Neural Machine Translation

Seetu Agarwal

UNIVERSITY *of* WASHINGTON

Introduction

- Machine Translation is the ability of machines to transform text from one language to another.

- It is the most prominent field in natural language processing which aims to improve human-computer interaction.

- Google translate is one of the popular tool widely used to convert text from one language to another. Currently it supports 108 languages.

Motivation

- At times when we have international meetings and conferences or when we visit different countries, a tool that can convert the audio into the text or audio that we can easily understand is really very helpful.
- Often we came across a video which is in language which we can’t understand. At this time, subtitles in a language which can be easily understood is necessary in understanding the content of the video.
- The machine translation tools will remove the language barrier and makes life easy for travelers in other countries

Existing Related Works

- In 1970's **Rule Based Machine Translation** (RBMT) was the major focus for the researchers. [1] In this method source text is converted into language-free conceptual representation. Information that was implicit in the source text was augmented to this representation. Finally the augmented representation is converted into target language. For the languages which have different structures (English-Japan),this technique is complicated.

Table 1. Categories of RBMT Systems

Type	**Description**
Direct Systems [1]	Input sentences are mapped directly to the output sentences
Transfer RBMT Systems [1]	These use morphological and syntactic analysis to translate sentences
Interlingual RBMT Systems [1]	Input sentence is transformed to an abstract representation

Existing Related Works

- **Statistical Machine Translation** (SMT) is introduced later.
- Suppose S is a sentence in the source language and T is the translation in target language. This system assigns each (S,T) sentence pair the probability P(T,S) which is the probability that sentence T is the translated equivalent in the target language of the sentence S in the source language.

- In the **word-based SMT** [1], the source text is partitioned into a set of fixed locations. Glossary and contextual information is used to select the corresponding set of fixed locations in the target language. The words of the target fixed location are arranged into a sequence that forms the target language.

- In **phrase-based SMT** [1] translation system, phase translation probability is defined to map phrases in source language to phrases in target languages.

Existing Related Works

- Neural machine translation which uses artificial neural networks forms the backbone of the most popular tool in machine translation that is google translate.
- For WMT-2014 English-French dataset, Table 2 describes some existing res

Table 2. Methods on WMT-2014 English-French dataset

Model	Description	BLEU score
Transformer [3]	Very deep transformers for neural machine translation - 60 encoder layers and 12 decoder layers	43.8
ConvS2S [3]	Convolutional sequence to sequence learning	41.3
CSLM + RNN + WP [3]	Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation	34.54
LSTM [3]	Sequence-to-Sequence Learning	34.8

Existing Related Work

- Google Neural Machine Translation System [2] : It uses multiple encoder decoder layers in LSTM (Long Short Term Memory) network.

Table Single model results on WMT En→Fr (newstest2014)

Model	BLEU	CPU decoding time per sentence (s)
Word	37.90	0.2226
Character	38.01	1.0530
WPM-8K	38.27	0.1919
WPM-16K	37.60	0.1874
WPM-32K	38.95	0.2118
Mixed Word/Character	38.39	0.2774
PBMT [15]	37.0	
LSTM (6 layers) [31]	31.5	
LSTM (6 layers + PosUnk) [31]	33.1	
Deep-Att [45]	37.7	
Deep-Att + PosUnk [45]	39.2	

Architecture

- Initial phase is data cleaning and data pre-processing in the design. In the cleaning phase all the lines starting with xml tags are removed and empty lines are stripped. Pre-processing mainly includes tokenization and padding.

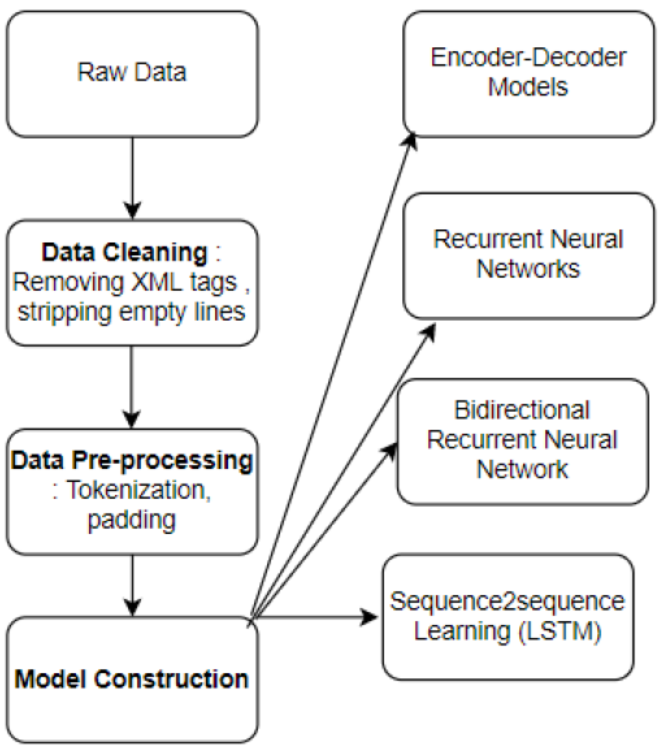


Figure 1 : System Architecture

Dataset Description

WMT2014 : It contains collection of datasets used in the workshop for on statistical machine translation for various tasks medical text translation task, news translation task etc.

