# Epic Chat Battles of Historyyy

## Simulating Conversations between Characters with Distinct Styles

CS 726 Advanced Machine Learning

---


| | |
|---|---|
| Anuj Shetty | 140260007 |
| Kumar Ayush | 140260016 |
| Sandesh Kalantre | 140260012 |
| Tejas Srinivasan | 140100025 |


```
OCTAVIA that have my heart parted betwixt two friends
Clown for young charbon the puritan and old poysam the
TOUCHSTONE justices could not take up a quarrel but when the
TOUCHSTONE of an if as if you said so then i said so and
ROSALIND of irish wolves against the moon
FLUELLEN if the enemy is an ass and a fool and a prating
CARDINAL WOLSEY let silence be commanded
PARIS come you to make confession to this father
ROSALIND i would love you if i could tomorrow meet me all together
HELENA what worser place can i beg in your love
HELENA than to be used as you use your dog
VIOLA longer some mollification for your giant sweet
Sixth Citizen mans voice
CONRADE that you frame the season for your own harvest
TOUCHSTONE justices could not take up a quarrel but when the
```

# Contents

# 1 Introduction

Natural Language Processing (NLP) has gained prominence in the last decade with a rising interest in artificial intelligence systems which are capable of communicating with human subjects in a realistic manner [Cambria and White, 2014]. Conventional techniques have focused on rule based hand-tuned methods for tasks such as parsing and part of speech tagging. However, recent advances in deep neural networks have opened a new vista for NLP tasks such as machine translation[Luong et al., 2017] and question answering. Deep learning techniques, particularly those involving recurrent neural networks are now being frequently used for text-generation tasks. One particular advance in recent years has been the sequence-to-sequence (seq2seq) encoder-decoder model which can be adopted and adapted to such problems.

In this project, we consider the problem of generating a self-sustaining dialogue between two or more actors who speak with their own distictive styles. Seq2seq models have been successfully used to train recurrent neural networks to learn the mapping from one sequence of words to another. Here, we aim to see if such networks can be trained to share a common context vector space to enable a common understanding between the two actors.

We pick short dialogues as responses (output) of our program, and a few words before this dialogue is our context (input) of our program. We train for some famous characters across Fiction/Non-Fiction literature, and then feed them all into a group chat. The predicted response is something picked out of the set of responses for a character, which best matches the context.

# 2 Prior Work

There has been a lot of work in related fields in NLP, namely controlled generation of realistic sentences [Hu et al., 2017] , the identification of a characteristic style given a corpus, and then possibly transferring the style to a different author [Singhal et al., ]. Some other models of style transfer also incorporate additional information of a dictionary mapping the 2 corpuses to each other [Jhamtani et al., 2017]. There have also been approaches for modeling threaded discussions on social media using a graph-structured bidirectional LSTM which represents both hierarchical and temporal conversation structure [Zayats and Ostendorf, 2017]. Seq2seq models have enjoyed great success in a variety of tasks such as machine translation, speech recognition, and text summarization [Luong et al., 2017]. However to apply it to our problem would require a large amount of processing power which was not available to us. There have been attempts at conversational modelling using LSTM RNNs, with the model conversing by predicting the next sentence given the previous sentence in a conversation [Vinyals and Le, 2015]. Some more advanced techniques in conversation modelling take into account attention and concepts from discourse analysis such as coherence and cohesion, to more accurately design intelligent neural models which reach beyond the sentence and utterance level [Pierre et al., 2016] .

# 3 Datasets

We considered a variety of possible datasets across all available texts and media, and settled upon the following:

- Shakespeare plays [Jeremy Hilton, ]
- Sigmund Freud - The Interpretation of Dreams [Smith, ]
- Kanye West lyrics [West, ]
- Aristotle's complete works [Aristotle, ]
- Bhagavad Gita [Anonymous, ]

We used the Natural Language Programming Toolkit (NLTK) to tokenize the dataset into sentences. Since the tokenizer takes care of English punctuation styles and hence frivolous splitting into incomplete causes is avoided.

