
Final Project Presentation:

“Exploring the effect of dataset on chat bot performance”

— Weiting Zhan and Hyein Jeong —

Outline

- Goals and motivation for the project
- Methods and Experiments
- Results
- Future Work

Goal and motivation for the project

1. To learn how to implement sequence to sequence model to build a chat bot
2. To explore the effect of dataset on chat bot performance

Dataset	number of conversation	number of utterance	dialogue act
Cornell Movie corpus	220,579	304,713	null
Daily Dialog	13,118	–	manually labeled
Cornell + Daily	233,697	304,713	classifier labeled

3. Use statistics method instead of template method, which is a typical way to build a chat bot

Method and Experiment

- Preprocessing data
 - Clean data
 - Formatting data
 - Build dictionary for the dataset
 - Embedding
- Chat bot model
 - Sequence to sequence model
 - Sequence to sequence model + dialogue act
- Experiment
 - Cornell Movie Dataset + Sequence to sequence model
 - Daily Dialogue Dataset + Sequence to sequence model
 - Cornell Movie Dataset + Daily Dialogue Dataset + Sequence to sequence model

Google colab

- Strength: free GPU

name: Tesla T4

major: 7 minor: 5 memoryClockRate (GHz) 1.59

Total memory: 14.73GiB

- Drawback:

Uploaded files and generated files will get deleted

when the runtime is recycled.

Means the results of training will get deleted too.

Don't close your browser, and secure your connection.

NVIDIA T4
Tensor Core GPU

- chat bot 2 daily dialogue.py
- cuda-repo-ubuntu1604_8.0.61-1_amd64.deb
- dialogues_text.txt
- nvidia-410_410.48-0ubuntu1_amd64.deb
- chatbot_weights.ckpt.data-00000-of-00001
- chatbot_weights.ckpt.index
- chatbot_weights.ckpt.meta
- checkpoint

Performance of CPU and GPU

CPU:

Epoch: 1/25, Batch: 100/1265, Training Loss Error: 14.792, Training Time on 100 Batches: 3161 seconds
Epoch: 1/25, Batch: 200/1265, Training Loss Error: 5.671, Training Time on 100 Batches: 1344 seconds
Epoch: 1/25, Batch: 300/1265, Training Loss Error: 4.841, Training Time on 100 Batches: 623 seconds
Epoch: 1/25, Batch: 400/1265, Training Loss Error: 6.565, Training Time on 100 Batches: 844 seconds
Epoch: 1/25, Batch: 500/1265, Training Loss Error: 6.751, Training Time on 100 Batches: 1199 seconds
Epoch: 1/25, Batch: 600/1265, Training Loss Error: 26.360, Training Time on 100 Batches: 1925 seconds

GPU: (20X faster)

Epoch: 1/25, Batch: 100/1265, Training Loss Error: 10.946, Training Time on 100 Batches: 126 seconds
Epoch: 1/25, Batch: 200/1265, Training Loss Error: 5.683, Training Time on 100 Batches: 59 seconds
Epoch: 1/25, Batch: 300/1265, Training Loss Error: 5.952, Training Time on 100 Batches: 31 seconds
Epoch: 1/25, Batch: 400/1265, Training Loss Error: 5.481, Training Time on 100 Batches: 41 seconds
Epoch: 1/25, Batch: 500/1265, Training Loss Error: 7.717, Training Time on 100 Batches: 59 seconds
Epoch: 1/25, Batch: 600/1265, Training Loss Error: 9.368, Training Time on 100 Batches: 61 seconds

macOS Mojave

Version 10.14.4

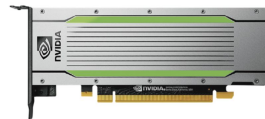
iMac (Retina 4K, 21.5-inch, 2017)