WMT2014 English-French dataset [3]

- ***Text in English***

```
new jersey is sometimes quiet during autumn , and it is snowy in april .
the united states is usually chilly during july , and it is usually freezing in november .
```

- ***Text in French***

```
new jersey est parfois calme pendant l' automne , et il est neigeux en avril .
les états-unis est généralement froid en juillet , et il gèle habituellement en novembre .
```

Data Pre-processing

- Tokenization: Neural Networks did not understand ASCII character encodings. Therefore text like “Apple” , “Cat” etc needs to be converted into numbers as Neural Network is just a sequence of addition and multiplication operations. Either each character can be assigned a number or each word can be assigned a number .

```
{'the': 1, 'quick': 2, 'a': 3, 'brown': 4, 'fox': 5, 'jumps': 6, 'over': 7, 'lazy': 8, 'dog': 9, 'by': 10, 'jove': 11, 'my': 12, 'study': 13, '

Sequence 1 in x
  Input:  The quick brown fox jumps over the lazy dog .
  Output: [1, 2, 4, 5, 6, 7, 1, 8, 9]
Sequence 2 in x
  Input:  By Jove , my quick study of lexicography won a prize .
  Output: [10, 11, 12, 2, 13, 14, 15, 16, 3, 17]
Sequence 3 in x
  Input:  This is a short sentence .
  Output: [18, 19, 3, 20, 21]
```

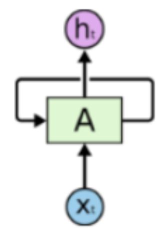
Data Pre-processing

- Padding : Sentences are dynamic in length , when aggregating word id’s together padding can be added to make the sentences of equal length.

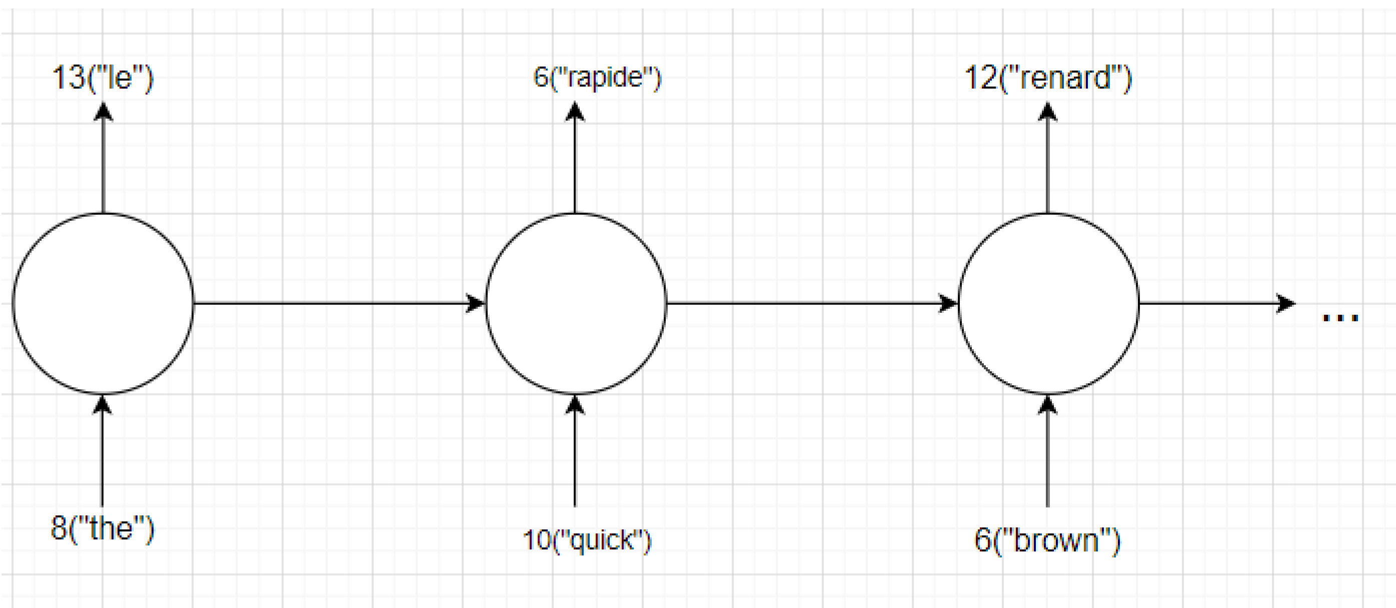
```
Sequence 1 in x
  Input:  [1 2 4 5 6 7 1 8 9]
  Output: [1 2 4 5 6 7 1 8 9 0]
Sequence 2 in x
  Input:  [10 11 12  2 13 14 15 16  3 17]
  Output: [10 11 12  2 13 14 15 16  3 17]
Sequence 3 in x
  Input:  [18 19  3 20 21]
  Output: [18 19  3 20 21  0  0  0  0  0]
```

Basic Recurrent Neural Network

- Outputs from previous steps are fed as input to the current step.



