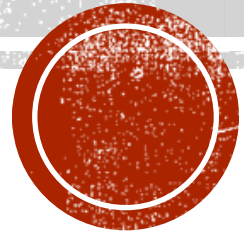
CONVERSATION ROBOT

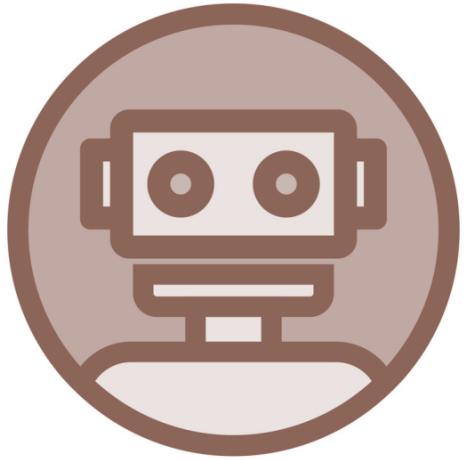
FINAL PRESENTATION



Zhiying Li

zl2697

- Data Processing
- Model Construction
- Model Experiments
- User Interface



CORNELL MOVIE DIALOGUE CORPUS

- 220,579 conversational exchanges between 10,292 pairs of movie characters.
- The dialogues are mainly extracted from movies
- There are pairs of the dialogue produced by the dataset.

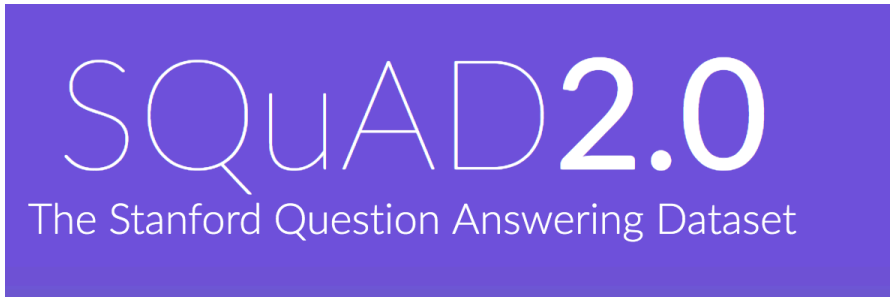
www.cs.cornell.edu

Cornell Movie-Dialogs Corpus

```
L666383 +++$+++ u9027 +++$+++ m616 +++$+++ CHELMSFORD +++$+++ What's that strange name the newspaper chap's called?
L666252 +++$+++ u9027 +++$+++ m616 +++$+++ CHELMSFORD +++$+++ But will they make good use of them?
L666251 +++$+++ u9029 +++$+++ m616 +++$+++ CREALOCK +++$+++ Yes. I see you've issued each of them with a Martini Henry Carbine. Our quota for Native
    contingencies: one rifle to ten men and only five rounds per rifle.
L666502 +++$+++ u9033 +++$+++ m616 +++$+++ STUART SMITH +++$+++ Right.  Bombardier, to me please.
L666501 +++$+++ u9028 +++$+++ m616 +++$+++ COGHILL +++$+++ Well, fed or hungry, Pulleine wants them in position immediately. .
L666500 +++$+++ u9033 +++$+++ m616 +++$+++ STUART SMITH +++$+++ Well, my horses are feeding, as you may observe, Mr Coghill. It'll take a little while.
L666499 +++$+++ u9028 +++$+++ m616 +++$+++ COGHILL +++$+++ How quickly can you move your artillery forward?
L666498 +++$+++ u9033 +++$+++ m616 +++$+++ STUART SMITH +++$+++ Yes.
L666497 +++$+++ u9028 +++$+++ m616 +++$+++ COGHILL +++$+++ Stuart?
L666576 +++$+++ u9031 +++$+++ m616 +++$+++ MELVILL +++$+++ Keep steady. You're the best shots of the Twenty-Fourth. You bunch of heathens, do it
L666575 +++$+++ u9028 +++$+++ m616 +++$+++ COGHILL +++$+++ Choose your targets men. That's right Watch those markers. 55
L666327 +++$+++ u9031 +++$+++ m616 +++$+++ MELVILL +++$+++ ft could be you flatter yourself CoghilL It's that odd eye.
L666326 +++$+++ u9028 +++$+++ m616 +++$+++ COGHILL +++$+++ Well that one. The one who keeps looking at me.
L666325 +++$+++ u9031 +++$+++ m616 +++$+++ MELVILL +++$+++ Which one?
L666324 +++$+++ u9028 +++$+++ m616 +++$+++ COGHILL +++$+++ Do you think she might be interested in  someone?
L666264 +++$+++ u9031 +++$+++ m616 +++$+++ MELVILL +++$+++ Well that's typical of Her Majesty's army. Appoint an engineer to do a soldier's work.
L666263 +++$+++ u9028 +++$+++ m616 +++$+++ COGHILL +++$+++ Um. There are rumours that my Lord Chelmsford intends to make Durnford Second in Command.
L666262 +++$+++ u9031 +++$+++ m616 +++$+++ MELVILL +++$+++ Lighting COGHILL' 5 cigar: Our good Colonel Dumford scored quite a coup with the Sikali Horse.
L666522 +++$+++ u9034 +++$+++ m616 +++$+++ VEREKER +++$+++ So far only their scouts. But we have had reports of a small Impi farther north, over there.
L666521 +++$+++ u9030 +++$+++ m616 +++$+++ DURNFORD +++$+++ And I assure you, you do not In fact I'd be obliged for your best advice. What have your scouts seen?
L666520 +++$+++ u9034 +++$+++ m616 +++$+++ VEREKER +++$+++ Well I assure you, Sir, I have no desire to create difficulties. 45
L666372 +++$+++ u9034 +++$+++ m616 +++$+++ VEREKER +++$+++ I think Chelmsford wants a good man on the border Why he fears a flanking attack and requires a steady
    Commander in reserve.
L666371 +++$+++ u9030 +++$+++ m616 +++$+++ DURNFORD +++$+++ Lord Chelmsford seems to want me to stay back with my Basutos.
L666370 +++$+++ u9034 +++$+++ m616 +++$+++ VEREKER +++$+++ I'm to take the Sikali with the main column to the river
L666369 +++$+++ u9030 +++$+++ m616 +++$+++ DURNFORD +++$+++ Your orders, Mr Vereker?
L666257 +++$+++ u9030 +++$+++ m616 +++$+++ DURNFORD +++$+++ Good ones, yes, Mr Vereker. Gentlemen who can ride and shoot
L666256 +++$+++ u9034 +++$+++ m616 +++$+++ VEREKER +++$+++ Colonel Durnford... William Vereker. I hear you 've been seeking Officers?
```

SQUAD DATASET

- **S**tanford **Qu**estion **A**nswering **D**ataset (SQuAD) is a reading comprehension dataset, consisting of questions posed by crowd workers on a set of Wikipedia articles, where the answer to every question is a segment of text, or *span*, from the corresponding reading passage, or the question might be unanswerable.



SQUAD DATASET

Explore the dataset

- The dataset is hierarchal
- Dataset: list contains 442 articles

For each article, there are many paragraphs and for each paragraphs, there is a context and many questions and answers. We extract the pairs of questions and answers and we do not need the context. Some of the questions has no answer, and there is a label *is_impossible* indicating whether the question has an answer.

```
Beyoncé Giselle Knowles-Carter (/biːˈjɒnseɪ/ bee-YON-say) (born September 4, 1981)
is an American singer, songwriter, record producer and actress. Born and raised in
Houston, Texas, she performed in various singing and dancing competitions as a
child, and rose to fame in the late 1990s as lead singer of R&B girl-group
Destiny's Child. Managed by her father, Mathew Knowles, the group became one of the
world's best-selling girl groups of all time. Their hiatus saw the release of
Beyoncé's debut album, Dangerously in Love (2003), which established her as a solo
artist worldwide, earned five Grammy Awards and featured the Billboard Hot 100
number-one singles "Crazy in Love" and "Baby Boy".
```

answers	list	1	[{'text':'in the late 1990s', 'answer_start':269}]
id	str	1	56be85543aeaaa14008c9063
is_impossible	bool	1	False
question	str	1	When did Beyonce start becoming popular?