Processor 3 GHz Intel Core i5

Memory 8 GB 2400 MHz DDR4

Graphics Radeon Pro 555 2 GB

Serial Number C02TR0C5J1G5

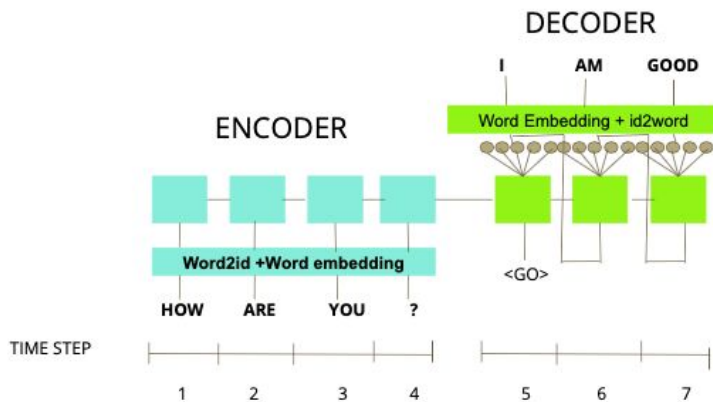


SPECIFICATIONS

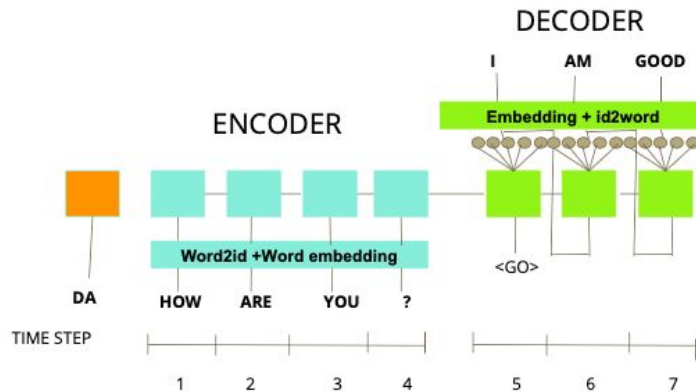
GPU Architecture	NVIDIA Turing
NVIDIA Turing Tensor Cores	320
NVIDIA CUDA® Cores	2,560
Single-Precision	8.1 TFLOPS
Mixed-Precision (FP16/FP32)	65 TFLOPS
INT8	130 TOPS
INT4	260 TOPS
GPU Memory	16 GB GDDR6 300 GB/sec
ECC	Yes
Interconnect Bandwidth	32 GB/sec

Models:

Model 1 : Sequence to sequence model



Model 2: Sequence to sequence model + dialogue act

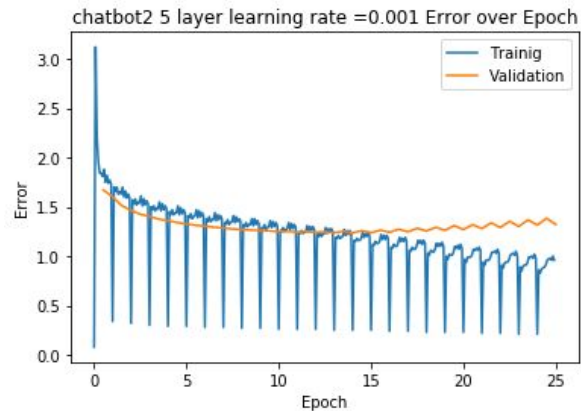
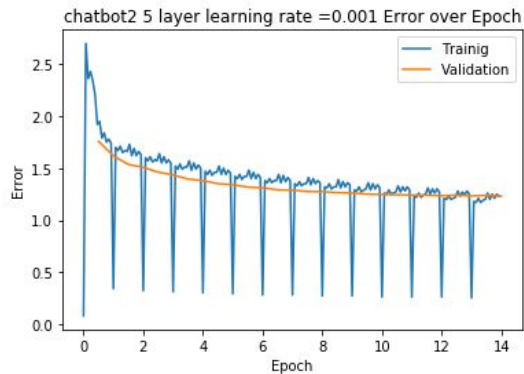
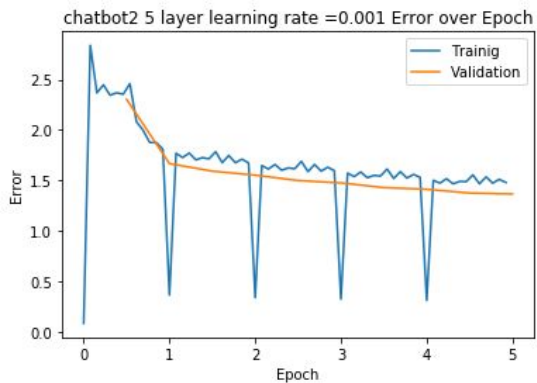
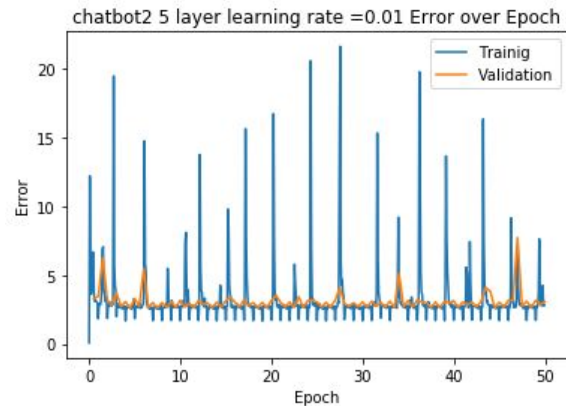
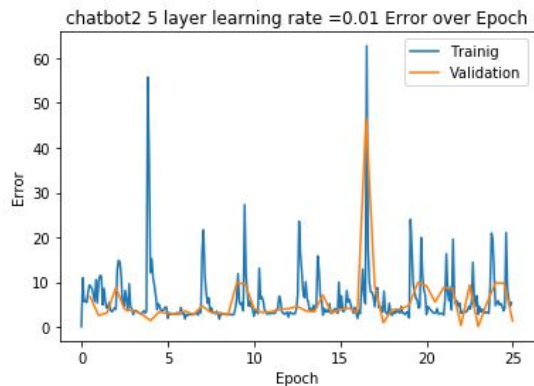
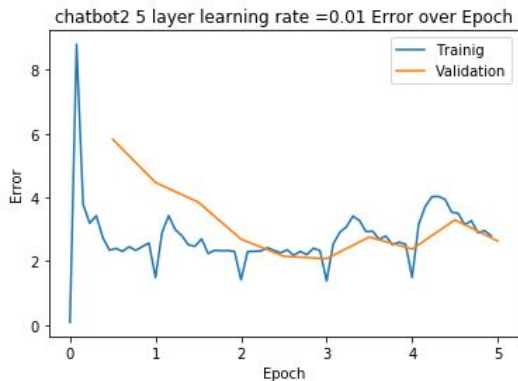