# 4 Methods

## 4.1 Naive Predictor

As a minimum test on our ideas, we decided to build a sentence predictor from a corpus given a context of words. We used the Shakespeare dataset in this section. Let $\mathcal{V}$ be the vocabulary of words in the dataset. $\mathcal{S} = \{\vec{s}_i\}$ be the set of sentences in the order as they appear in the dataset. $\vec{x}$ is a collection of words from $\mathcal{V}$, defined as the context. Let $d$ be a positive integer called the dialogue depth.

### 4.1.1 Probability Model

The task is given a context $\vec{x}$, predict a most probable set of surrounding sentences $\{\vec{s}_i\}_{i=j-d}^{j-1}$ such that words from $\vec{x}$ are contained in the set.

As a start, the probability is used.

$$P(\vec{x}, j|d) \propto \sum_{i=1, i\neq j}^{|x|} \mathcal{I}(x_i \in \cup_{i=j-d}^{j-1} \vec{s}_i)$$

where $\mathcal{I}(...)$ is the indicator function It is basically the count of the number of times words from $\vec{x}$ appear in sentences $\{\vec{s}_i\}_{i=j-d}^{j-1}$.

### 4.1.2 Dialogue Prediction

The final output dialogue $\mathcal{D}$ is defined as:

$$j^* = \arg\max_j P(\vec{x}, j|d)$$
$$\mathcal{D} = s_{j^*}$$

### 4.1.3 Problems with the naive approach

We observed that our naive approach faces problems because the absence of a training procedure to learn a context of the statement. It works only on word level matching and fails to pick up the semantics of the conversation. Moreover, since the score of the context is a fixed quantity (based on the word count), it is prone to get caught in conversation loops.

```
OCTAVIA that have my heart parted betwixt two friends
Clown for young charbon the puritan and old poysam the
TOUCHSTONE justices could not take up a quarrel but when the
TOUCHSTONE of an if as if you said so then i said so and
ROSALIND of irish wolves against the moon
FLUELLEN if the enemy is an ass and a fool and a prating
CARDINAL WOLSEY let silence be commanded
PARIS come you to make confession to this father
ROSALIND i would love you if i could tomorrow meet me all together
HELENA what worser place can i beg in your love
HELENA than to be used as you use your dog
VIOLA longer some mollification for your giant sweet
Sixth Citizen mans voice
CONRADE that you frame the season for your own harvest
TOUCHSTONE justices could not take up a quarrel but when the
```

*Figure 1: A sample conversation predicted using the naive sentence predictor using the Shakespeare dataset. The intial context of [`king`, `slave`] was used.*

```
Aristortle  :  And he acts willingly (for he acts in a sense with knowledge both of what he does and of the end to wh
ich he does it), but is not wicked, since his purpose is good; so that he is half-wicked.
Gita  :  Show me Thy very self, the Eternal God!
Freud  :  There can be no doubt that it is they that are the true disturbers of sleep and not dreams, which, on the c
ontrary are concerned to guard it.
Shakespeare  :  captain macmorris i think look you under your
Aristortle  :  It is evident also that if pleasure, i.e.
Gita  :  This--Pritha's Son!--   Is born of Tamas, "dark" and miserable!
Freud  :  By analysing dreams we can take a step forward in our understanding of the composition of that most marvell
ous and most mysterious of all instruments.
Shakespeare  :  did i ever call for thee to pay thy part
Aristortle  :  But sometimes we praise the ambitious man as being manly and a lover of what is noble, and the unambit
ious man as being moderate and self-controlled, as we said in our first treatment of the subject.
Gita  :  Krishna.
```

*Figure 2: A sample conversation predicted using the naive sentence predictor using four datasets: Shakespeare, Aristotle, Gita and Freud dataset. The intial context of [`virtue`, `man`, `woman`, `soul`, `wrong`, `right`, `evil`] was used.*

## 4.2 Encoder-Decoder Mixing

Ideally, we wanted to generalize the framework defined in the previous section with a seq2seq model of conversation between all actors. This is, however, difficult in practice because we cannot define a loss metric for a conversation between widely different actors e.g. Sigmund Freud and Aristotle, as we do not have data for a valid conversation between the two. We can do this for characters from Harry Potter, but the number of lines for each character if you consider conversations between any two characters is too low for training a seq2seq model.

