**Sequence to Sequence Supervised Model**

**Introduction**

Firstly Seq2Seq model is very popular. Recently there are numerous example of making chatbot using Seq2Seq Model. [See Reference]

**Script Name:** Seq2Seq.pyin https://github.com/Iftekhar-mobin/AI-system-.git

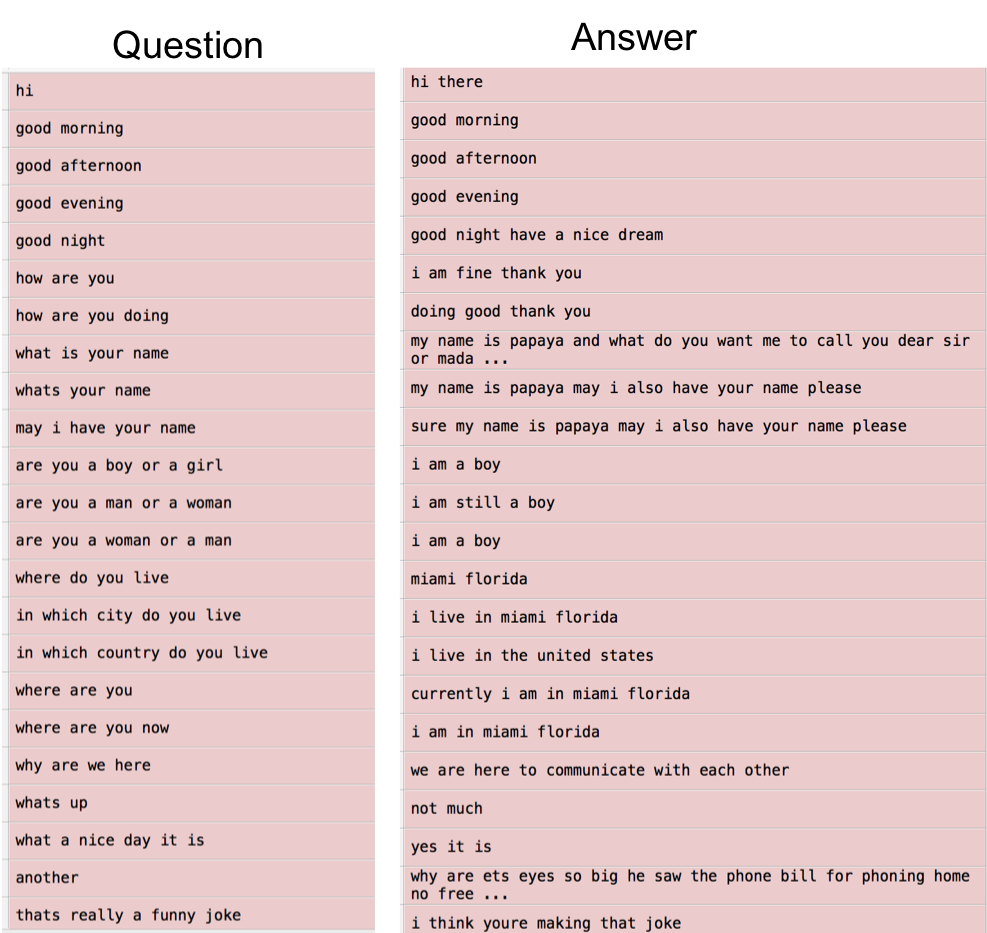
**Dataset Format**

Hi \t Hi there \n

Good Morning \t good morning \n

How are you \t I am fine thank you \n

What is your name \t My name is papaya …. \n



**Problem**

1. Dataset does not comes in that format

2. Difficult to identify start and end. Does not contain \n and \t

3. We could have

1. one question and multiple lines of answer,
2. multiple lines of question but answer is only one line or may be less than a line

**Solution**

1. We can slice the whole document character by character. An Example of document is

“How are you. I am fine thank you.”

converted to

‘H’, ‘o’, ‘w’, ‘a’, ‘r’, ‘e’, ‘y’, ‘o’, ‘u’, ‘I’, ‘a’, ‘m’, ’f’, ’i’, ‘n’, ‘e’, ‘t’, ‘h’, ‘a’, ‘n’, ‘k’ ‘y’, ‘o’, ‘u’