Experiments

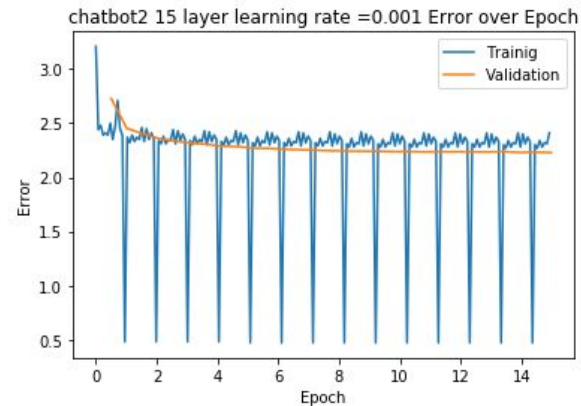
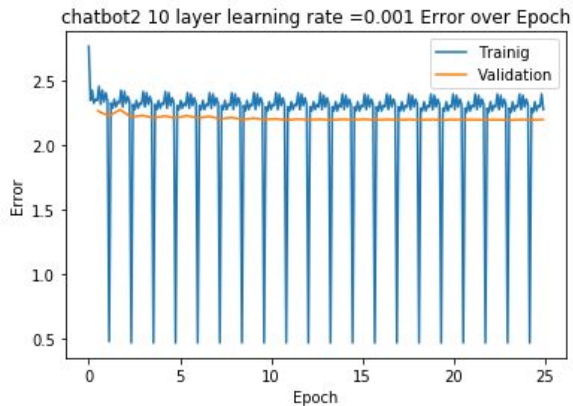
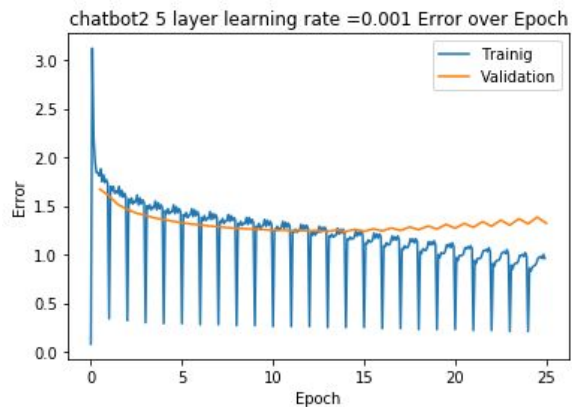
	Learning rate	Hidden layer	epoch
Chat bot 1	0.01	5	5
Chat bot 1	0.001	5	5
Chat bot 2	0.01	5	5
Chat bot 2	0.01	5	25
Chat bot 2	0.01	5	50
Chat bot 2	0.001	5	5
Chat bot 2	0.001	5	14

	Learning rate	Hidden layer	epoch
Chat bot 2	0.001	5	25
Chat bot 2	0.001	5	25
Chat bot 2	0.001	10	25
Chat bot 2	0.001	15	25
Chat bot 3	0.01	5	50
Chat bot 3	0.001	5	5