### 4.2.1 Disjoint seq2seq training

To evade the problems mentioned in the previous paragraph, we try a non-conventional approach. We train the corpus for each individual actor for itself, as is done for generating new sentences that the actor is likely to speak. The $k^{\text{th}}$ sentence is used as the output versus the $(k-1)^{\text{th}}$ sentence as the input. Note that the global minimum for this model would be to generate the full corpus starting from the input sentence. However, providing an input which has not been seen by the model before would lead it to generate new sentences as we would be sampling an unexplored area in the context vector space. Note that this training is disjoint for each actor. How does this help to generate a conversation?

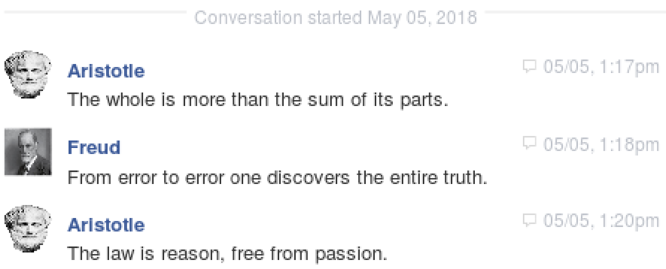



*Figure 3: A sample conversation between Aristotle and Sigmund Freud, manually written for illustration*

Consider a sample conversation between *Aristotle* and *Freud* as seen in Fig 3. Our model would take the first sentence of Aristotle, feed it to the encoder for Aristotle, and decode it using the decoder for Freud. And the same process would be repeated for Aristotle replying to Freud. Of course, the dimensionality of their context vector spaces is the same. This is illustrated in Fig 4.

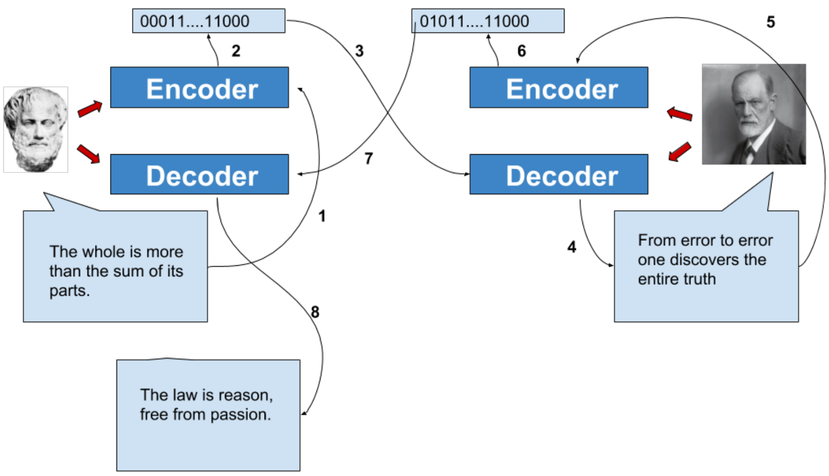



*Figure 4: Illustrating the model of conversation generation between the two actors Aristotle and Freud. The numbers represent the steps. We begin at step 1, with Aristotle saying* `The whole...` *and being encoded to the context vector by Aristotle's encoder in step 2. The context vector is passed to the decoder of Freud in 3 and used to generate the output* `From error to error`*. In steps 5 and 6, the encoder output of Freud is passed to decoder input of Aristotle and the next sentence is generate in 8 as* `The law...`*.*

#### 4.2.2 Intuition

We are led to believe that this model would work due to the following intuitive hypothesis. Assume two actors $A$ and $B$, whose context vector spaces have dimension **3**. Note the following possible scenario.

- $A$ starts by saying `Apples are nice`. Assume that this maps to $\vec{c} = (1, 0, 0)$ in the context vector space.
- Since we are feeding this directly to the decoder for $B$, this implicitly defines a one to one mapping between the context vector spaces of $A$ and $B$. Suppose that $(1, 0, 0)$ corresponds to `Oranges are nice` for $B$.
- $B$ might then reply with `Yes, I like eating oranges too`. Let's say this maps to $\vec{c'} = (1, 0, 1)$ in the context vector space of $B$.
- Here enters the crux of our idea. Because there is an implicit mapping between the two context vector spaces, there is a chance that $A$ would implicitly understand this as `Yes, I like eating apples too` and then it might respond with `Let's eat some apples`.