Recurrent Neural Networks have loops.



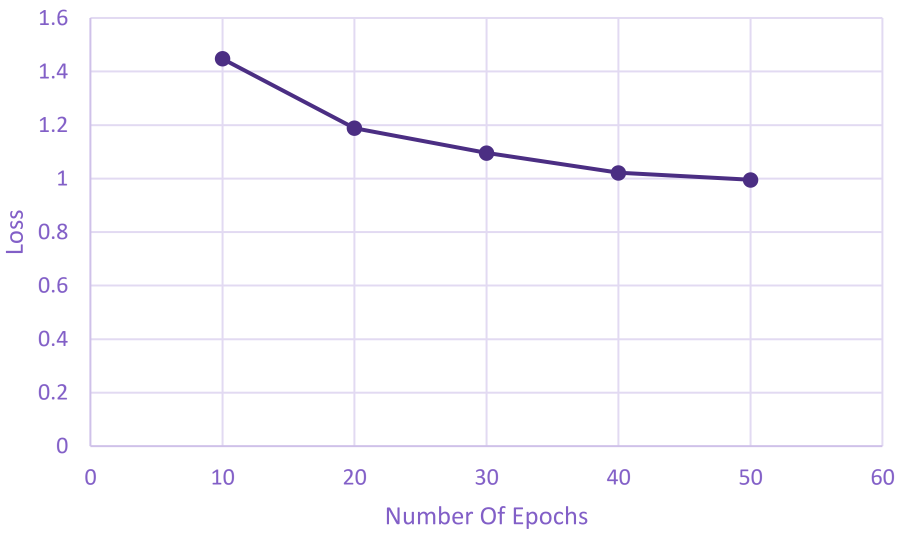
Translation Results

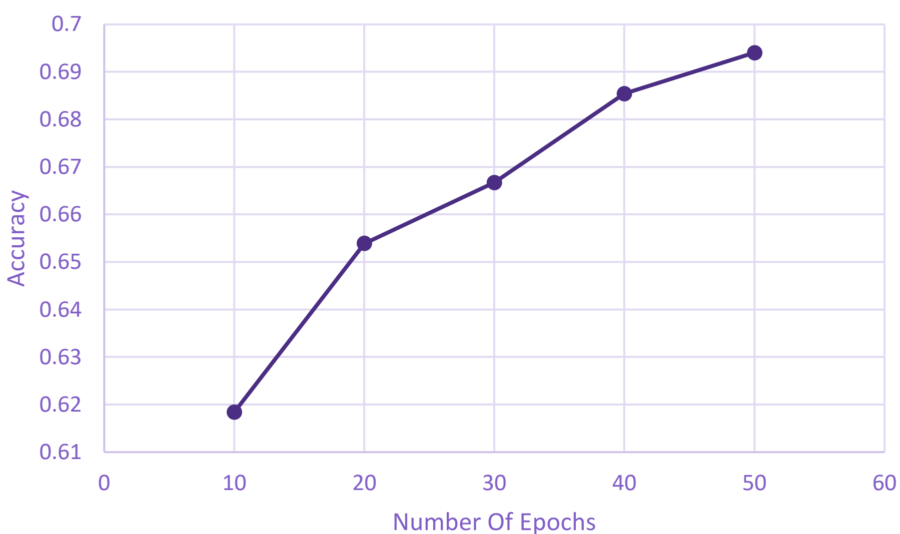
```
# Print prediction(s)
print("Prediction:")
print(logits_to_text(simple_rnn_model.predict(tmp_x[:1])[0], french_tokenizer))

print("\nCorrect Translation:")
print(french_sentences[:1])

print("\nOriginal text:")
print(english_sentences[:1])
```

```
Prediction:
new jersey est parfois calme en mois et il et il en agréable <PAD> <PAD> <PAD> <PAD> <PAD> <PAD> <PAD> <PAD>

Correct Translation:
["new jersey est parfois calme pendant l' automne , et il est neigeux en avril ."]

Original text:
['new jersey is sometimes quiet during autumn , and it is snowy in april .']
```

Evaluation

Number of Epochs	Validation Loss	Validation Accuracy
10	1.4477	0.6184
20	1.1885	0.6539
30	1.0956	0.6667
40	1.0217	0.6854
50	0.9957	0.6940