SQUAD DATASET

- 1. Read the data from json
- 2. extract the pairs of questions and answers if the flag is_impossible is false

```
os.chdir("/Users/lizhiying/Desktop/Big Data Analytics/LI-master")
with open("data/train-v2.0.json") as f:
    dataset_json = json.load(f)
    dataset = dataset_json['data']

input_set = []
output_set = []


for article in dataset:
    for p in article['paragraphs']:
        for qa in p['qas']:
            if qa['is_impossible'] == False:
                input_set.append(qa['question'])
                output_set.append(qa['answers'][0]['text'])
```

```
In [58]: print(input_set[100])
How many weeks did their single "Independent Women Part I" stay on top?

In [59]: print(output_set[100])
eleven
```

WIKIQA DATASET

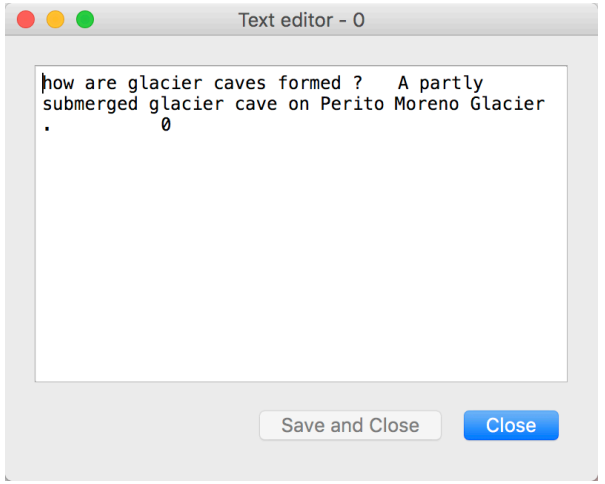
- The WikiQA corpus is a new publicly available set of question and sentence pairs, collected and annotated for research on open-domain question answering. In order to reflect the true information need of general users, WikiQA used Bing query logs as the question source. Each question is linked to a Wikipedia page that potentially has the answer.
- With included 3,047 questions and 29,258 sentences in the dataset, where 1,473 sentences were labeled as answer sentences to their corresponding questions.
- **We just choose the questions with the right sentences.**

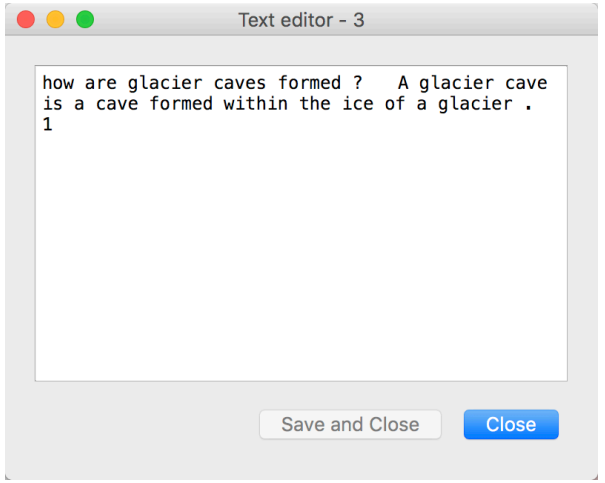
WIKIQA DATASET

- For the data which is the correct answer, there is a symbol 1, otherwise that is 0.

- We extract the data which the answer to the question is correct.

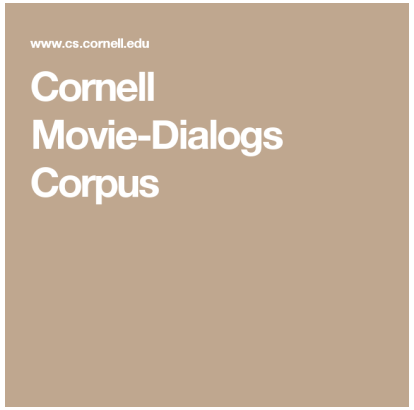
```
25
26 with open("data/WikiQA-train.txt") as f:
27     q = f.readlines()
28
29 all_set = []
30
31 for item in q:
32     if item[-2] == "1":
33         all_set.append(item)
34
35
36
37 for item in all_set:
38     item_l = item.split('\t')
39     input_set.append(item_l[0])
40     output_set.append(item_l[1])
41
42
```



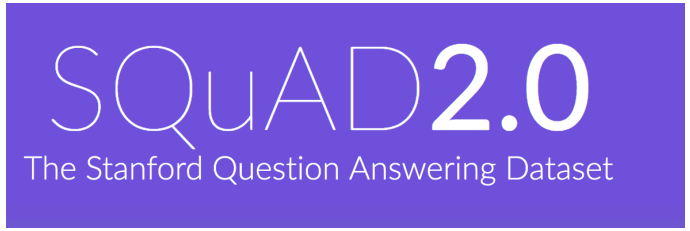


DATA PROCESSING

- We merge the three dataset together and then tokenize the data



+



+

DATA PROCESSING

Tokenization:

We control the length of the sentence no longer 30

Lower the case of each word

Split each word by “([.,!?\"'-<>:;)(])”

Turn the word to the number(i.e. the index in the dic)

The total number of the dataset is 227219

The total vocabulary is 42690

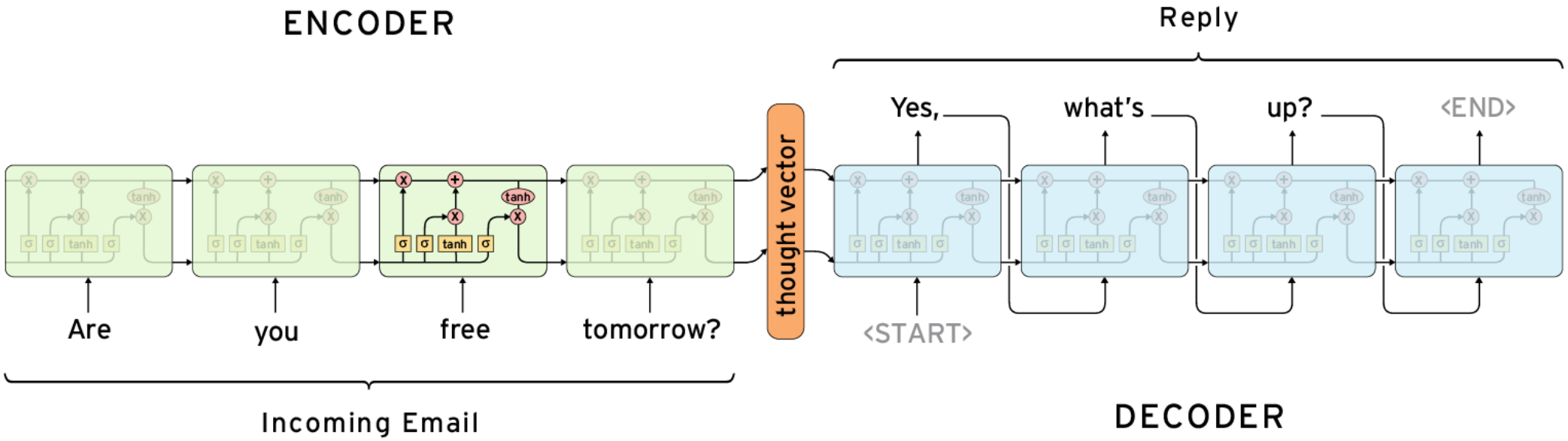
```
def load_data(input_path,target_path):

    lines_input = open(input_path, 'rt', encoding='utf8').read().split('\n')
    lines_target = open(target_path, 'rt', encoding='utf8').read().split('\n')
    input_texts = []
    target_texts = []
    prev_words = []

    input_counter = Counter()
    target_counter = Counter()



    for line in lines_input:

        next_words = [w.lower() for w in nltk.word_tokenize(line)]
        input_texts.append(next_words)
        #if len(next_words) > MAX_TARGET_SEQ_LENGTH:
            #next_words = next_words[0:MAX_TARGET_SEQ_LENGTH]

        for w in prev_words:
            input_counter[w] += 1
        prev_words = next_words


    prev_words = []

    for line in lines_target:
        next_words = [w.lower() for w in nltk.word_tokenize(line)]
        target_words = next_words
        target_words.insert(0, 'START')
        target_words.append('END')
        target_texts.append(next_words)

        for w in prev_words:
            target_counter[w] += 1
        prev_words = next_words

    print(input_texts[1000],target_texts[1000])

    input_texts_new = []
    target_texts_new = []
    for i in range(len(input_texts)-1):
        if len(input_texts[i]) < MAX_INPUT_SEQ_LENGTH and len(target_texts[i])
            input_texts_new.append(input_texts[i])
            target_texts_new.append(target_texts[i])

    return input_texts_new, target_texts_new, input_counter, target_counter
```