This is likely to happen because `apples` and `oranges`, being fruits are likely to map to similar vectors in the context vector space. But, realistically, they can map to entirely different classes e.g. `apples` can map to `bikes`. In this case, `Yes, I like riding bikes` becomes a valid response to `Apples are nice.`. And then $A$'s response would likely be far from the correct context. However, note that the semantic structure of the conversation holds even if the pragmatics are destroyed. Thus, we expect that this model will output a comprehensible conversation. We also believe that the contextual integrity is dependent on the intersection of the dense regions of the context vector spaces of the actors.

# 5 Experiments and Results

## 5.1 Naive Prediction

Repository URL: `https://github.com/sandeshkalantre/epic-chat-battles`

We used NLTK for tokenizing the datasets into sentences We used Python 3.5 for codes in this model and total of 150 lines of code. This simple code was written from scratch and no specific library was used. It takes 20 seconds for prediction of each response given the context vector of words.

The results for this model are shown in Fig.1 and Fig.2.

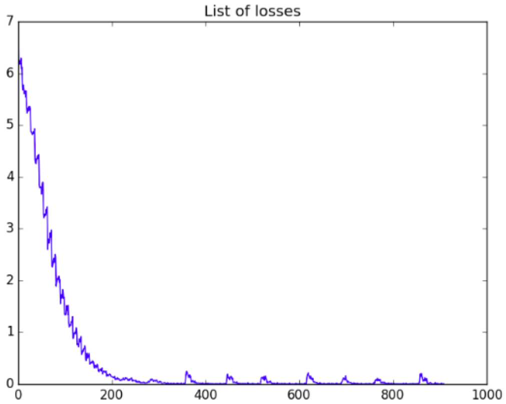
## 5.2 Disjoint Seq2Seq model

Repository URL: `https://github.com/tejas1995/EpicChatBattles`

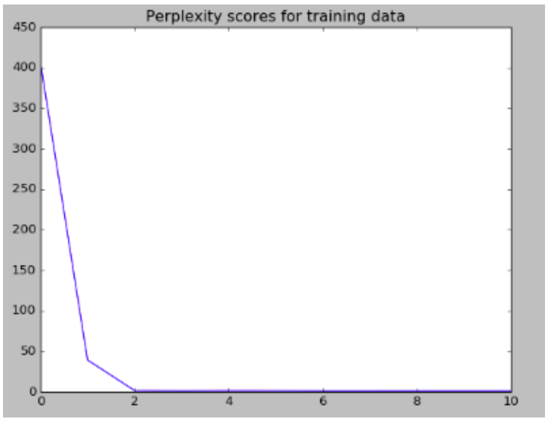
This Seq2Seq model was coded in PyTorch, based on this official PyTorch Seq2Seq tutorial. We trained individual language models for each character (Hamlet from

Shakespeare's 'Hamlet', Henry V from Shakespeare's 'Henry V' and Kanye West from Chicago). These Shakespearean characters were chosen because of the easy availability of a large number of dialogues spoken by them (around 1500 lines spoken by each). We took over 2500 lines of Kanye West rap lyrics, due to his distinctive speech style was a stark contrast to that of Hamlet and Henry V.

To show evidence of the training, below are the cross-entropy loss and the perplexity curves when training the encoder-decoder models of Henry V.




*(a) Loss for training on the Henry V dataset*




*(b) Perplexity for training on the Henry V dataset*

When generating a conversation, after speaker A finishes a dialogue ending with line $L_A$, if the next speaker is B, we mix the encoder of A and decoder of B (as described in Section 4.2) to generate the first line of B's dialogue, $L_B^1$. The following lines of B's dialogue, $L_B^{2...n}$, are generated using the encoder and decoder of B, with previous line acting as input at each stage.