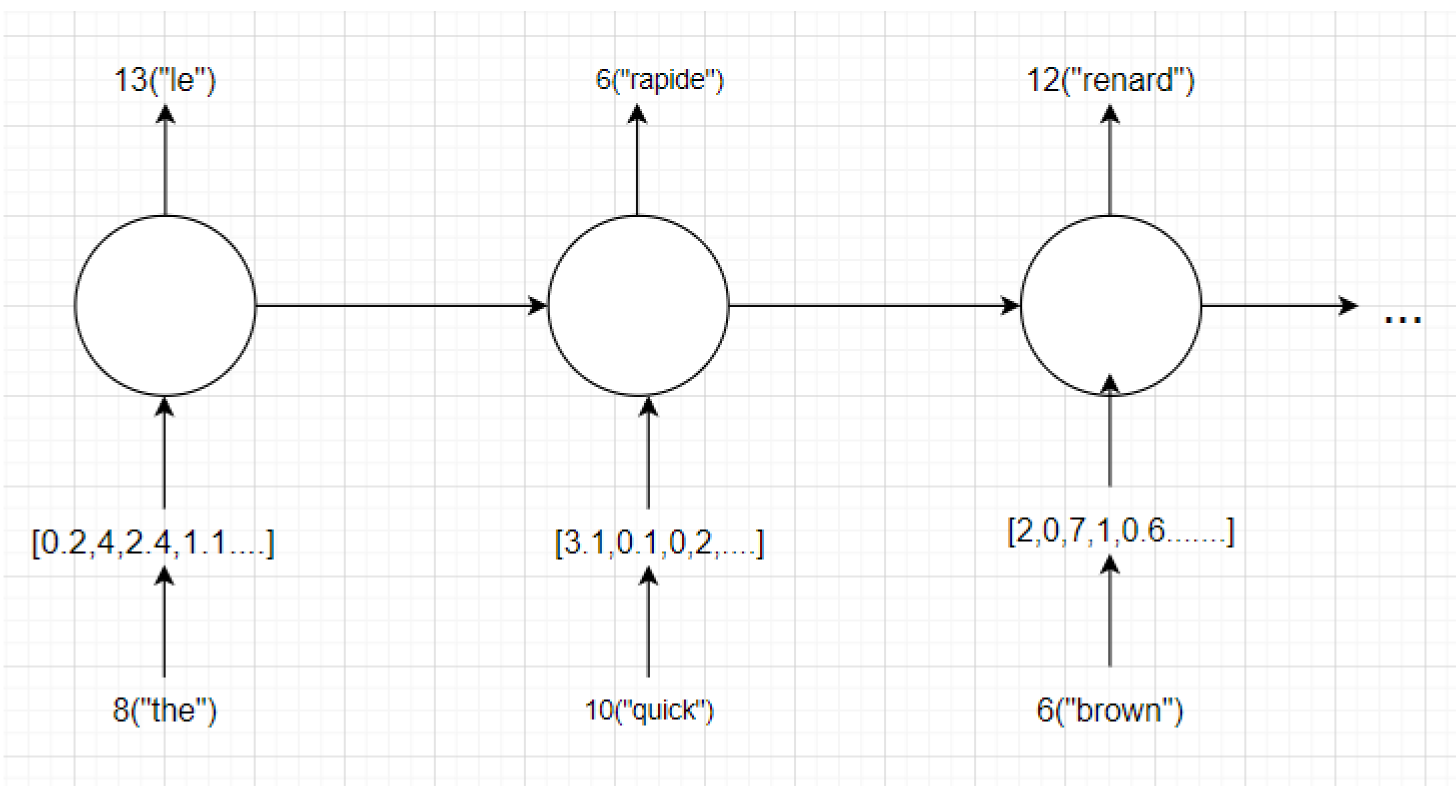




BLEU SCORE : 0.20

Recurrent Neural Network with Word Embedding

- Word Embedding : is a technique where words are represented as vectors in a predefined vector space.