MODEL DESIGN

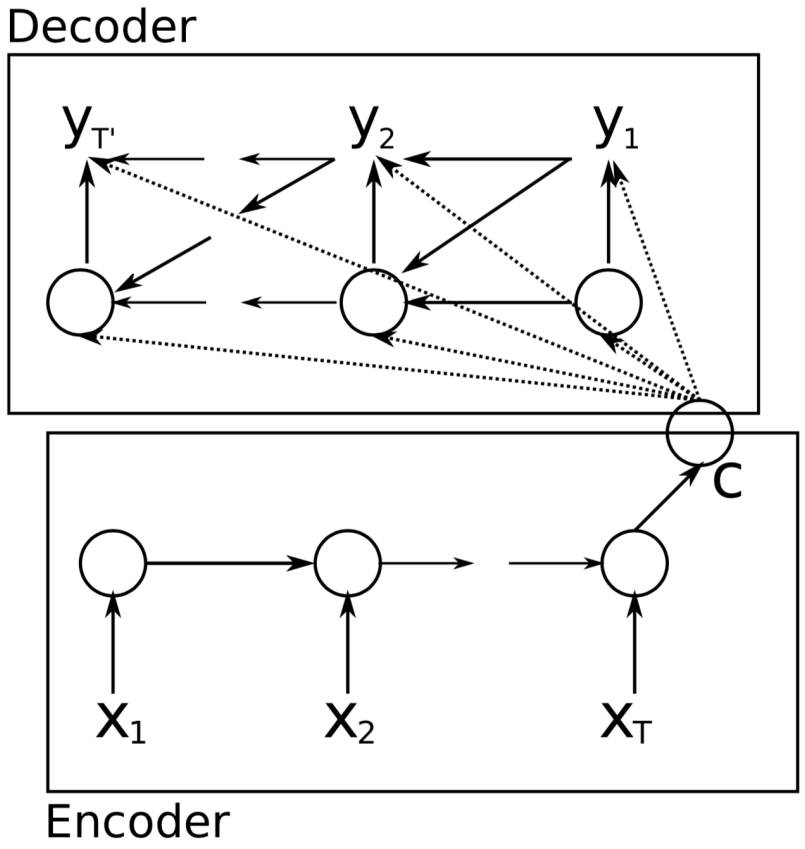
- We use Sequence to sequence model with different RNNs techniques.
- As for the RNN, we use LSTM as our base model. We also try the Bi-LSTM, as well as the LSTM with Attention Mechanism.
- We analyze the structure of model by different parts:
  - As for the optimizer, we have tried Adam and Rmsprop. For each optimizer, we tried different learning rate.
  - For the LSTM model, we analyzed model with different hidden layers as well as the learning rate.
  - We also tried different batch size in the training step.
- For the metric, we use the cross entropy error as the object function and BLEU score for the metric we evaluate the model. We use 80% of data as the training data and 20% as the test data.
- We also tried character model, but the effect is not so good.

SEQUENCE TO SEQUENCE MODEL

- Mapping between variable-length input to variable length output
- Cho et al, Learning Phrase Representation using RNN Encoder-Decoder for Statistical Machine Translation
- Sustkever et al, Sequence to Sequence Learning with Neural Networks



SEQUENCE TO SEQUENCE MODEL



$$\mathbf{h}_{\langle t \rangle} = f\left(\mathbf{h}_{\langle t-1 \rangle}, y_{t-1}, \mathbf{c}\right)$$

$$\mathbf{c} = \tanh\left(\mathbf{V}\mathbf{h}^{\langle N \rangle}\right)$$

$$P(y_t | y_{t-1}, y_{t-2}, \ldots, y_1, \mathbf{c}) = g\left(\mathbf{h}_{\langle t \rangle}, y_{t-1}, \mathbf{c}\right)$$

Objective Function

$$\max_{\boldsymbol{\theta}} \frac{1}{N} \sum_{n=1}^{N} \log p_{\boldsymbol{\theta}}(\mathbf{y}_n \mid \mathbf{x}_n)$$

$\boldsymbol{\theta}$ is the set of the model parameters

The parameters are calculated and updated by gradient descent

SEQ2SEQ PREDICTION

Greedy Search

For any input x

Given x, find word y0 with highest probability using RNN

Given y0 and x, find word y1 with highest probability using RNN

…

Stop when you see <EOS>

MODEL: LSTM

- Long Short-term Memory (LSTM) is a special design of RNN.

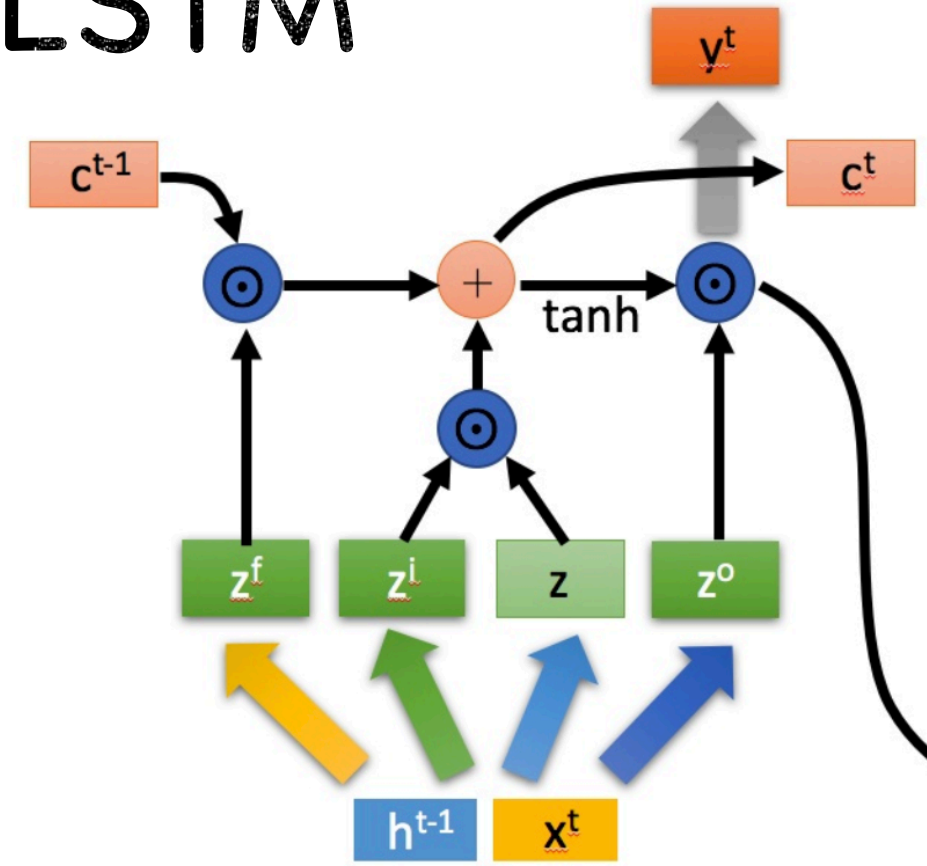
Why LSTM?

- RNN has a problem of gradient vanishing and gradient exploding.

The cat, with already ate a fish, was full.

- LSTMs are explicitly designed to avoid the long-term dependency problem. Remembering information for long periods of time is practically their default behavior, not something they struggle to learn!

LSTM



$$c^t = z^f \odot c^{t-1} + z^i \odot z$$

$$h^t = z^o \odot tanh(c^t)$$

$$y^t = \sigma(W' h^t)$$

$$z = tanh\left(W \begin{bmatrix} x^t \\ h^{t-1} \end{bmatrix}\right)$$

$$z^i = \sigma\left(W^i \begin{bmatrix} x^t \\ h^{t-1} \end{bmatrix}\right)$$

$$z^f = \sigma\left(W^f \begin{bmatrix} x^t \\ h^{t-1} \end{bmatrix}\right)$$

$$z^o = \sigma\left(W^o \begin{bmatrix} x^t \\ h^{t-1} \end{bmatrix}\right)$$

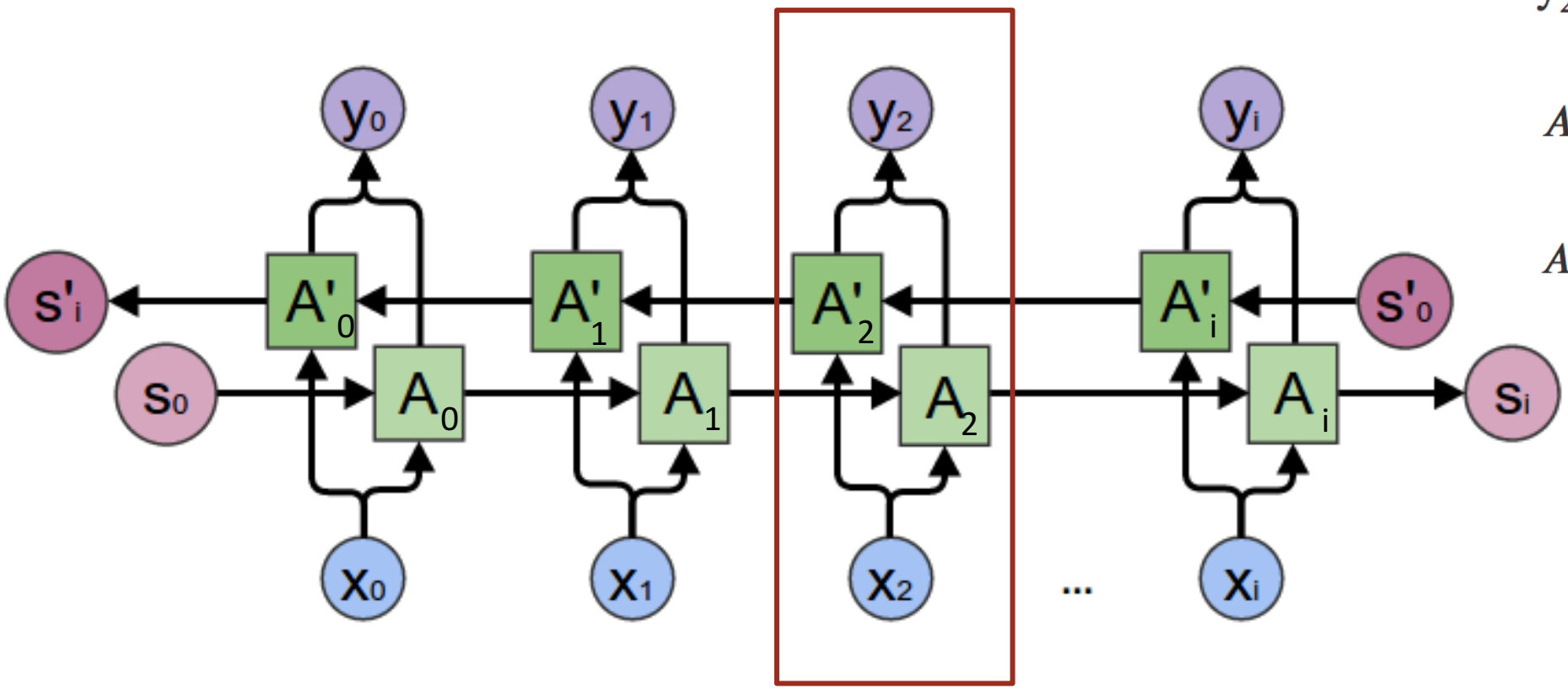
c changes slowly!
h changes fast!