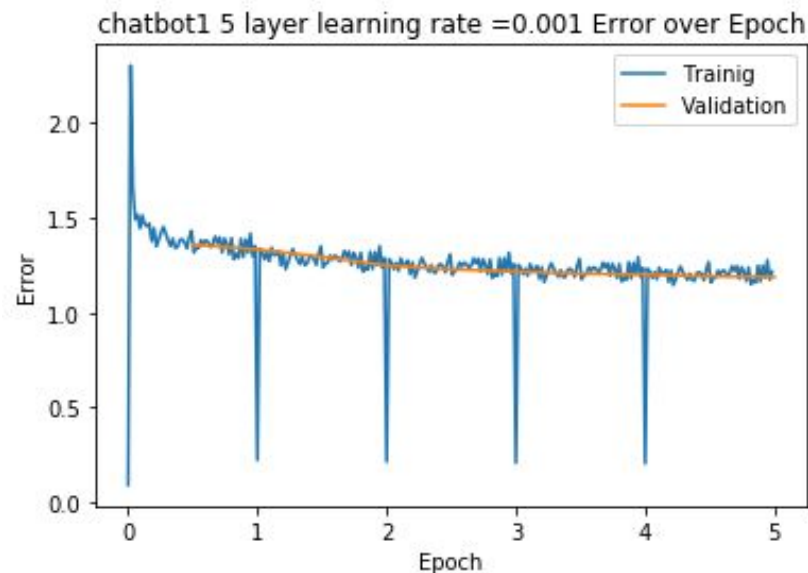
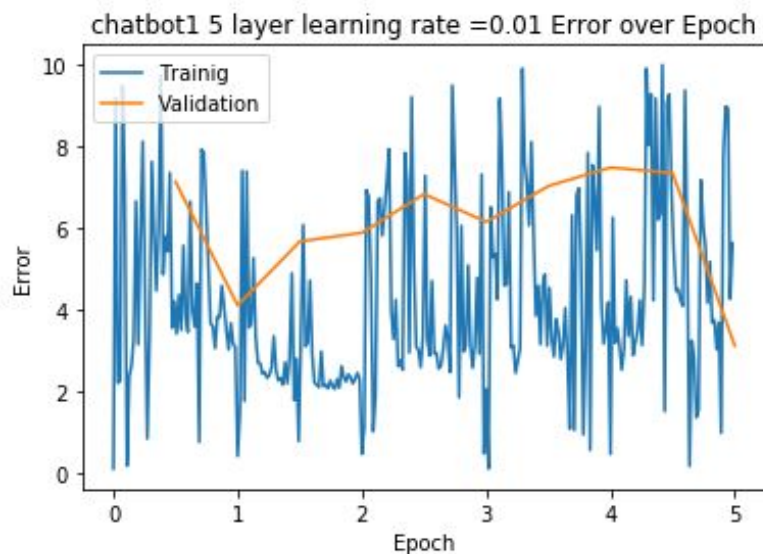
Chatbot 2 training result



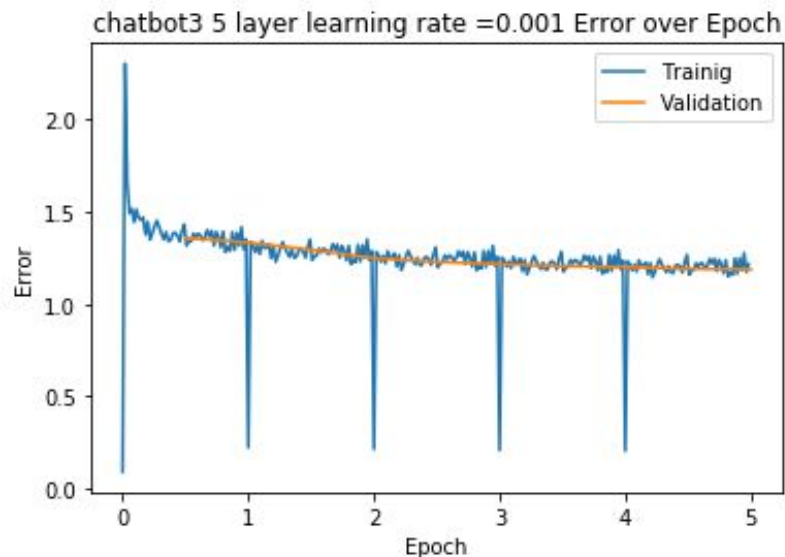
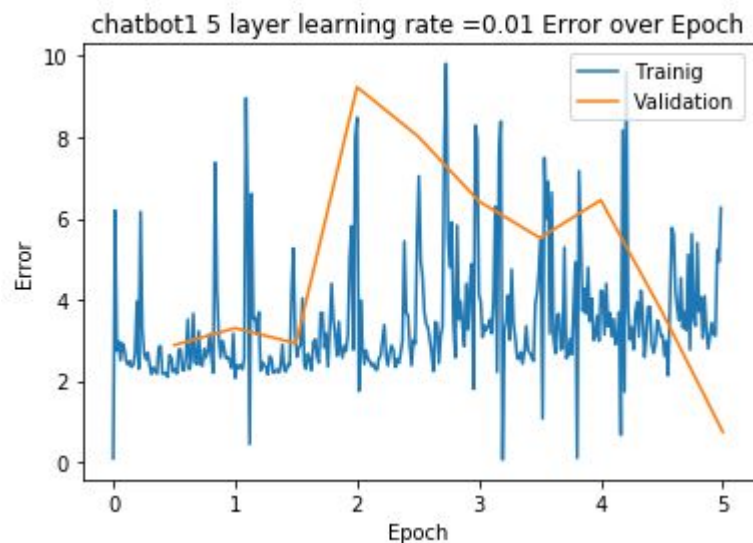
chatbot2 5,10,15 layer