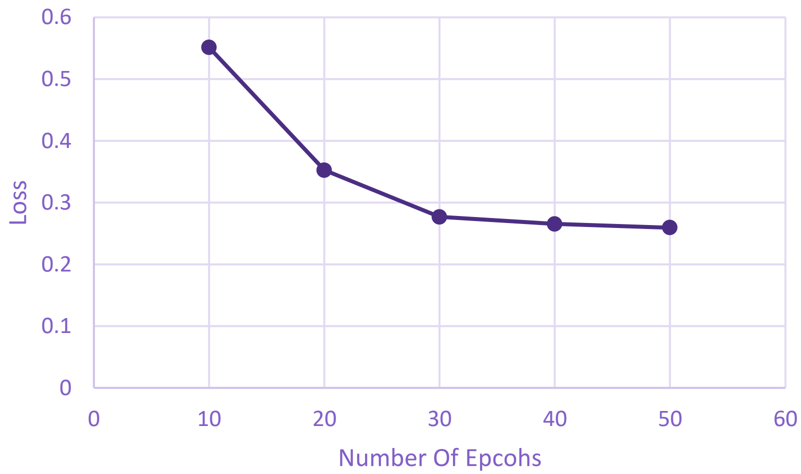
Translation Result

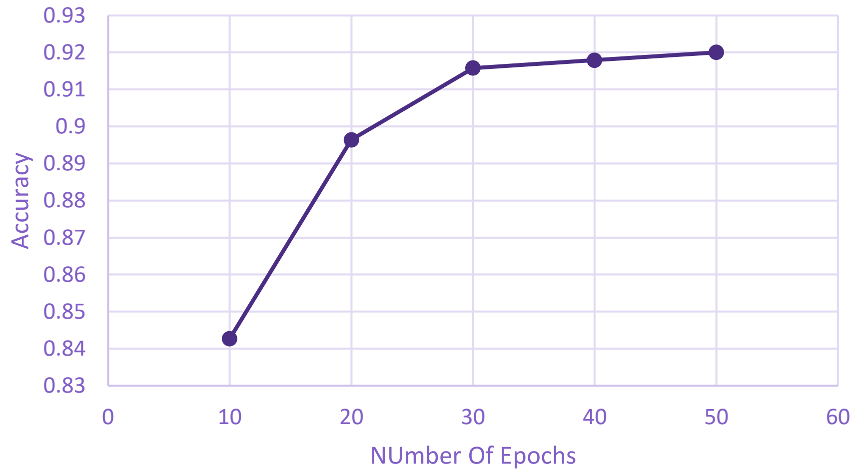
```
# Print prediction(s)
print("Prediction:")
print(logits_to_text(embeded_model.predict(tmp_x[:1])[0], french_tokenizer))

print("\nCorrect Translation:")
print(french_sentences[:1])

print("\nOriginal text:")
print(english_sentences[:1])
```

```
Prediction:

new jersey est parfois calme en l' automne et il est neigeux en avril <PAD> <PAD> <PAD> <PAD> <PAD> <PAD> <PAD>

Correct Translation:
["new jersey est parfois calme pendant l' automne , et il est neigeux en avril ."]

Original text:
['new jersey is sometimes quiet during autumn , and it is snowy in april .']
```

Evaluation

Number Of Epochs	Validation loss	Validation accuracy
10	0.5514	0.8427
20	0.3526	0.8964
30	0.2770	0.9158
40	0.2656	0.9179
50	0.2597	0.9200

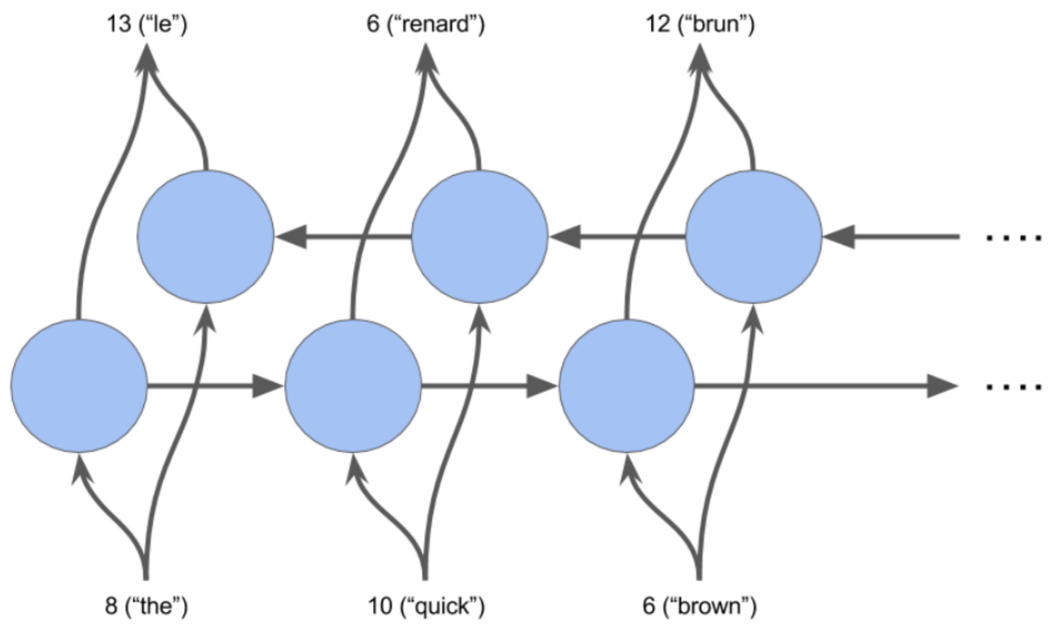




BLEU SCORE : 0.27

Bi-directional Recurrent Neural Network

- It connects two hidden layers of opposite direction to the same output, that is the output layer can get the information from past (backwards) and future (forward) states simultaneously.