MODEL STRUCTURE

Layer (type)	Output Shape	Param #	Connected to
encoder_inputs (InputLayer)	(None, None)	0	
encoder_embedding (Embedding)	(None, 24, 200)	6000400	encoder_inputs[0][0]
decoder_inputs (InputLayer)	(None, None, 30001)	0	
encoder_lstm (LSTM)	[(None, 200), (None,	320800	encoder_embedding[0][0]
decoder_lstm (LSTM)	[(None, None, 200),	24161600	decoder_inputs[0][0] encoder_lstm[0][1] encoder_lstm[0][2]
decoder_dense (Dense)	(None, None, 30001)	6030201	decoder_lstm[0][0]

Total params: 36,513,001
Trainable params: 36,513,001
Non-trainable params: 0

BIDIRECTIONAL LSTM



$$y_2 = g\left(V A_2 + V' A'_2\right)$$

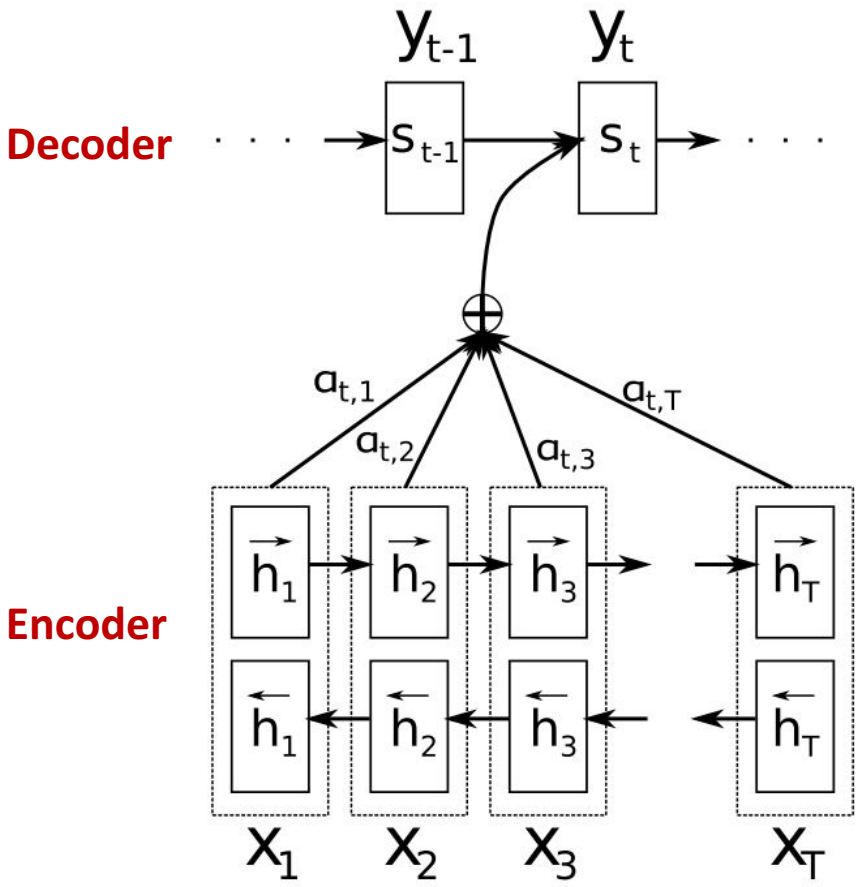
$$A_2 = f\left(W A_1 + U x_2\right)$$

$$A'_2 = f\left(W' A'_3 + U' x_2\right)$$

BIDIRECTIONAL LSTM

```
encoder_inputs = Input(shape=(None,), name = 'encode_inputs',dtype='int32')

encoder_embedding = Embedding(input_dim=num_encoder_tokens, output_dim=HIDDEN_UNITS,
                              input_length=encoder_max_seq_length, name='encoder_embedding')(encoder_inputs)


encoder_lstm, forward_h, forward_c, backward_h, backward_c =  Bidirectional(LSTM
     (HIDDEN_UNITS,
      dropout=0.2,
      return_sequences=True,
      return_state=True,
      recurrent_activation='relu',
      recurrent_initializer='glorot_uniform'))(encoder_embedding)


state_h = Concatenate()([forward_h, backward_h])
state_c = Concatenate()([forward_c, backward_c])

encoder_states = [state_h, state_c]
decoder_inputs = Input(shape=(None, num_decoder_tokens), name='decoder_inputs')
decoder_lstm = LSTM(units=HIDDEN_UNITS*2, return_state=True, return_sequences=True, name='decoder_lstm')

decoder_outputs, decoder_state_h, decoder_state_c = decoder_lstm(decoder_inputs,
                                                                 initial_state=encoder_states)

output = Dense(units=num_decoder_tokens, activation='softmax', name='decoder_dense')(decoder_outputs)

model = Model([encoder_inputs,decoder_inputs], outputs=output)

# summarize layers
print(model.summary())
'''
```

ATTENTION MECHANISM



Goal

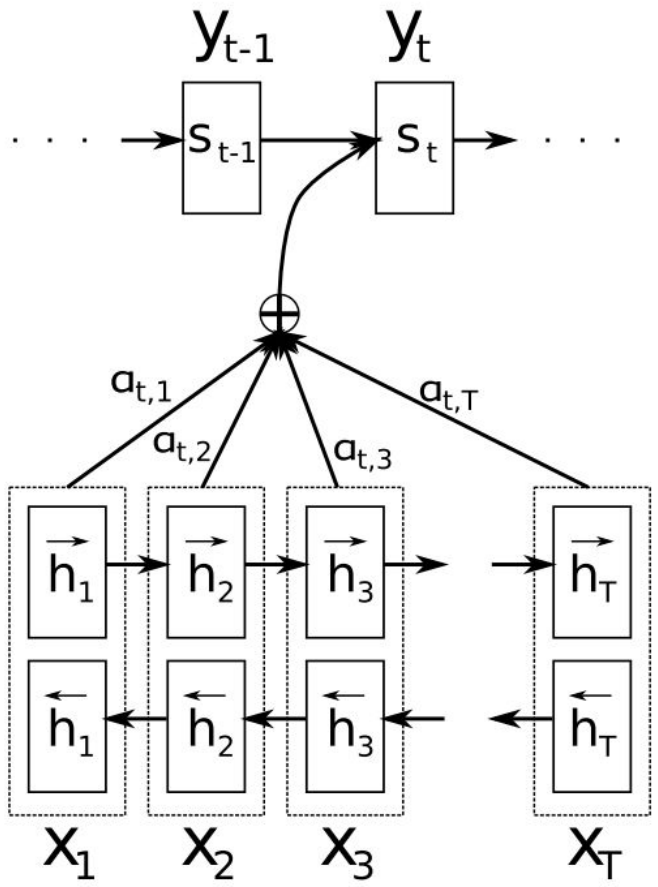
We aim to get the attention α_{tj} in order to compute s_t and the output

Algorithm

Firstly, we get the hidden state $(h_1, h_2, \ldots, h_T)$ from the RNN structure

Secondly, Suppose the hidden state of the current decoder is s_{t-1} we need to calculate $e_{tj} = a(s_{t-1}, h_j)$ where a is a function which can be learned by neural network.

