1. **Introduction**

Machine Translation is one the most important areas in Natural language Processing because we have several languages across the world and it becomes difficult to communicate in a language that we don’t understand. For instance, suppose someone posted a very good article on neural network in a foreign language. If we see this in conventional way, we need to learn that language first before we can read that article. But suppose, how helpful it would be if we have an intelligent learning algorithm at our disposal which can convert text from any language to the language we want. The objective of this project is Neural machine translation from French to English language.

1. **Related Work**

There are several approaches to accomplish Neural machine translation. Recently there has been a transformation in the machine translation by the introduction of deep learning sequential models such as RNN and LSTM. Singh et al. [4] proposed a deep learning model where they proposed usage of different deep learning models for doing different tasks such as usage of Fully Connected Neural Network for word alignment and rule selection, RNN for reordering the sentences Language Modeling and translation.

The use of Encoder-Decoder Sequence model which is well known to remember the long-term context and it is a very important factor for machine translation as the text from source language cannot be transferred word by word because we need to know the context.

1. **Project Topic and Proposed Solution**

Through this project, I am going to perform translation between English and French languages using encoder decoder model. I am going to use the dataset from various sources for better accuracy and performance as cited in [1]. Also, different data cleaning and preprocessing on the text will be performed for making it suitable for the model.  I am going to try multiple different approaches by improving text processing and also calculate the accuracy and loss functions to evaluate the performance. Achieving higher performance and accuracy and error free translation text is the main goal of this project.

1. **Project Timeline**

|  |  |  |
| --- | --- | --- |
|  | Date | Milestone achieved |
| 1 | Jan 8th | Project Proposal submitted |
| 2 | Feb 7th | Project progress submitted |
| 3 | Feb 29th | Final Project due |

1. **Team Roles and Contributions**

|  |  |  |
| --- | --- | --- |
|  | Team Member Name (800 id) | Responsible For |
| 1 | Shilpa Goel(801104237) | Will work on every step starting from preprocessing and cleaning of data to implementation to generate the required results (i.e. language translation) including documentation. |

1. **What have you accomplished so far? What have you changed or updated with respect to your project? Provide a clear description of any issues you faced and how you overcame them and how you changed your strategies.**

Firstly, I read a lot of online material to get in depth knowledge of the Neural machine translation. For data, I have used tab delimited bilingual pair dataset which is in format English, tab space ,French ,tab space, attribution. I have used French language sentences as input and English sentences as output. I have used encoder decoder sequence model using keras library. While installing keras, I faced a lot of issues on my machine like version upgrading and downgrading. Due to larger dataset, I was not able to execute the code on my machine, the kernel crashed, then I switched to google Colab to execute my code. I have restricted my dataset to 35k records approx. for now due to longer execution time and used train-test ratio of (70:30) and epochs 15 to predict the results. I have calculated my BLEU score using corpus bleu function which is giving me accuracy of 65% approx and I was getting varied results on every execution. For the final project, I will increase my dataset and will change the train-test ratios and epochs to calculate the score and try to show the results graphically in code.

1. **References**
2. *[1]Tab-delimited Bilingual Sentence Pairs*

[*http://www.manythings.org/anki/*](http://www.manythings.org/anki/)

[*https://lionbridge.ai/datasets/25-best-parallel-text-datasets-for-machine-translation-training/*](https://lionbridge.ai/datasets/25-best-parallel-text-datasets-for-machine-translation-training/)

1. *[2]Sequence to Sequence modelling*

[*https://www.analyticsvidhya.com/blog/2018/03/essentials-of-deep-learning-sequence-to-sequence-modelling-with-attention-part-i/*](https://www.analyticsvidhya.com/blog/2018/03/essentials-of-deep-learning-sequence-to-sequence-modelling-with-attention-part-i/)

1. *[3]Introduction to Sequence 2 Sequence Learning*

[*https://blog.keras.io/a-ten-minute-introduction-to-sequence-to-sequence-learning-in-keras.html*](https://blog.keras.io/a-ten-minute-introduction-to-sequence-to-sequence-learning-in-keras.html)

1. *[4]S. P. Singh, A. Kumar, H. Darbari, L. Singh, A. Rastogi, and S. Jain, “Ma-chine translation using deep learning: An overview,” in2017 InternationalConference on Computer, Communications and Electronics (Comptelix),July 2017, pp. 162–167*
2. [*https://towardsdatascience.com/neural-machine-translation-with-python-c2f0a34f7dd*](https://towardsdatascience.com/neural-machine-translation-with-python-c2f0a34f7dd)