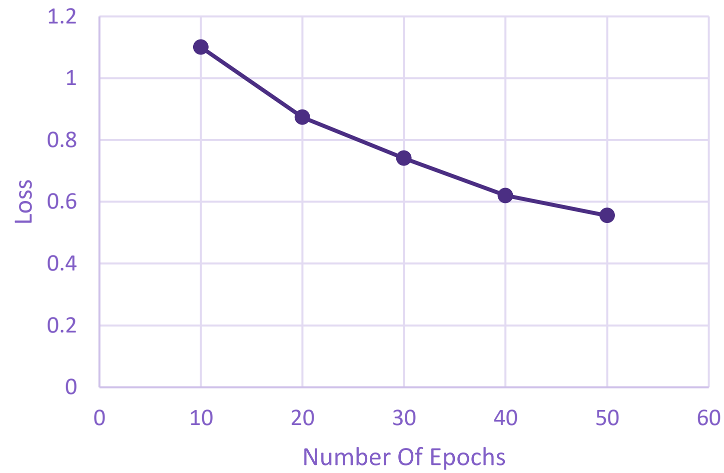
Translation Result

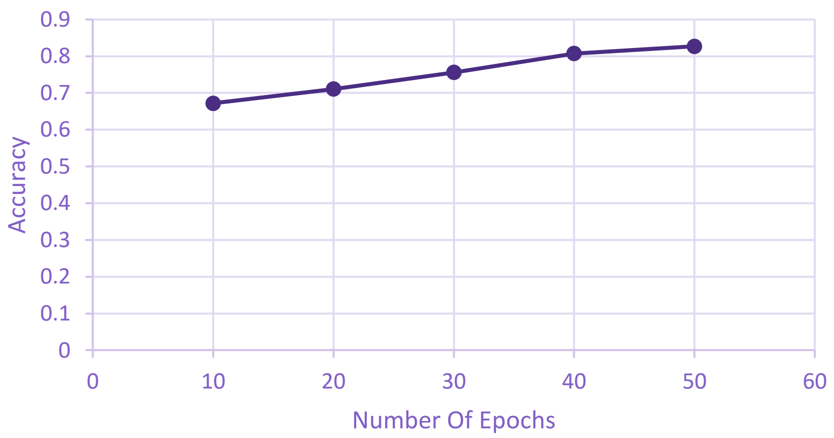
```
# Print prediction(s)
print("Prediction:")
print(logits_to_text(bidi_model.predict(tmp_x[:1])[0], french_tokenizer))

print("\nCorrect Translation:")
print(french_sentences[:1])

print("\nOriginal text:")
print(english_sentences[:1])
```

```
Prediction:
new jersey est parfois occupé en printemps mais il gèle agréable l' mai <PAD> <PAD> <PAD> <PAD> <PAD> <PAD> <PAD> <PAD>

Correct Translation:
["new jersey est parfois calme pendant l' automne , et il est neigeux en avril ."]

Original text:
['new jersey is sometimes quiet during autumn , and it is snowy in april .']
```

Evaluation

Number Of Epochs	Validation loss	Validation Accuracy
10	1.1006	0.6720
20	0.8737	0.7105
30	0.7411	0.7557
40	0.6202	0.8071
50	0.5555	0.8266

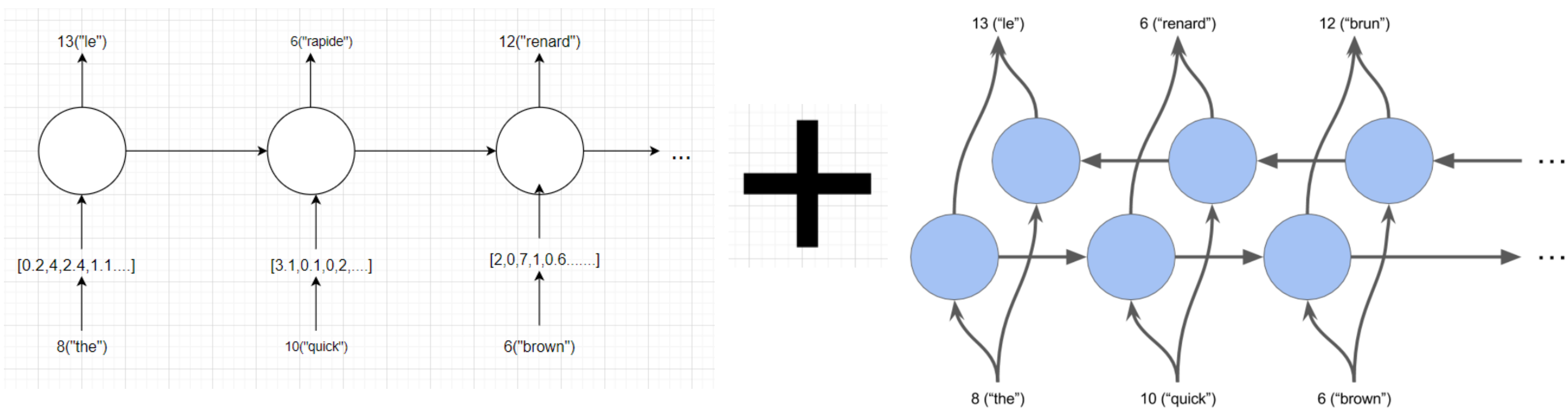




BLEU SCORE : 0.29

Word Embedding + Bidirectional RNN

- Combines both of the learned techniques into one RNN.