Following are the results when we mix the encoders and decoders of Hamlet and Kanye West to generate a conversation.

| | |
|---|---|
| HAMLET | excellent i faith of the chameleons dish i eat |
| KANYE | yea the been<br>the tell we talk the school or the fame<br>baby the way school get what though<br>and my money money them now how cold looks than a time<br>and when they said and act like they ass it |
| HAMLET | peace could you make peace no more and my hopes<br>thrown well i know it himself what thou hear<br>thou farewell dear of my compass but o cursed spite<br>that ever i know to divide him inventorially would |
| KANYE | but she got to leave it cause they they on me<br>but she said she cant thats that ass me that they got me in they<br>now i gotta move now everybody now i started<br>you too the soul and i was living<br>everything your family on the wake i cant talking back good do it wrong |
| HAMLET | the first well like<br>the first row of the pious chanson but<br>one beast may his heels make you eyes two what<br>if you to be too though you will me |

*Table 1: Results of conversation between only 2 characters, Hamlet and Kanye West*

Below is a conversation between all 3 characters, where we mix encoder of the previous speaker and decoder of the new speaker at each stage.

| | |
|---|---|
| HAMLET | two thousand souls and twenty thousand ducats |
| KANYE | when we been she call you ask me no way<br>how its just control<br>like in the people<br>right will probably cause on the any i cant buy her<br>why do the mall with the dope with a lil girl throw only called |
| HENRY V | that you are like him down the kings quarrel<br>the taste not the king keeps the fault<br>my father of a king and his son<br>and with crispian but his thought |
| HAMLET | come here to the to<br>when he hath let me comply with you in this<br>that was in your own two help all<br>thus i should have and within of nature<br>so horridly to shake our disposition and<br>they are be the as as by my father |
| KANYE | go from it we gon be face<br>and i wont mean no things she got they ass for money<br>and when i know it take closed |
| HAMLET | digested the king and the suits of woe<br>must to the very comment of thy soul<br>observe mine uncle if his occulted guilt<br>do not itself unkennel in one speech<br>it is a damned ghost that we have seen<br>and my imaginations are as foul |
| HENRY V | they are hard of this<br>the man have their nobles that them him<br>no yet it of god death |

*Table 2: Results of conversation between 3 characters, Hamlet, Henry V and Kanye West*

# 6 Challenges

The most challenging part of the project was deciding how to model conversations since it is hard to define a metric for what constitutes a valid conversation. We did not succeed in resolving this in a satisfactory manner, and agreed to carry on with an intuitive hypothesis as has been described in previous sections.

Other than this, we faced the following challenges with our project:

- The training time for NLP models is large. We were unable to find sufficient cluster time to try more experiments with our hypothesis.
- Identifying characters and their lines within a book is non-trivial pre-processing. Additionally, even if we could successfully do this, the number of lines per character is limited. Thus, we were forced to use datasets where either the character lines are segregated already or the complete volume of work is predominantly the thoughts of a solo actor.