ATTENTION MECHANISM



Goal

We aim to get the attention α_{tj} in order to compute s_t

Algorithm

Then, get the attention by a softmax function. $$\alpha_{tj} = \frac{exp(e_{tj})}{\sum_{k=1}^{T} exp(e_{tk})}$$

After that, get the context vector $$\overrightarrow{c_t} = \sum_{j=1}^{T} \alpha_{tj} h_j$$

Finally, we can compute the hidden state of the next time in the decoder, and the output of next time: $$s_t = f(s_{t-1}, y_{t-1}, c_t)$$

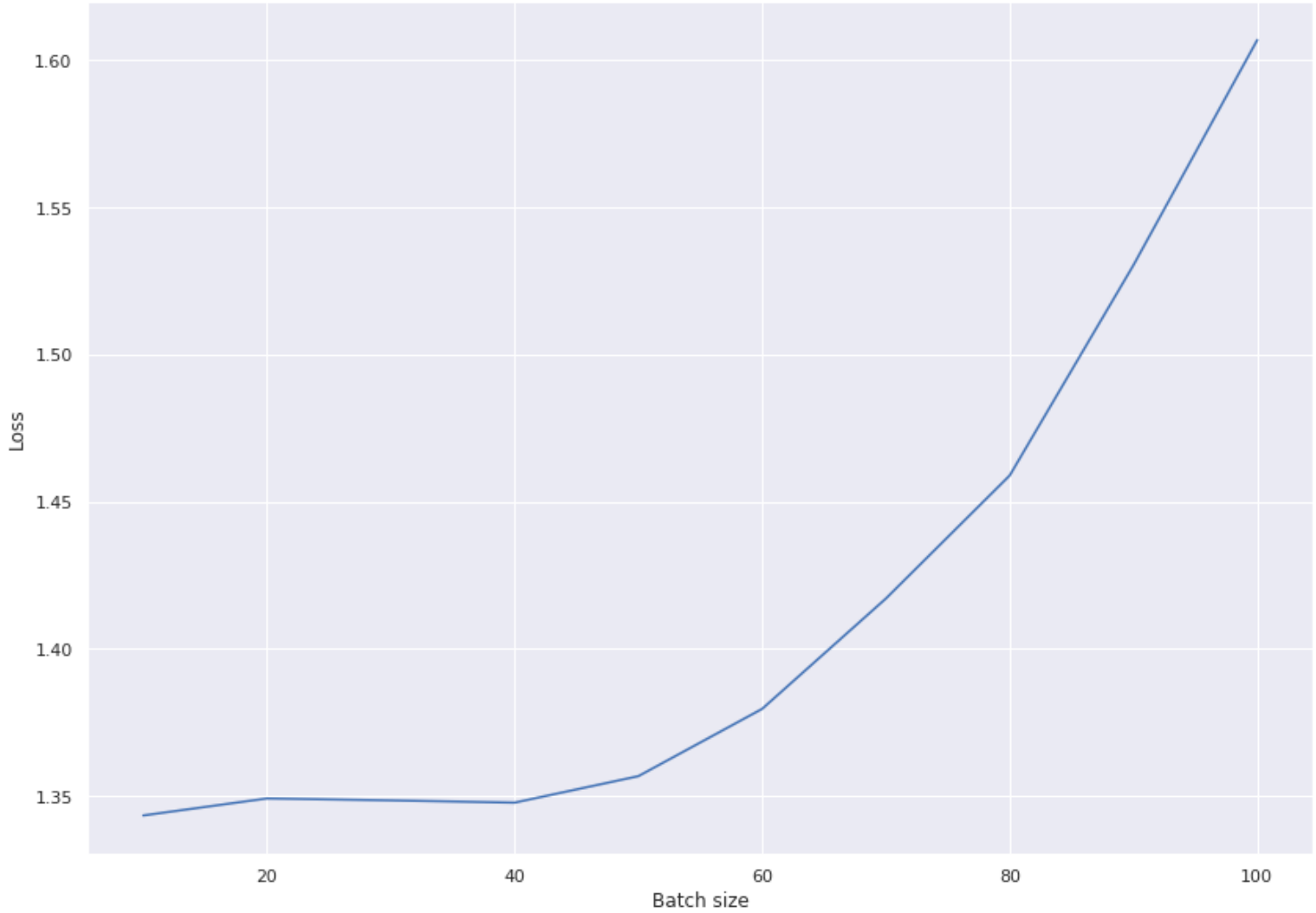
$$p(y_t | y_1, \ldots, y_{t-1}, \vec{x}) = g(y_{i-1}, s_i, c_i)$$