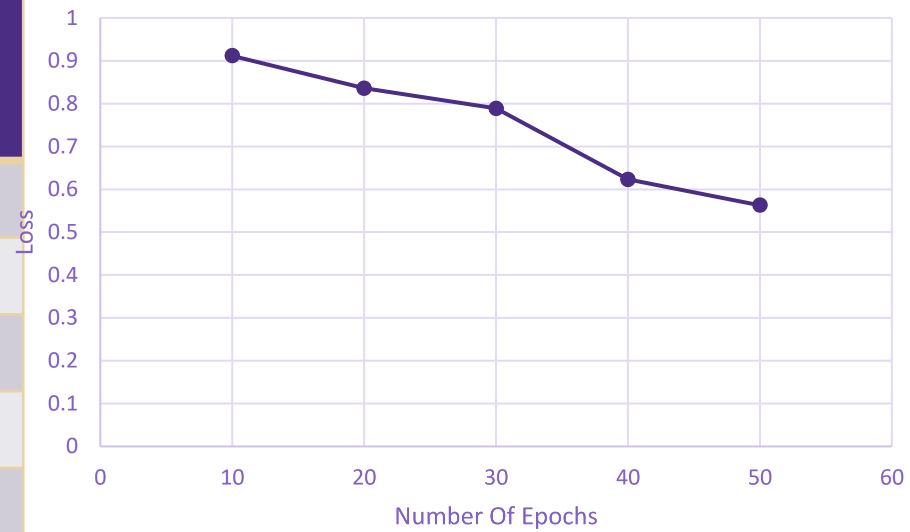
Translation Result

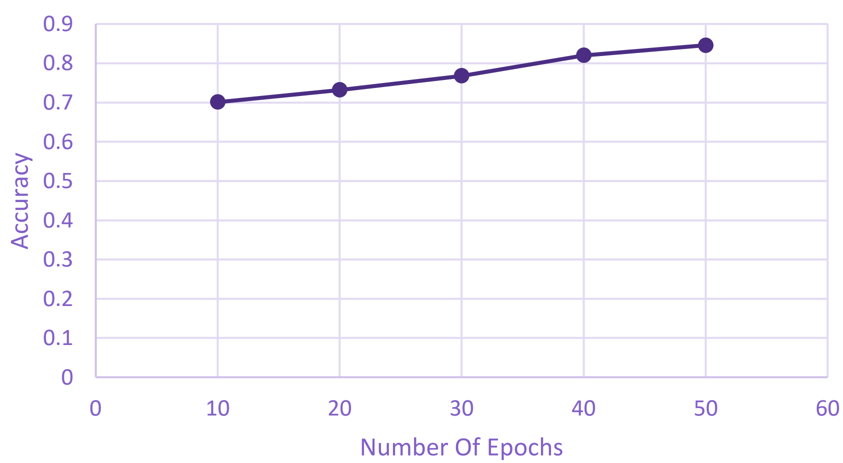
```
# Print prediction(s)
print("Prediction:")
print(logits_to_text(embeded_model.predict(tmp_x[:1])[0], french_tokenizer))

print("\nCorrect Translation:")
print(french_sentences[:1])

print("\nOriginal text:")
print(english_sentences[:1])
```

```
Prediction:

new jersey est parfois calme en l' automne et il est neigeux en avril <PAD> <PAD> <PAD> <PAD> <PAD> <PAD> <PAD>

Correct Translation:
["new jersey est parfois calme pendant l' automne , et il est neigeux en avril ."]

Original text:
['new jersey is sometimes quiet during autumn , and it is snowy in april .']
```

Evaluation

Number of Epochs	Validation Loss	Validation Accuracy
10	0.912	0.701
20	0.836	0.732
30	0.789	0.768
40	0.623	0.820
50	0.563	0.846