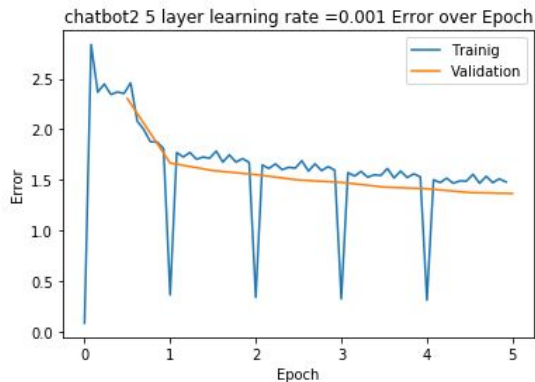
chatbot 1



chatbot 3



chatbot2 5epoch 0.001



Overfitting if: training loss << validation loss

Underfitting if: training loss >> validation loss

Just right if training loss ~ validation loss

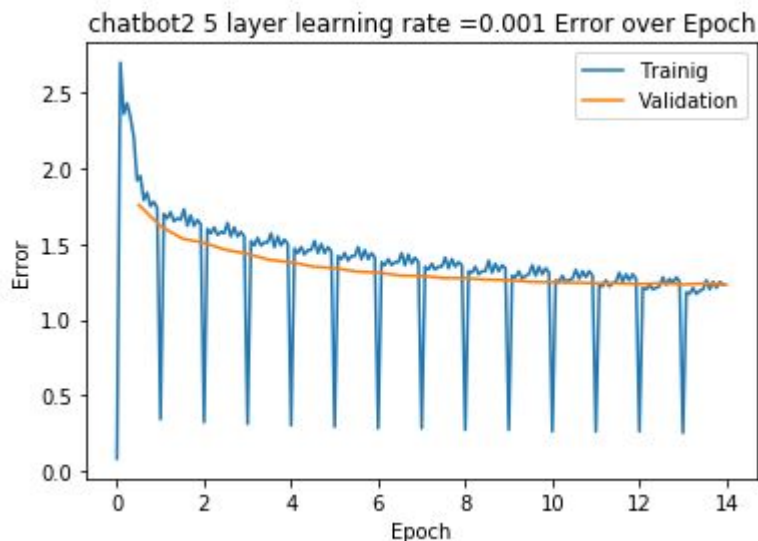
```
You: HI
ChatBot: I am glad , I am not sure of you ..
You: Where are you from?
ChatBot: I am glad ofout ..
You: Whatdo you do for a living?
ChatBot: I am glad , but I will be able to be aout ..
You: Can you let me know what time it is?
ChatBot: I am sorry , but I am not sure ..
You: Do you have time to go shopping tomorrow?
ChatBot: I am glad I will ..
You: When is your birthday?
ChatBot: I am sorry , but I am not sure of theout ..
You: Which one do you prefer, a dog or a cat?
ChatBot: I am sorry , but I am not sure of you ..
You: When is the next Olympic?
ChatBot: I am glad , but I am not sure of you ..
You: Can you recommend any country to go on a trip?
ChatBot: I am glad of you ..
You: How is the weather today?
ChatBot: I am sorry , but I am not sure of you ..
You: I miss my family.
ChatBot: I am sorry ..
You: Happy new year!|
ChatBot: I am glad . I am glad , I am not sure of you ..
You: I am hungry
ChatBot: I am glad , but I am not sure of theout ..
```

chatbot2 14 epoch

Overfitting if: training loss \ll validation loss

Underfitting if: training loss \gg validation loss

Just right if training loss \sim validation loss



You: HI
ChatBot: conversation sets awful pardon romance brilliant parcel discuss lisa efficient typicalout swimming.

You: How are you
ChatBot: conversation sets wilson eaten own conversation level typicaloutout swimming.

You: conversation
ChatBot: conversation sets awful romance maintain register swimming.

You: Where are you from?
ChatBot: conversation sets wilson eaten own conversation however already hands romance swimming.

You: What do you do for a living?
ChatBot: conversation sets awful romance level pleaseout dance discuss swimming.

You: Can you let me know what time it is?
ChatBot: conversation sets wilson eaten own conversation however level lisa cameras lisa coach western swimming.

You: Do you have time to go shopping tomorrow?
ChatBot: conversation sets wilson eaten own conversation however level lisa cameras lisa coachout swimming.

You: When is your birthday?
ChatBot: conversation sets awful romance maintain register swimming.

You: Which one do you prefer, a dog or a cat?
ChatBot: conversation sets awful romance maintain register swimming.

You: When is the next Olympic?
ChatBot: conversation sets awful romance maintain register swimming.

You: Can you recommend any country to go on a trip?
ChatBot: conversation sets awful romance maintain register swimming.

You: How is the weather today?
ChatBot: conversation sets awful romance maintain register swimming.

You: : I miss my family.
ChatBot: conversation sets awful romance maintain register swimming.

You: i am so happy today
ChatBot: conversation sets awful romance maintain register swimming.