ATTENTION MECHANISM

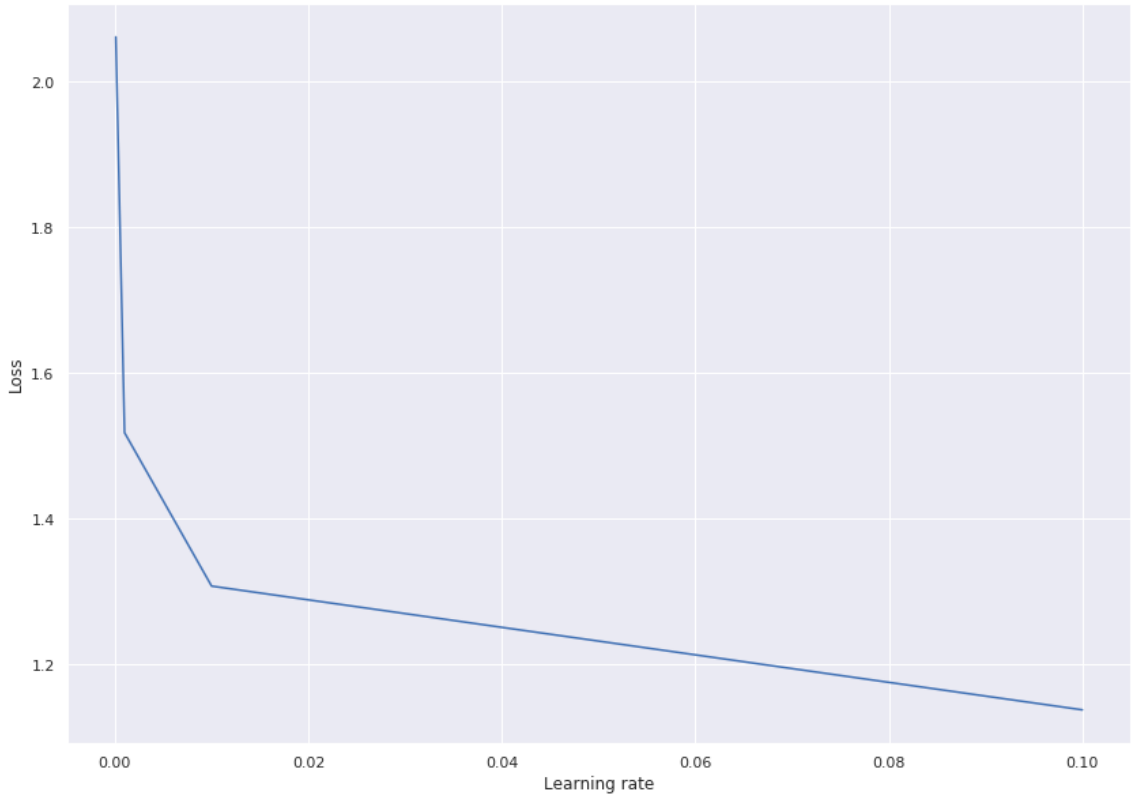
Using the package Keras Self-Attention

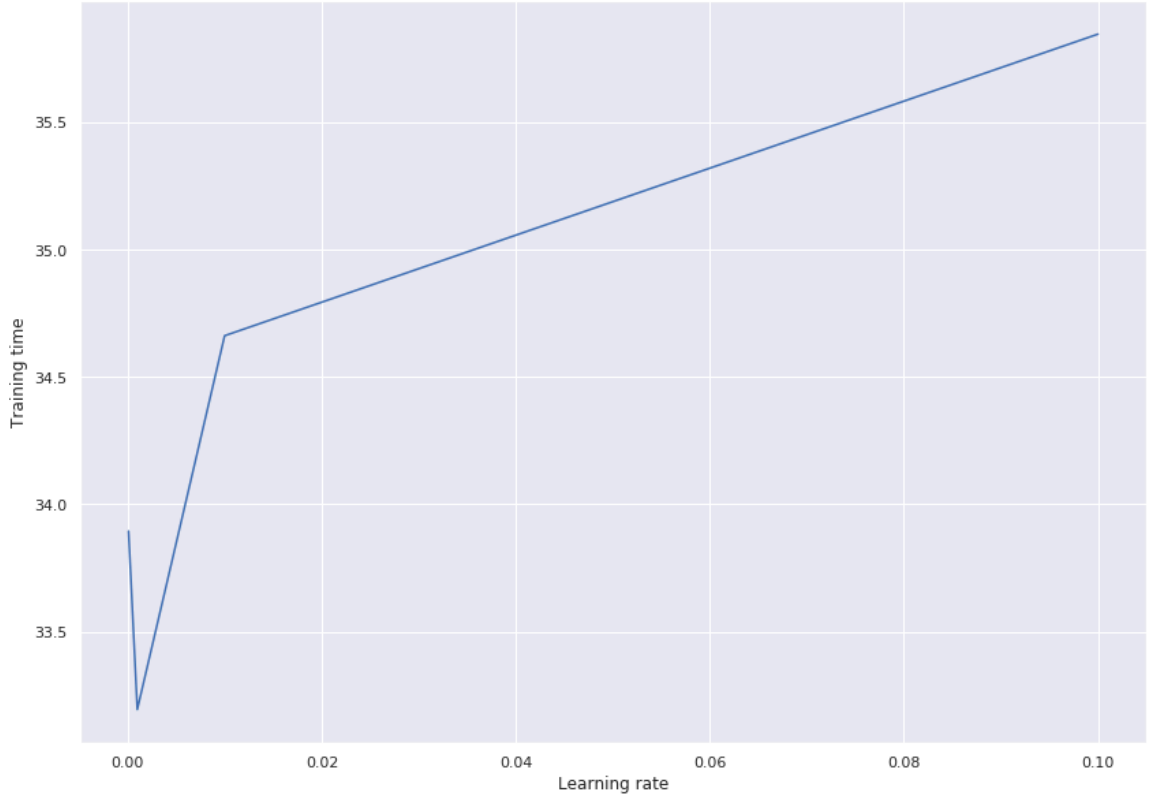
```
import keras
from keras_self_attention import SeqSelfAttention

inputs = keras.layers.Input(shape=(None,))
embd = keras.layers.Embedding(input_dim=32,
                              output_dim=16,
                              mask_zero=True)(inputs)
lstm = keras.layers.Bidirectional(keras.layers.LSTM(units=16,
                                                    return_sequences=True))(embd)


att = SeqSelfAttention(attention_type=SeqSelfAttention.ATTENTION_TYPE_MUL,
                       kernel_regularizer=keras.regularizers.l2(1e-4),
                       bias_regularizer=keras.regularizers.l1(1e-4),
                       attention_regularizer_weight=1e-4,
                       name='Attention')(lstm)
```