# 7 Effort

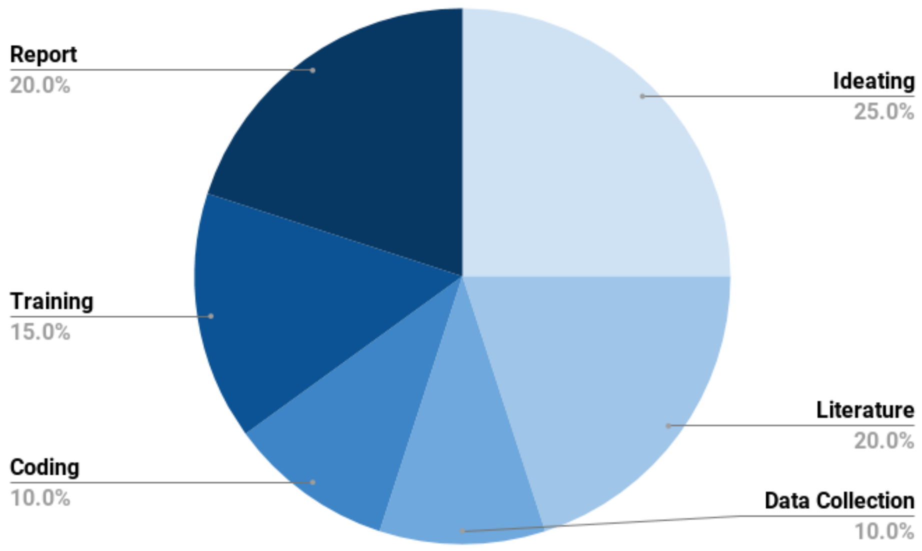



*Figure 6: Fraction of time spent in different parts of the project*

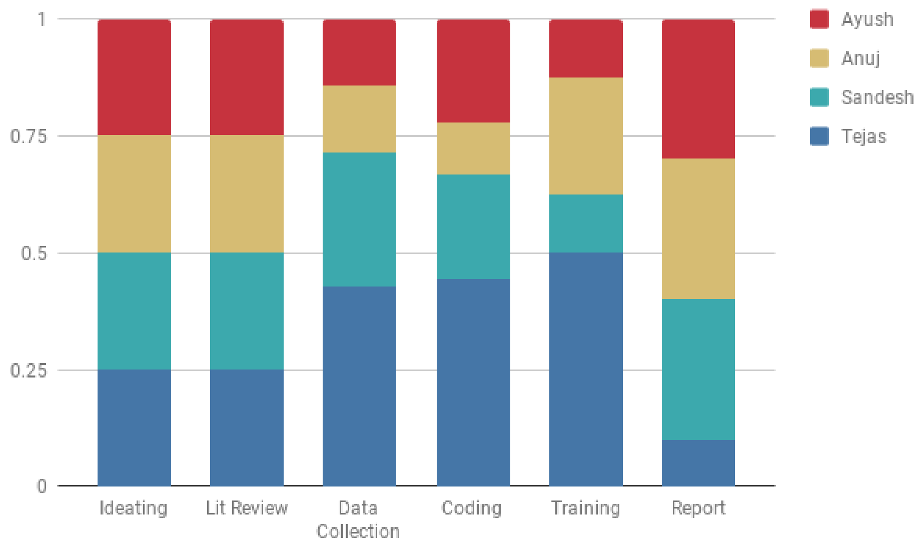



*Figure 7: Fraction of work done by different team members*

# References


[Anonymous, ] Anonymous. The bhagavad-gita
https://www.gutenberg.org/files/2388/2388.txt
last accessed on [03-05-2018].

[Aristotle, ] Aristotle. Nicomachean ethics
http://classics.mit.edu/Aristotle/nicomachaen.mb.txt
last accessed on [03-05-2018].

[Cambria and White, 2014] Cambria, E. and White, B. (2014). Jumping nlp curves: A review of natural language processing research. *IEEE Computational intelligence magazine*, 9(2):48–57.

[Hu et al., 2017] Hu, Z., Yang, Z., Liang, X., Salakhutdinov, R., and Xing, E. P. (2017). Controllable text generation. *arXiv preprint arXiv:1703.00955*.

[Jeremy Hilton, ] Jeremy Hilton, M. The complete works of william shakespeare
http://shakespeare.mit.edu/
last accessed on [03-05-2018].

[Jhamtani et al., 2017] Jhamtani, H., Gangal, V., Hovy, E. H., and Nyberg, E. (2017). Shakespearizing modern language using copy-enriched sequence-to-sequence models. *CoRR*, abs/1707.01161.

[Luong et al., 2017] Luong, M., Brevdo, E., and Zhao, R. (2017). Neural machine translation (seq2seq) tutorial. *https://github.com/tensorflow/nmt*.

[Pierre et al., 2016] Pierre, J. M., Butler, M., Portnoff, J., and Aguilar, L. (2016). Neural discourse modeling of conversations. *CoRR*, abs/1607.04576.

[Singhal et al., ] Singhal, S., Siddarth, K., Agarwal, P., and Garg, A. Text generation and style transfer
http://home.iitk.ac.in/ soumye/cs771a/CS771A_Report.pdf
last accessed on [03-05-2018].

[Smith, ] Smith, I. Complete works of sigmund freud
https://holybooks-lichtenbergpress.netdna-ssl.com/
wp-content/uploads/Sigmund-Freud-The-Complete-Works.pdf
last accessed on [03-05-2018].

[Vinyals and Le, 2015] Vinyals, O. and Le, Q. V. (2015). A neural conversational model. *CoRR*, abs/1506.05869.

[West, ] West, K. Lyrics of the complete discography of kanye west
https://www.azlyrics.com/w/west.html
last accessed on [03-05-2018].

[Zayats and Ostendorf, 2017] Zayats, V. and Ostendorf, M. (2017). Conversation modeling on reddit using a graph-structured LSTM. *CoRR*, abs/1704.02080.