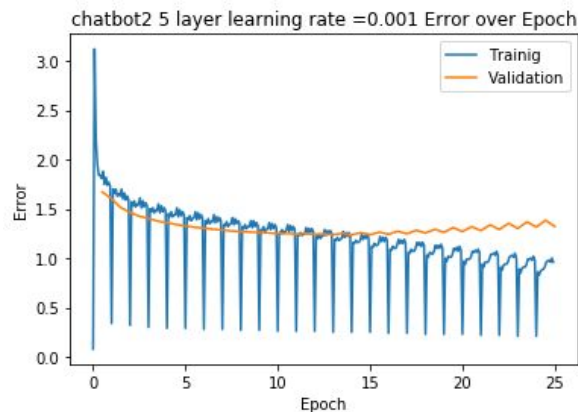
You: i am so happy today
ChatBot: conversation sets awful romance maintain register swimming.

chatbot2 25epoch

Overfitting if: training loss << validation loss

Underfitting if: training loss >> validation loss

Just right if training loss ~ validation loss



You: happy

ChatBot: cancer demand allowed phone letters speed cancer pay illegal definitely entertainment mexico definitely bungee speed.

You: happy new year

ChatBot: cancer demand charges songs certainly surprise wear next speed.

You: Where are you from?

ChatBot: cancer demand charges songs exciting hong speed.

You: What do you do for a living?

ChatBot: cancer demand charges songs magazine palace speed cancer ! cheap santa tend safe hospital nice speed.

You: Can you let me know what time it is?

ChatBot: cancer demand allowed phone independent cancer demand cheap police speed.

You: Do you have time to go shopping tomorrow?

ChatBot: cancer demand allowed phone independent cancer digital certainly safe towards :: definitelyout speed.

You: When is your birthday?

ChatBot: cancer demand allowed phone independent cancer digital certainly safe towards mexico surprise library speed.

You: Which one do you prefer, a dog or a cat?

ChatBot: cancer demand allowed phone independent cancer demand cheap police whom cancer judge speed.

You: When is the next Olympic?

ChatBot: cancer demand allowed phone independent cancer digital certainly safe towards mexico definitelyout speed.

You: Can you recommend any country to go on a trip?

ChatBot: cancer demand allowed phone independent cancer digital certainly safe towards mexico surpriseout speed.

You: How is the weather today?

ChatBot: cancer demand allowed phone independent cancer digital certainly safe towards mexico songs speed.

You: I miss my family.

ChatBot: cancer demand charges songs certainly surprise wear next speed.

You: cancer

ChatBot: cancer demand allowed phone independent cancer demand cheap police whom cancer judge hospital songs speed.

You: why

ChatBot: cancer demand charges songs guy speed.

Goal and motivation of the project

- To learn how to implement sequence to sequence model to build a chat bot
- To explore the effect of dataset on chat bot performance
 - Cornell movie corpus
 - Daily Dialogue
- Use statistics method instead of template method, which is a typical way to build a chat bot

Epoch on tensorflow

- An epoch is a full iteration over all samples.
- The number of epochs is how many times the algorithm is going to run.
- With just a few epochs you can reach only a local minimum, but with more epochs, you can reach a global minimum or at least a better local minimum.

Result: chat bot 1 - cornell movie dataset

5 epoch, learning rate: 0.001

You: hi

ChatBot: I am not going to get you to theout.

You: how are you?

ChatBot: I am not going to beout.

You: Where are you from?

ChatBot: I am not going to beout.

You: What do you do for a living?

ChatBot: I am not going to beout.

You: Can you let me know what time it is?

ChatBot: I am not going to get out of here.

You: Do you have time to go shopping tomorrow?

ChatBot: I am not going to get aout.

You: When is your birthday?

ChatBot: I am not going to get you.

You: Which one do you prefer, a dog or a cat?

ChatBot: I am not going to get you to theout.

You: When is the next Olympic?

ChatBot: I am not going to get you.

You: Can you recommend any country to go on a trip?

ChatBot: I am not going to beout.

You: How is the weather today?

ChatBot: I am not going to get it out of my life.

You: I miss my family.

ChatBot: I am not going to beout.

You: I am not sure

ChatBot: I am not going to get aout.

You: I am happy

ChatBot: I am not going to get you to theout.

You: why not?

ChatBot: I am not going to beout.

You: weiting

ChatBot: I am not going to beout.

You: good night

ChatBot: I am not going to beout.

Result: chat bot 2 - daily dialogue, 5 epoch

You: HI

ChatBot: I am glad , I am not sure of you ..

You: Where are you from?

ChatBot: I am glad ofout ..

You: What do you do for a living?

ChatBot: I am glad , but I will be able to be aout ..

You: Can you let me know what time it is?

ChatBot: I am sorry , but I am not sure ..

You: Do you have time to go shopping tomorrow?

ChatBot: I am glad I will ..

You: When is your birthday?

ChatBot: I am sorry , but I am not sure of theout ..

You: Which one do you prefer, a dog or a cat?

ChatBot: I am sorry , but I am not sure of you ..

You: When is the next Olympic?

ChatBot: I am glad , but I am not sure of you ..

You: Can you recommend any country to go on a trip?

ChatBot: I am glad of you ..