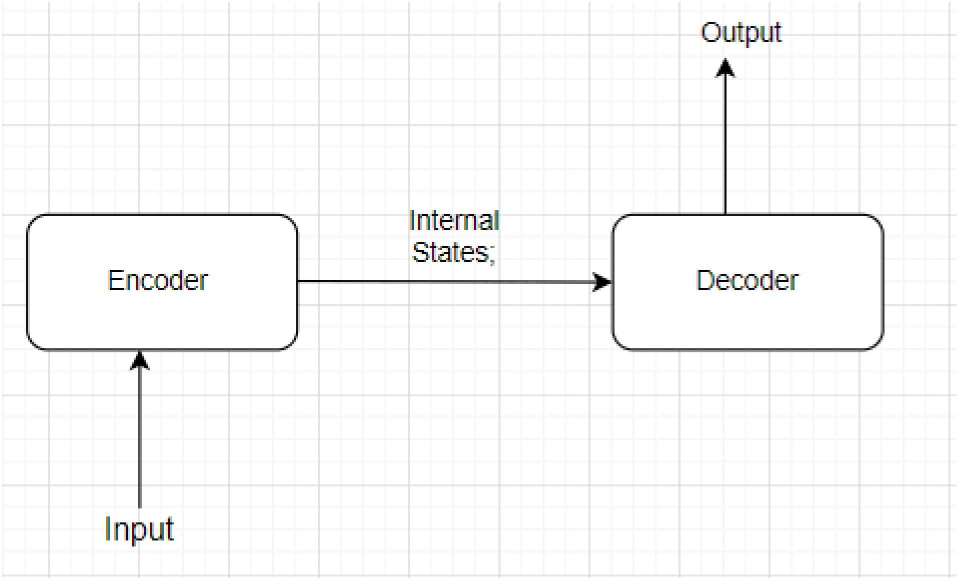




BLEU SCORE:0.32

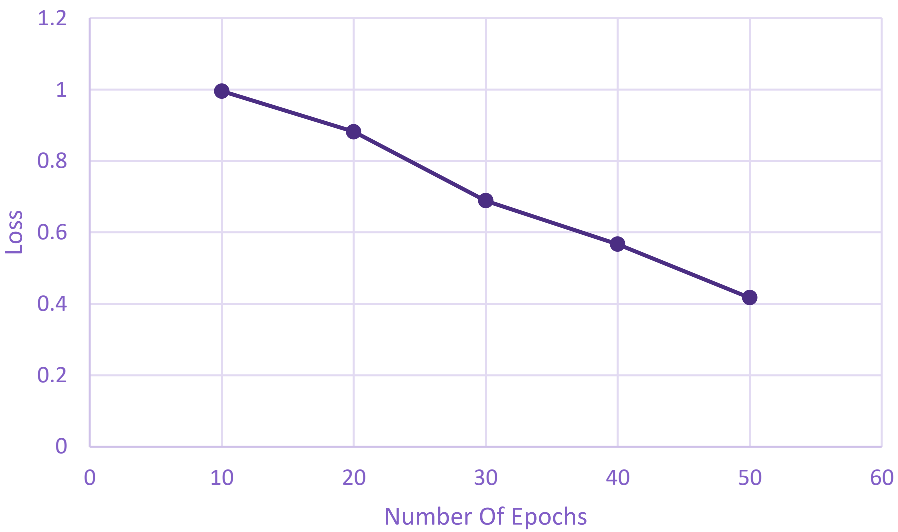
Encoder – Decoder Model

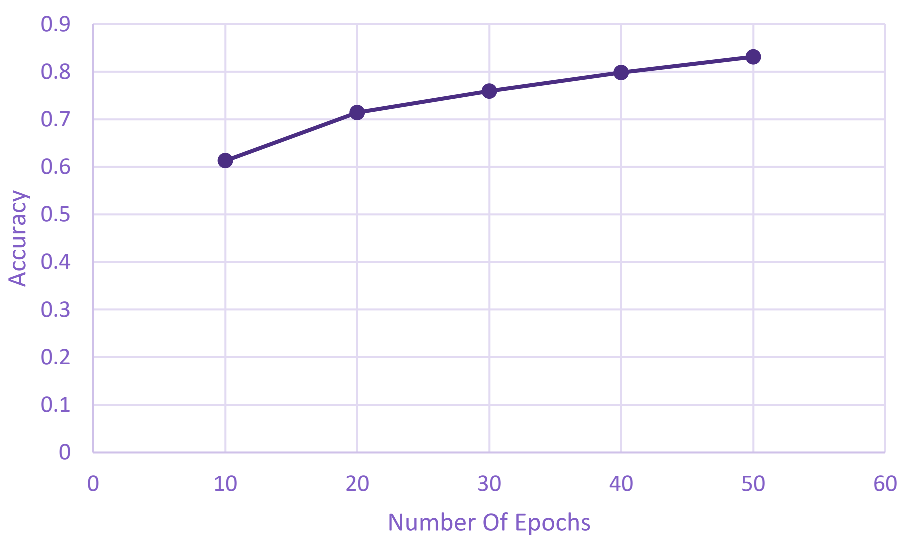
- Contains one encoder which creates the matrix representation of the sentence and one decoder which takes the matrix representation as input and predicts the translation as output.



Evaluation

Number Of Epochs	Validation loss	Validation Accuracy
10	0.996	0.613
20	0.882	0.714
30	0.689	0.759
40	0.567	0.798
50	0.4173	0.8310





BLEU Score : 0.29

Demo link

- https://drive.google.com/file/d/1UdVyXtytl_MjN2lMr87dPBnrK6UhUhB4/view?usp=sharing

Conclusion

- Of all the explored models, RNN (word embedding + bidirectional) has highest translation accuracy and BLEU score

- With increase in number of epochs , slight increase in accuracy and decrease in loss was observed.

Future work

- Various other deep learning models like transformers , LSTM , sequence to sequence learning models can be explored.

- Application can be integrated with video or image captioning or generating subtitles in video in target language.

- Application can further be enhanced to convert text obtained in different languages to speech so that along with subtitles user can also listen the video in different language. Text-to-Speech models can be explored.

- Apart from French ,support for other languages like German , Spanish can be provided.

References

[1] A. Garg and M. Agarwal, “Machine translation: A literature review,” 2018

[2] K. Stevens, G. Kurian, N. Patil, Y. Wu, M. Schuster,Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun,Y. Cao, Q. Gao, K. Macherey, J. Klingner, A. Shah,M. Johnson, X. Liu, Łukasz Kaiser, S. Gouws, Y. Kato, T. Kudo, H. Kazawa, W. Wang, C. Young, J. Smith,J. Riesa, A. Rudnick, O. Vinyals, G. Corrado, M. Hughes,and J. Dean, “Google’s neural machine translation system:Bridging the gap between human and machine translation,”2016.

[3] “Machine translation on wmt-2014 english-French availaible on https://paperswithcode.com/sota/machinetranslation- on-wmt2014-english-french.”

[4] “Wmt-2014 dataset availaible on http://www.statmt.org.”

Thank You !