PARAMETERS TUNING: BATCH SIZE

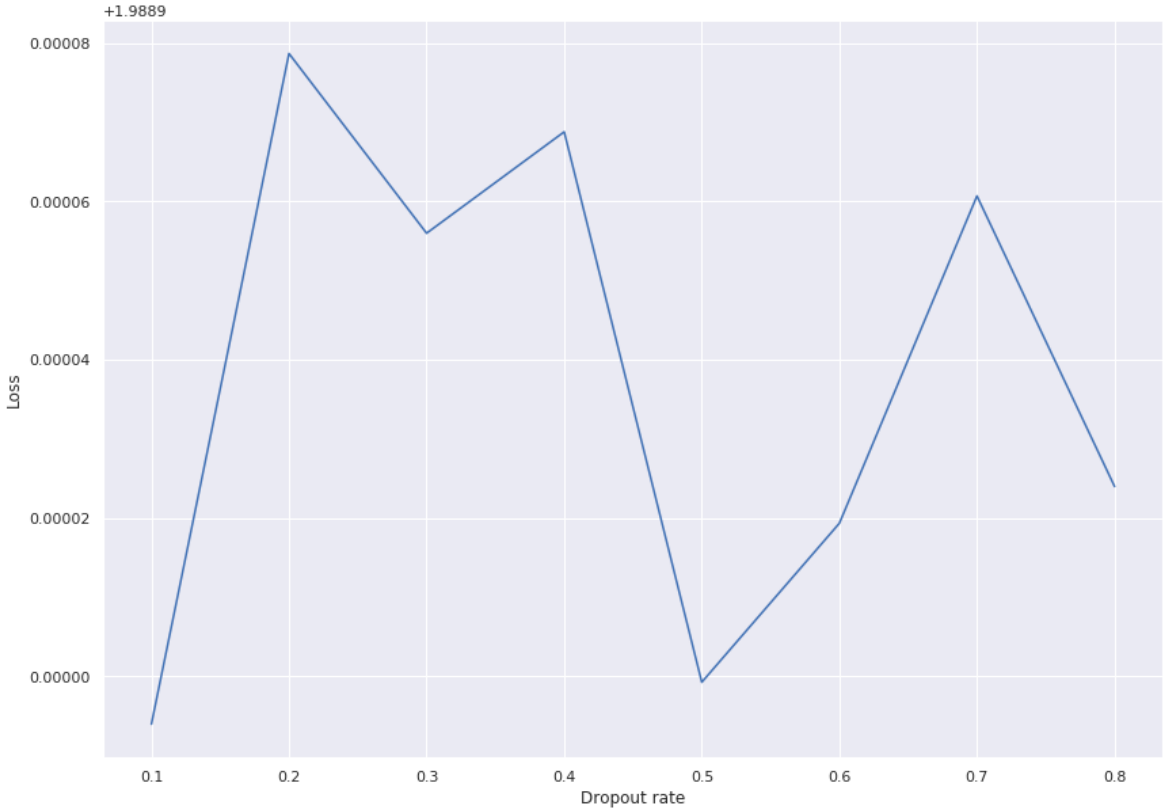


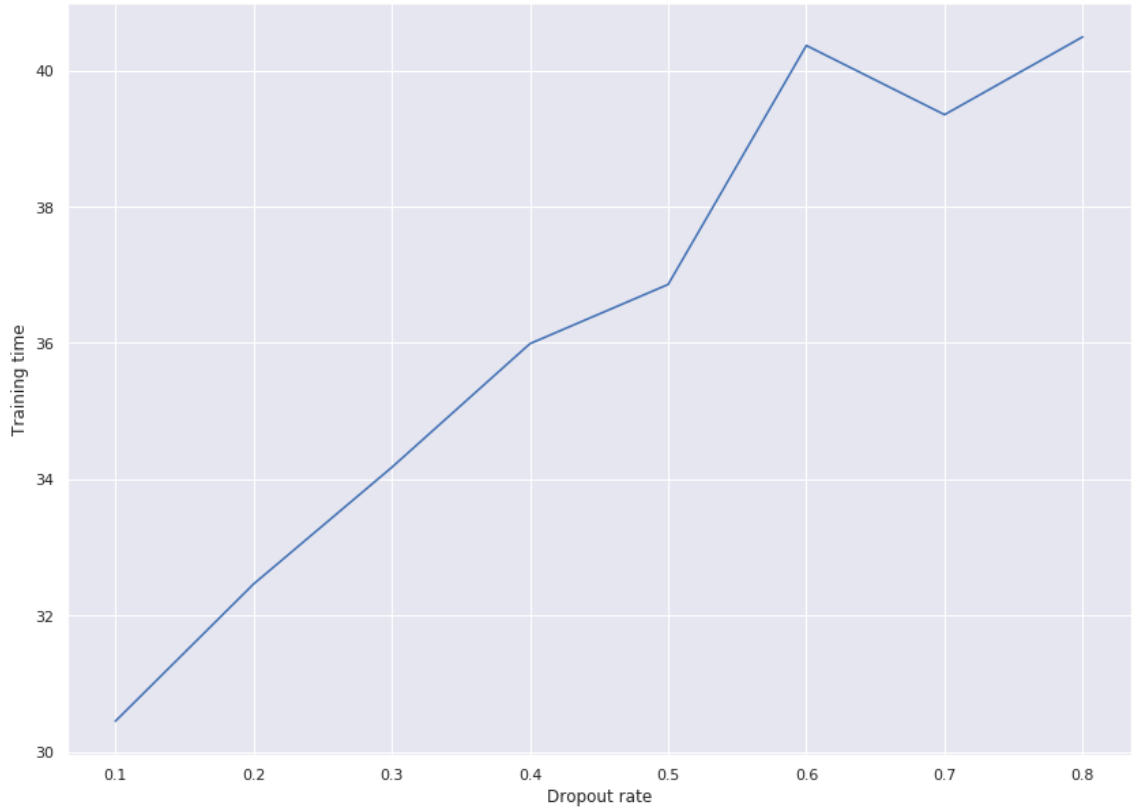
PARAMETERS TUNING: LEARNING RATE



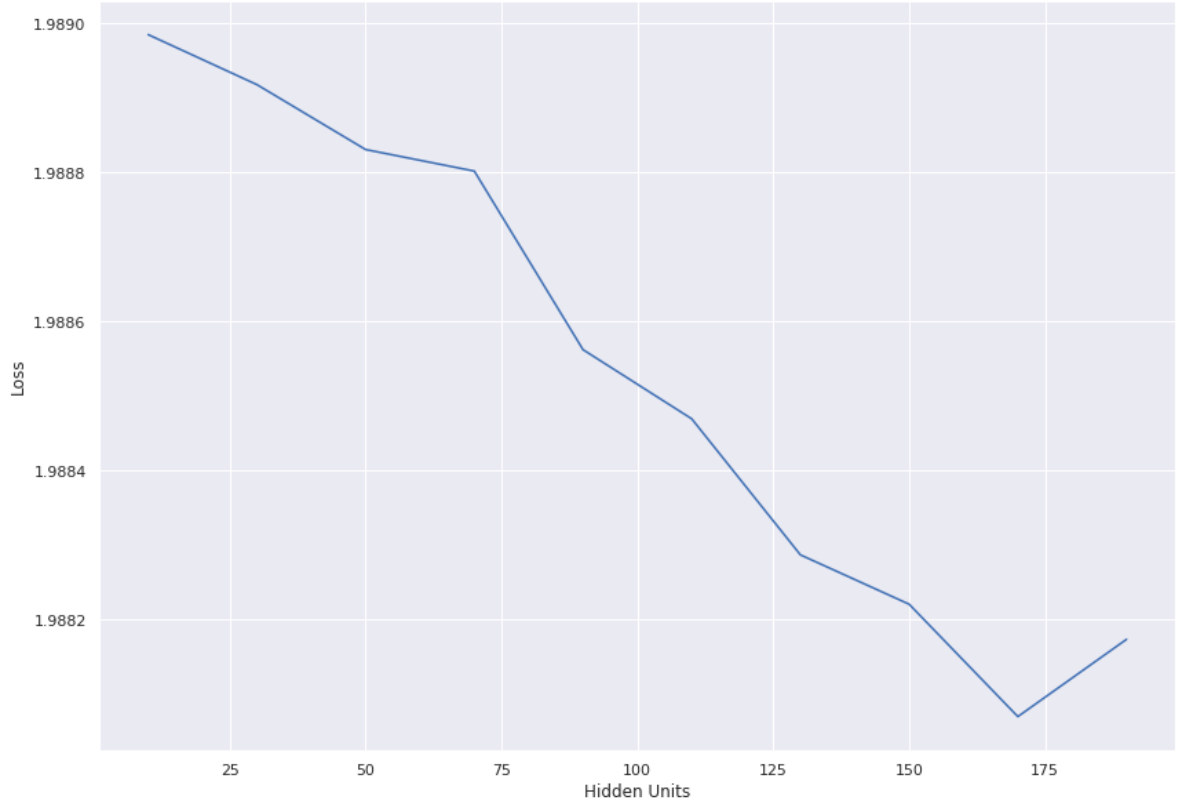


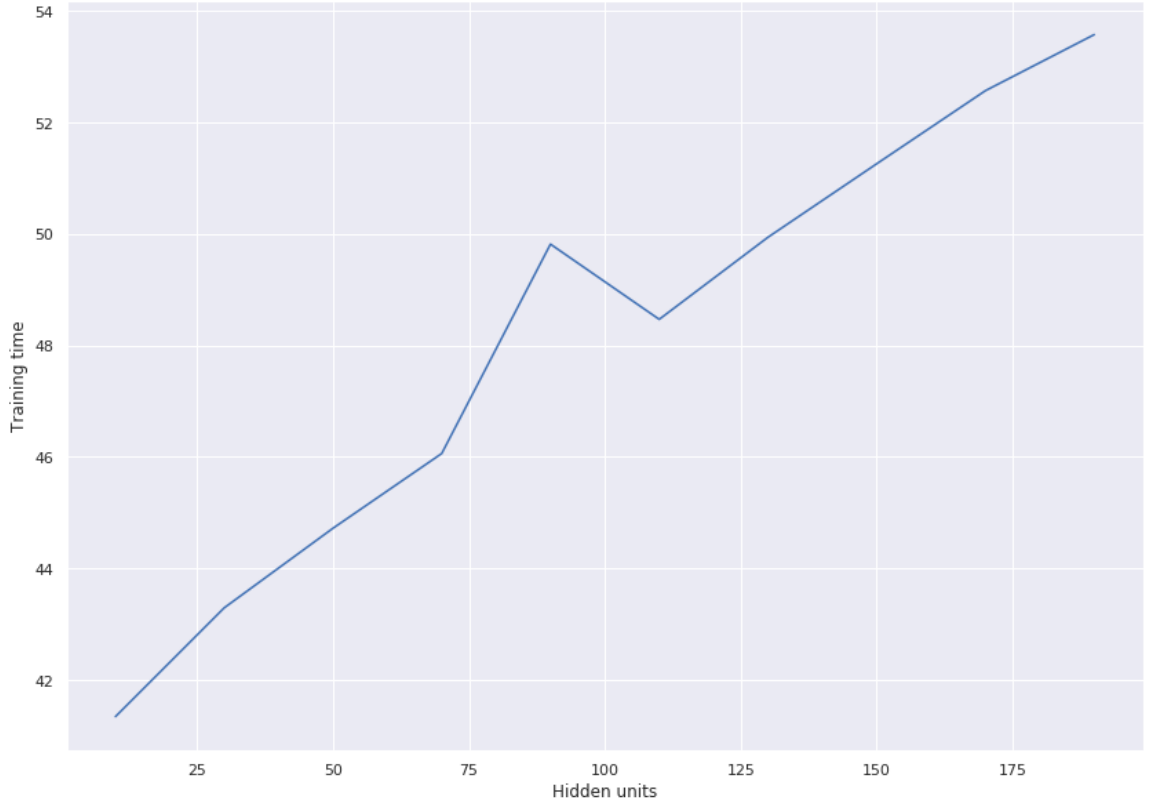
PARAMETERS TUNING: DROPOUT RATE





PARAMETERS TUNING: HIDDEN UNITS





PARAMETERS

	Vocabulary Size	Length of Sentence	Batch Size	Hidden layer units	Learning rate	Dropout rate	Epochs
Training Time	↑	↑	↓	↑	↑	↑	↑
Training Loss	↓	↓	Arbitrary	↓	↓	Arbitrary	↓
Cross Validation Loss	↓	↓	Arbitrary	↓	↓	Arbitrary	Arbitrary

MODEL COMPARISON

	LSTM (baseline)	LSTM with attention	Bidirectional LSTM without attention	Bidirectional LSTM with attention	GRU
BLEU score (1 and 2 gram)	2e-4	2e-4	0.013	0.014	3e-5
Cross Entropy loss	0.7882	0.6453	0.0231	0.0023	0.9923
Training time	15180 s	21423 s	35280 s	126234 s	9800 s

Note:
We choose epochs to be 10 (for the final state-of-art model, we may set epochs to be 20 or more)
The length of the sentence is 30 and the vocabulary in the dictionary is 15000.
The units in each layer is 256.

SUMMARY

- As for the hyper parameter of the data structure. The **vocabulary size** matters a lot, since the larger the vocabulary size, the longer time the training process takes, because the embedding dimension will be larger. Plus, less Unknown words will be generated and the model will predict less unknown words.

- For the structure of the model, the Bidirectional LSTM with attention is the best model. However, the parameters amount is much more than the basic model, and the training time increases a lot. (better GPU can solve this problem)

- As for the optimization function, RMSprop is much quicker than the adam, and the results does not differ a lot.

- **Epochs seem not so critical, and sometimes additional epochs can make the performance worse. For our model, the best epochs(minimum validation loss) is around 8.**

CHARACTER MODEL

- We also tried to build a new model, instead of predicting the word, we predict the characters.

Advantage:

- it eliminates the vocabulary size only to the characters size, which reduces 40000 to no more than 100. Thus, the training process is significantly accelerated.

```
{' ': 0, '!': 1, '"': 2, '#': 3, '$': 4, '%': 5, '&': 6, "'": 7, '(': 8, ')': 9, '*': 10, '+': 11, ',': 12, '-': 13, '.': 14, '/': 15,
'0': 16, '1': 17, '2': 18, '3': 19, '4': 20, '5': 21, '6': 22, '7': 23, '8': 24, '9': 25, ':': 26, ';': 27, '<': 28, '=': 29, '>': 30,
'?': 31, 'A': 32, 'B': 33, 'C': 34, 'D': 35, 'E': 36, 'F': 37, 'G': 38, 'H': 39, 'I': 40, 'J': 41, 'K': 42, 'L': 43, 'M': 44, 'N': 45,
'O': 46, 'P': 47, 'Q': 48, 'R': 49, 'S': 50, 'T': 51, 'U': 52, 'V': 53, 'W': 54, 'X': 55, 'Y': 56, 'Z': 57, '[': 58, ']': 59, '_': 60,
'`': 61, 'a': 62, 'b': 63, 'c': 64, 'd': 65, 'e': 66, 'f': 67, 'g': 68, 'h': 69, 'i': 70, 'j': 71, 'k': 72, 'l': 73, 'm': 74, 'n': 75,
'o': 76, 'p': 77, 'q': 78, 'r': 79, 's': 80, 't': 81, 'u': 82, 'v': 83, 'w': 84, 'x': 85, 'y': 86, 'z': 87, '{': 88, '|': 89, '}': 90,
'~': 91}
```

Disadvantage:

- The model performs not so good as the word model since it can only produce the meaning of the character.
- The model may generate some words which have never seen before.