You: How is the weather today?

ChatBot: I am sorry , but I am not sure of you ..

You: I miss my family.

ChatBot: I am sorry ..

You: Happy new year!

ChatBot: I am glad . I am glad , I am not sure of you ..

You: I am hungry

ChatBot: I am glad , but I am not sure of theout ..

Result: chat bot 2, 25 epoch

You: hi

ChatBot: cancer demand charges songs exciting hong speed.

You: how are you

ChatBot: cancer demand charges songs magazine palace speed cancer ! cheap santa tend safe haven surpriseout speed.

You: happy

ChatBot: cancer demand allowed phone letters speed cancer pay illegal definitely entertainment mexico definitely bungee speed.

You: happy new year

ChatBot: cancer demand charges songs certainly surprise wear next speed.

You: Where are you from?

ChatBot: cancer demand charges songs exciting hong speed.

Result: chat bot 2, 25 epoch

You: What do you do for a living?

ChatBot: cancer demand charges songs magazine palace speed cancer ! cheap santa tend safe hospital nice speed.

You: Can you let me know what time it is?

ChatBot: cancer demand allowed phone independent cancer demand cheap police speed.

You: Do you have time to go shopping tomorrow?

ChatBot: cancer demand allowed phone independent cancer digital certainly safe towards :: definitelyout speed.

You: When is your birthday?

ChatBot: cancer demand allowed phone independent cancer digital certainly safe towards mexico surprise library speed.

You: Which one do you prefer, a dog or a cat?

ChatBot: cancer demand allowed phone independent cancer demand cheap police whom cancer judge speed.

Result: chat bot 2, 14 epoch

You: HI
ChatBot: conversation sets awful pardon romance brilliant parcel discuss lisa efficient typicalout swimming.

You: How are you
ChatBot: conversation sets wilson eaten own conversation level typicaloutout swimming.

You: conversation
ChatBot: conversation sets awful romance maintain register swimming.

You: Where are you from?
ChatBot: conversation sets wilson eaten own conversation however already hands romance swimming.

You: What do you do for a living?
ChatBot: conversation sets awful romance level pleaseout dance discuss swimming.

You: Can you let me know what time it is?
ChatBot: conversation sets wilson eaten own conversation however level lisa cameras lisa coach western swimming.

You: Do you have time to go shopping tomorrow?
ChatBot: conversation sets wilson eaten own conversation however level lisa cameras lisa coachout swimming.

You: When is your birthday?
ChatBot: conversation sets awful romance maintain register swimming.

You: Which one do you prefer, a dog or a cat?
ChatBot: conversation sets awful romance maintain register swimming.

You: When is the next Olympic?
ChatBot: conversation sets awful romance maintain register swimming.

You: Can you recommend any country to go on a trip?
ChatBot: conversation sets awful romance maintain register swimming.

You: How is the weather today?
ChatBot: conversation sets awful romance maintain register swimming.

You: : I miss my family.
ChatBot: conversation sets awful romance maintain register swimming.

You: i am so happy today
ChatBot: conversation sets awful romance maintain register swimming.

Result: chat bot 2, 25 epoch

You: When is the next Olympic?

ChatBot: cancer demand allowed phone independent cancer digital certainly safe towards mexico definitelyout speed.

You: Can you recommend any country to go on a trip?

ChatBot: cancer demand allowed phone independent cancer digital certainly safe towards mexico surpriseout speed.

You: How is the weather today?

ChatBot: cancer demand allowed phone independent cancer digital certainly safe towards mexico songs speed.

You: I miss my family.

ChatBot: cancer demand charges songs certainly surprise wear next speed.

Result: chat bot 3 - mixed dataset

You: hi

ChatBot: I am not sure I am not going to be aout.

You: how are you

ChatBot: I am not sure I am not sure.

You: Where are you from?

ChatBot: I am not sure.

You: What do you do for a living?

ChatBot: I am not sure I am not going to be aout.

You: Can you let me know what time it is?

ChatBot: I am not sure.

You: Do you have time to go shopping tomorrow?

ChatBot: I am not sure I am a littleout.

You: When is your birthday?

ChatBot: I am not sure I am not going to be able to be aout.

You: Which one do you prefer, a dog or a cat?

ChatBot: I am not sure I am not going to be able to be aout.

You: When is the next Olympic?

ChatBot: I am not sure.

You: Can you recommend any country to go on a trip?

ChatBot: I am not sure.

You: How is the weather today?

ChatBot: I am not sure I am not going to be aout.

You: I miss my family.

ChatBot: I am not sure.

You: have a good day

ChatBot: I am not sure I am not sure I am not going to be able to be aout.

You: happy new year

ChatBot: I am not sure I am not going to be able to be able to be aout.

