

## 1 Introduction

- Group members

Kelsi Riley  
Sakthi Vetrivel

- Team name

Caltech Earthquakes

- Division of labour

Kelsi implemented the HMM, and focused on improving it. She also did the preprocessing for the HMM, and the unsupervised algorithm. Sakthi implemented the RNN and worked on improvements for it, as well as the visualization and interpretation of the HMM.

## 2 Pre-Processing

## 3 Unsupervised Learning

## 4 Poetry Generation, Part 1: Hidden Markov Models

## 5 Poetry Generation, Part 2: Recurrent Neural Networks

### Initial Implementation

For our preprocessing for the RNN, we decided to first break down the text file by line, and for each line we parsed, we made sure the line was not empty (not just an endline character) and removed any special characters from it. For example, we wanted "Hello!" and "hello" to be processed as the same sequence of characters. We then finished each line with an endline character and added it to an accumulating string, which held the contents of the processed text file. After processing the input, we created dictionaries to convert each character found in the processed text to an integer, and a dictionary that converted integers to characters. We then generated a data set splitting this string into sequences of 40 consecutive characters, and converting the sequence of characters into a sequence of integers, and used the 41st character of the sequence as the y value, again, after converting it to an int.

We implemented a recurrent neural network using the Keras package for Python3. Using a sequential model, we had two dense LSTM layers of size 200, and one output layer. We calculated our loss using categorical cross-entropy loss, and 'adam' optimization. We then trained this model for 125 epochs, converging to a loss of 0.4146 from a loss of 3.45 after the first epoch. To improve our model, we also looked at more complex RNN, using two dropout layers between the dense layers, and fine tuned the dropout probability to eventually choose 0.4. Initially, the model was only trained for 20 epochs, but we found that the loss still hadn't converged so we continued training until the loss did not improve for 2 epochs. We used a window size of 40 as the instruction suggested, but later moved to smaller window size of 25 in our attempts to improve our model.

To generate our poems, we used our seed sequence "Shall I compare thee to a summer's day?" and processed it, and for a sliding window of 40 characters, predicted the next letter in the sequence. Our

model gives us an array of probabilities for the next character, given the previous 40 characters, so using this array of probabilities, we sample from the population of characters accordingly for a given diversity value. We also counted the number of newline characters that were predicted in the entire generation process, stopping after 14 newline characters, giving us 14 lines in the poem.

Despite training for so many iterations, our LSTM did not successfully learn sentence structure or sonnet structure, as we see in the poems below, few of the words produced are real words. However, for the brief segments of the poems that contain real words, it seems to follow some loose sentence structure. For example: "i love you" and "of the braid".

### Poems

— diversity: 1.5

— Generating with seed: " shall i compare thee to a summers day

"

shall i compare thee to a summers day  
weat kerp tellv in these i would to hath  
my self iil iiddsu your thlltt my self brane ereed  
but ceatt that weal i love you be teildde  
and me altereasonse of the braid and lind  
which treals mind eye is is a lawvereo  
at tenmn lines bety paming iimeshsy  
or at dolg pdr tiat weaddddleditg tiink  
eor shamls iampeut sp will he's shcd might  
sevoy seep tidu thou 'liserpeas nor  
j thak oyck tine suelt so lem so ku haln  
and nock i tas of fold cach or pattry  
but be thy liate me thrugh mights me sn botn  
and in holaskeri hrln with the trwe doth green  
and confoundane farth in thee tie live"

— diversity: 0.75

— Generating with seed: " shall i compare thee to a summers day

"

shall i compare thee to a summers day  
when that wilt nottncry that fell asd feidt  
that it 'bol gor moctatd i do dispilts  
and diary my self i'ck tren to the most  
but when your changent of this weil  
byt sickt his tputatt mot i loow more eeee  
to say toe borcmest whereup the bear  
thy presclv ceauiies ifart he lile artire  
for whose winter's enoling on the rulp're  
when sesimg a betteiry ocrure of thy deeds

therefo thy putlok dead trealed thou art  
o what a worthsed wron delive no mane sehmnts  
oo aly of these falsehe move's fresh ceserity  
then the means me with vinter did stansed  
and dotnt and in habkt and it gaults light"

— diversity: 0.25

— Generating with seed: " shall i compare thee to a summers day  
"

shall i compare thee to a summers day  
aid uosthfr this wirte doth beauty stail  
thou mayst be thy oudsent'st a linit sade  
but when your count in these cannot chind  
o carve norer mine him though mews the even  
but day doth daily draw my sorrows line  
so thy freat gift woon be forgouingnl  
for higheo of line ow well my heart deegines  
so fotth the blow of with dupy steet selbit  
ald my hoade fyen siln liss lysbs'bd and were orisit  
oatt reason haved the stard or thy sweet graces  
beauteous all fellls tine world have erreemed  
more that my self but was donf iis oun  
gow many lambs kild and they acvodance seegng  
and all the dead no nore drtbl dole"

## 6 Additional Goals

## 7 Visualization and Interpretation