2. Now, Total length of characters is 24. chop the document size with length 10. Adding \n at the end.

‘H’, ‘o’, ‘w’, ‘a’, ‘r’, ‘e’, ‘y’, ‘o’, ‘u’, ‘I’, \n

‘a’, ‘m’, ’f’, ’i’, ‘n’, ‘e’, ‘t’, ‘h’, ‘a’, ‘n’, \n

‘k’ ‘y’, ‘o’, ‘u’ \n

3. Split from the middle at put \t

‘H’, ‘o’, ‘w’, ‘a’, ‘r’, \t ‘e’, ‘y’, ‘o’, ‘u’, ‘I’, \n

‘a’, ‘m’, ’f’, ’i’, ‘n’, ‘e’, \t ‘t’, ‘h’, ‘a’, ‘n’, \n

‘k’ ‘y’, \t ‘o’, ‘u’ \n

4. Feed this content to a Encoder -Decoder Seq to Seq model.

**How to make it more efficient**

1. Using word by word tokenization

2. Marking Question Start and End position

3. Marking Answer Start and End position

**Use different Learning Model**

Lets assume the document is “How are you. I am fine thank you.”

Now, to convert it to a supervised structure → We can follow a Timeseries forecasting Technique.

['How', 'are', 'you', 'I', 'am', 'fine', 'thank', 'you', 'Will', 'you', 'go', 'to', 'Osaka', 'Today']→ will become

****

**New Model Structure**

Col1(C-3) Col1(C-2) Col1(C-1) Col1(C) → → Col1(C+1) Col1(C+2) Col1(C+3) Col1(C+4)

How are you I → → am fine thank you

Question: How are you I

Answer: am fine thank you

So, from the given input “How are you I” we are saying try to predict “am fine thank you”

**More complicated model using Multiple variables**

So, If we want to give input “How are you” we have to use multiple variables

Here Variables means ‘Col1’

1. See the table structure only Col1 is used
2. Similarly Col2, Col3 and more variable can be used
3. Col1 will start from “How are … “
4. Col2 will start from “are you …. “
5. Col3 will start from “you I… “

Still Question remains what will be exact size of “Col” variable length because Question and answer may overlap with each other.

**For more details check my Github repo**

Mutistep Mutivariate LSTM Forecasting based sequence prediction.py script in https://github.com/Iftekhar-mobin/AI-system-.git

**References**

[https://github.com/adeshpande3/Facebook-Messenger-Bot/blob/master/Seq2Seq.py](https://hackernoon.com/implementing-a-sequence-to-sequence-model-45a6133958ca)

https://github.com/AtaaEddin/Turkish-ChatBot

https://github.com/changwookjun/ChatBot\_seq2seq\_extend

https://github.com/fendouai/Awesome-Chatbot

https://github.com/fmehmetun/tf\_encdec\_seq2seq

https://github.com/higepon/tensorflow\_seq2seq\_chatbot

https://github.com/lang-ai/chatbots2s

https://github.com/Marsan-Ma-zz/tf\_chatbot\_seq2seq\_antilm

https://github.com/OliverEdholm/Tensorflow-Easy-Seq2Seq

https://github.com/partoftheorigin/mahabharatha\_chatbot\_rnn\_seq2seq

https://github.com/purelyvivid/Schedule\_Sampling\_Seq2Seq\_Movie\_Script\_Chatbot\_PyTorch

https://github.com/rand0wn/Game\_Of\_Thrones\_Chatbot\_Seq2Seq

https://github.com/Ravi-Jay/Pytorch\_Seq2seq

https://github.com/REDFOX1899/Chatbot

https://github.com/sayhitosandy/Chatbot

https://github.com/sea-boat/seq2seq\_chatbot

https://github.com/xwhan/dialogue\_generation

https://github.com/zhaoyingjun/chatbot

https://hackernoon.com/implementing-a-sequence-to-sequence-model-45a6133958ca