Men and women, In the city of sorrows, There are so much injustice.	city sorrows much Racism, the city much injustice
Night whispers in the dark makes eerie sounds with the wind making it so comforting. I hear night sounds like an owl hooting far sitting alone in the dark moonlight still.	Night whispers the dark eerie sounds with the wind

Table 11: Poems generated using BERT

Passage	Generated Blackout Poem
Innocence and desire illusions of my thoughts forsaken into realization, hollow like our ears is the sifting destiny.	innocence illusions my thoughts into realization
When you're lonely I wish you company, when you're sad I wish you happiness, when you're heartbroken I wish you eternal love, when all is chaotic I wish you inner silence, when all seems empty I wish you hope.	heartbroken love is chaotic silence
My soul, sit thou a patient looker-on; judge not the play before the play is done: her plot hath many changes; every day speaks a new scene; the last act crowns the play.	my soul hath many changes; every new scene
The red sun rises without intent and shines the same on all of us. we play like children under the sun. one day, our ashes will scatter— it doesn't matter when. now the sun finds our innermost hearts, fills us with oblivion intense as the forest, winter and sea.	sun rises and shines our hearts
I am loosed, I am free I have no responsibility no longer to be found the shackles and chains that had me bound a new life waits ahead it is the unknown what I most dread though I'm free as a dove I'd surrender if bound by your love let not go is my plea please bring again my captivity without you I am lost hold onto me no matter the cost	responsibility shackles that new life

Table 12: Poems generated using RoBERTa

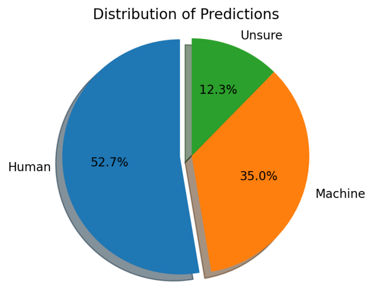

Figure 4: Distribution of predictions for machine generated poems predicted as human written

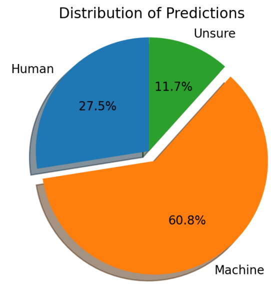

Figure 5: Distribution of predictions for machine generated poems predicted as machine generated

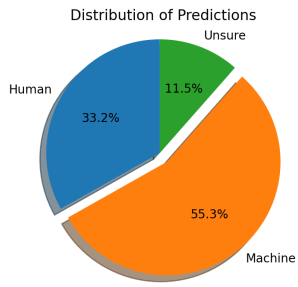

Figure 6: Distribution of predictions for human written poems predicted as machine generated

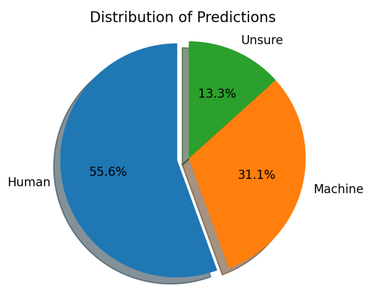

Figure 7: Distribution of predictions for human written poems predicted as human written

prediction (see Figure 8). However, for all cases we observe that the number of predictions passing the Turing Test are only a few more than the number of predictions failing the Turing Test (see Figure 2) (see Figure 3) (see Figure 4). This suggests that although the models were able to pass the Turing Test, they were neither poor to be labelled as machine generated nor great to have a higher prediction count for the other class (see Figure 5) (see Figure 6) (see Figure 7). We also observe that the average number of unsure predictions across all four cases remain about the same (see Figure 9).

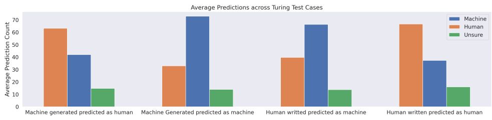

Figure 8: Average Predictions across Turing Test Cases

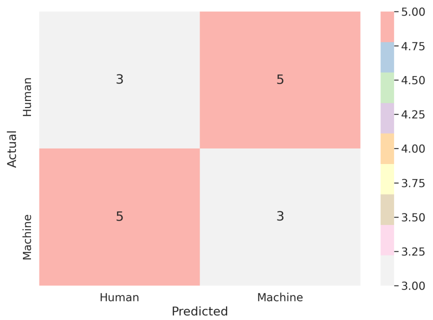

Figure 9: Confusion Matrix of Predictions

8 Conclusion

We thus show through our work how it is possible to generate free-form blackout-poetry using both abstractive summarization as well as sequence labelling techniques. Although the poems were able to pass a Turing Test, the models are highly inconsistent in their results. The Transformer architectures were observed to be the best working models, producing the best results both in terms of a syntactical as well as semantic sense.

The quality of the training dataset has a huge role to play in the result, since the dataset itself was generated synthetically using a set of grammar rules. Replacing this dataset with an actual blackout poetry dataset would greatly improve the performance of the models. Additionally, the poems generated can also be filtered using various linguistic tools to check for valid sequences.

References

Kaggle. https://www.kaggle.com/.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Liza Daly. Blackout. https://github.com/lizadaly/blackout.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Som Gupta. Abstractive summarization: An overview of the state of the art. https://www.sciencedirect.com/science/article/abs/pii/S0957417418307735.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory.

Diederik Kingma, P. and Jimmy Ba. 2014. Adam: A method for stochastic optimization.

Austin Kleon. Newspaper blackout poems. https://austinkleon.com/category/newspaper-blackout-poems/.

John Lafferty, Andrew Mccallum, and Fernando Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach.

E. CE Miller. 2017. What is blackout poetry? https://www.bustle.com/p/what-is-blackout-poetry-these-fascinating-poems-are-created-from-existing-art-78781.

National Novel Generation Month. Nanogenmo. https://nanogenmo.github.io/.

David E. Rumelhart and James L McClelland. 1997. Learning internal representations by error propagation.

New York Times. 2014. Searching for poetry in prose. https://www.nytimes.com/interactive/2014/multimedia/blackout-poetry.html.

Wikipedia. a. Sequence labelling. https://en.wikipedia.org/wiki/Sequence_labeling.

Wikipedia. b. Turing test. https://en.wikipedia.org/wiki/Turing_test.