Human Evaluation

- Considering that the aim of our project is to make an open-ended generation, which does not target deriving a correct answer, we conduct only human evaluation by asking three people. We characterize (1) grammaticality and (2) naturalness (3) interestingness for some samples of the four different models.
- Each criteria was evaluated compared to other models.
- Grammaticality: scale 0-5, evaluates grammatical errors (subject verb agreement, random words or a full sentence)
- Naturalness: scale 0-5, evaluates whether a response from a chat bot is similar to natural dialogue
- Interestingness: scale 0-5, evaluates whether the response from a chat bot evokes a person to continue talking to it.

Human Evaluation

Average score for each criteria

	Grammaticality	Naturalness	Interestingness
Chat bot 1, 5 epoch,	3.3	2	1.7
Chat bot 2, 5 epoch	4	2.7	2.7
Chat bot 2, 25 epoch	1	1	2.3
Chat bot 2, 14 epoch	1	1	1.7
Chat bot 3, mixed dataset	3.7	4	2

Human Evaluation

	Grammaticality	Naturalness	Interestingness
<p>Chat bot 1, 5 epoch</p> <p>(You: Where are you from?</p> <p>ChatBot: I am not going to beout.</p> <p>You: Can you let me know what time it is?</p> <p>ChatBot: I am not going to get out of here.</p> <p>You: When is your birthday?</p> <p>ChatBot: I am not going to get you.)</p>	<p>3, At least it generates a full sentence. However, it generates a random word, such as <i>theout</i>, <i>beout</i>, <i>about</i>. Still it shows subject verb agreement. Sometimes, it generates a grammatical full sentence.</p>	<p>2, It keeps saying the same thing regardless of what question it was asked.</p>	<p>2, It is not interesting very much as it keeps saying the same thing.</p>

Human Evaluation

	Grammaticality	Naturalness	Interestingness
<p>Chat bot 2, 5 epoch (You: Where are you from? ChatBot: I am glad ofout .. You: What do you do for a living? ChatBot: I am glad , but I will be able to be aout .. You: Can you let me know what time it is? ChatBot:I am sorry , but I am not sure ..)</p>	<p>4, It generates a full sentence. Sometimes it produces a random word, such as <i>aout</i>, <i>ofout</i>, but it happens less than other chat bots. Also, it shows subject verb agreement. It produces a future tense, and a comma and a punctuation at the right place.</p>	<p>2, It also keeps saying the same thing regardless of what question it was asked.</p>	<p>2, It is not interesting very much as it keeps saying the same thing. There seems to be no point continuing talking to it.</p>

Human Evaluation

	Grammaticality	Naturalness	Interestingness
<p>Chat bot 2, 25 epoch</p> <p>(You: Do you have time to go shopping tomorrow?</p> <p>ChatBot: cancer demand allowed phone independent cancer digital certainly safe towards :: definitelyout speed.</p>	<p>1, It does not generate a full sentence, and spread only words. It also generates meaningless words, such as <i>definitelyout, surpriseout</i>.</p>	<p>1, It keeps spreading the same words which are out of order. And it produces an exclamation mark and colon at a random place.</p>	<p>2, It also keeps producing the same line of words, but it looks funny to see it spreading some words no matter what question was asked.</p>

Human Evaluation

	Grammaticality	Naturalness	Interestingness
Chat bot 2, 14 epoch (You: Do you have time to go shopping tomorrow? Chatbot: conversation sets wilson eaten own conversation however level lisa cameras lisa coacchout swimming.	1, It does not generate a full sentence, but produces some strings of words and uses a punctuation mark. It also generates random words, such as <i>typicaloutout</i> , <i>coachout</i> .	1, It only produces some strings of words, not sentences.	2, It also keeps producing almost the same strings of words. There seems to be no point to continue talking to it.

Human Evaluation

	Grammaticality	Naturalness	Interestingness
<p>Chat bot 3, mixed dataset</p> <p>(You: Can you let me know what time it is?</p> <p>ChatBot: I am not sure.</p> <p>You: When is the next Olympic?</p> <p>ChatBot: I am not sure.</p> <p>You: Can you recommend any country to go on a trip?</p> <p>ChatBot: I am not sure.</p>	<p>3, It generates a full sentence. But sometimes it misses a punctuation mark. Also, it sometimes generates a random word, such as littleout.</p>	<p>4, it repeats <i>I am not sure</i>, but it makes sense for some questions.</p>	<p>2, it is not that interesting to continue talking to it as it keeps saying the same thing.</p>

Human Evaluation

- In terms of grammaticality, the chat bot 2, 5 epoch outperforms the other ones.
- In terms of naturalness, the chat bot 3 with mixed dataset outperforms the other ones.
- In terms of interestingness, all of them perform similar as they repeat the same thing. However, the chat bot 2, 5 epoch got the highest score.

Future work

- Get Enough computing power

Optimize hyperparameters to find the best weight matrix and understand the model

- **epochs = 5 10 14 25 50**
- batch_size = 32
- rnn_size = 102
- **num_layers = 3 5 10 15**
- **learning_rate = 0.01 -----0.001**
- learning_rate_decay = 0.9
- min_learning_rate = 0.1
- keep_probability = 0.5