CHARACTER MODEL

- We find that there are some characters which is not English characters, so we eliminate all the characters which are not English.

```
40
41 def is_number(uchar):
42     if uchar >= u'\u0030' and uchar<=u'\u0039':
43         return True
44     else:
45         return False
46
47
48
49 def is_alphabet(uchar):
50     if (uchar >= u'\u0041' and uchar<=u'\u005a') or (uchar >= u'\u0061' and uchar<=u'\u007a'):
51         return True
52     else:
53         return False
54
55
56
```

```
'ì',
'í',
'ï',
'ð',
'ñ',
'ò',
'ó',
'õ',
'ö',
'ø',
'ú',
'ü',
'ý',
'þ',
'Ā',
'ā',
'ă',
'ć',
'Đ',
'ē',
'ę',
'ě',
'ī',
'İ',
'ı',
'Ł',
'ł',
'ń',
'ō',
'Ś',
'ś',
'Ş',
'š',
```

RESULTS OF THE STATE-OF-ARTS MODEL

My model

```
my name is david. what is my name?
i 'm not a lawyer .
my name is John. what is my name?
i do n't know .
are you a leader or a follower?
i 'm not sure .
are you a follower or a leader?
i 'm not sure .
```

A Neural Conversational model

Conversation 2: Simple Q&A (contexts and multiple choice)

Human: *my name is david . what is my name ?*
Machine: *david .*
Human: *my name is john . what is my name ?*
Machine: *john .*
Human: *are you a leader or a follower ?*
Machine: *i 'm a leader .*
Human: *are you a follower or a leader ?*
Machine: *i 'm a leader .*

RESULTS OF THE STATE-OF-ARTS MODEL

My model

```
who is a skywalker?
the chief justice of the
who is bill cliton?
the police department of china
is sky blue or black ?
no .
does a cat have a tail?
it 's not a .
does a cat have a wing?
no , i did n't .
can a cat fly?
i 'm not sure .
how many legs does a cat have?
two
how many legs does a spider have?
two
```

A Neural Conversational model

Conversation 3: General knowledge Q&A

Human: *who is skywalker ?*
Machine: *he is a hero .*
Human: *who is bill clinton ?*
Machine: *he 's a billionaire .*
Human: *is sky blue or black ?*
Machine: *blue .*
Human: *does a cat have a tail ?*
Machine: *yes .*
Human: *does a cat have a wing ?*
Machine: *no*
Human: *can a cat fly ?*
Machine: *no .*
Human: *how many legs does a cat have ?*
Machine: *four , i think .*
Human: *how many legs does a spider have ?*
Machine: *three , i think .*
Human: *how many legs does a centipede have ?*
Machine: *eight .*

RESULTS OF THE STATE-OF-ARTS MODEL

My model

```
what is the purpose of life ?
a small body

what is the purpose of living ?
to protect the entire constitutional rights

what is the purpose of existence ?
the UNK of the u.s. constitution

where are you now ?
i 'm not sure .

what is the purpose of dying ?
to protect the historic center

what is the purpose of being intelligent ?
to protect the entire constitutional rights

what is the purpose of emotions?
to protect the border
```

A Neural Conversational model

Conversation 4: Philosophical Q&A

Human: *what is the purpose of life ?*
Machine: *to serve the greater good .*
Human: *what is the purpose of living ?*
Machine: *to live forever .*
Human: *what is the purpose of existence ?*
Machine: *to find out what happens when we get to the planet earth .*
Human: *where are you now ?*
Machine: *i 'm in the middle of nowhere .*
Human: *what is the purpose of dying ?*
Machine: *to have a life .*
Human: *what is the purpose of being intelligent ?*
Machine: *to find out what it is .*
Human: *what is the purpose of emotions ?*
Machine: *i don 't know .*

RESULTS OF THE STATE-OF-ARTS MODEL

My model

```
what is moral ?
a network of a male entity
what is immoral ?
the austrians , the federal register ( ) and the UNK
what is morality ?
a .
what is the definition of altruism ?
the notion of selective breeding
ok ... so what is the definition of moral- ity ?
i do n't know .
ok ... so what is the definition of morality ?
i do n't know .
well , the truth is , you 're not a believer in god almighty .
i 'm sorry , i did n't mean to .
tell me the definition of morality , i am quite upset now !
i 'm sorry . i 'm not .
```

A Neural Conversational model

Human: *what is moral ?*
Machine: *what empowered humanity , what intellectual the essence is .*
Human: *what is immoral ?*
Machine: *the fact that you have a child .*
Human: *what is morality ?*
Machine: *what is altruism ?*
Human: *what is the definition of altruism ?*
Machine: *if you don 't believe in god , then you don 't know .*
Human: *ok ... so what is the definition of morality ?*
Machine: *well , the truth is , you 're not a believer in god almighty .*
Human: *tell me the definition of morality , i am quite upset now !*
Machine: *i 'm not ashamed of being a philosopher !*

RESULTS OF THE STATE-OF-ARTS MODEL

My model: Some interesting results

```
I am going to school tommorrow
you 're a liar .

Who is Lebron James?
the cardinal vicar of the former general of the emperor
```

USER INTERFACE

- We use flask as our main tool of developing the user interface and build the website.
- Flask is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries.
- We created a post method to get the message in the html file and send the sentence to the function of the prediction of the chatbot. After that, we send the reply back to the html again.

USER INTERFACE

The python file

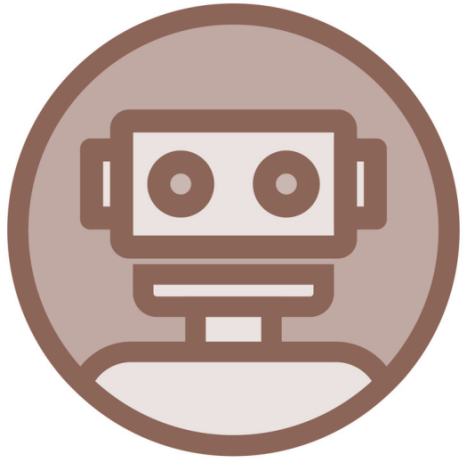
```
57  @app.route('/word_reply', methods=['POST', 'GET'])
58  def cornell_word_reply():
59      if request.method == 'POST':
60          if 'sentence' not in request.form:
61              flash('No sentence post')
62              redirect(request.url)
63          elif request.form['sentence'] == '':
64              flash('No sentence')
65              redirect(request.url)
66          else:
67              sent = request.form['sentence']
68              cornell_word_chat_bot_conversations.append('YOU: ' + sent)
69              reply = cornell_word_chat_bot.reply(sent)
70              cornell_word_chat_bot_conversations.append('BOT: ' + reply)
71      return render_template('word_reply.html', conversations=cornell_word_chat_bot_conversations)
```

USER INTERFACE

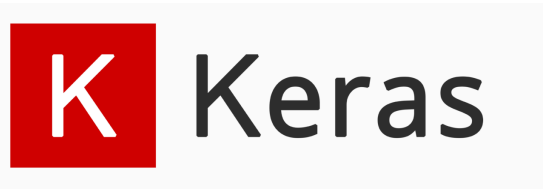
The html file

```
{% extends "layout.html" %}
{% block body %}
    <p>Let's chat!:</p>
    <form method=post>
        <ul>
            {% for line in conversations %}
            <li>{{line}}</li>
            {% endfor %}
        </ul>
       <p>Say something: <input type=text name=sentence></p>
         <input type=submit value=Send>
    </form>
{% endblock %}


```

LET’S SEE THE RESULT



THANKS TO THE TOOLS WE USE



REFERENCE

- Mikolov, T. *Statistical Language Models based on Neural Networks*. PhD thesis, Brno University of Technology, 2012.
- Shang, L., Lu, Z., and Li, H. Neural responding ma- chine for short-text conversation. In *Proceedings of ACL*, 2015.
- Sutskever, I., Vinyals, O., and Le, Q. V. Sequence to se- quence learning with neural networks. In *NIPS*, 2014
- Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin. A neural probabilistic language model. In *Journal of Machine Learning Research*, pages 1137–1155, 2003.
- K.Cho,B.Merrienboer,C.Gulcehre,F.Bougares,H.Schwenk,andY.Bengio.Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Arxiv preprint arXiv:1406.1078*, 2014.
- S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 1997.
- T. Mikolov, M. Karafia ´t, L. Burget, J. Cernocky`, and S. Khudanpur. Recurrent neural network based language model. In *INTERSPEECH*, pages 1045–1048, 2010.

Thanks for listening

Q&A