

Machine Translation form French to English

Sunday Okechukwu
Dhruv Jain

March 9, 2023

1 Introduction

- Machine Translation is the process of translating a source language to a target language using computers. Despite significant advances in Machine Translation over the past few years, the quality of translation still remains a challenge. The primary goal of this project is to improve the quality of Machine Translation using Neural Network Architecture. [6] [5]
- French to English translation in NLP (Natural Language Processing) refers to the process of automatically translating text language using machine learning algorithms and statistical models. This task falls under the category of machine translation, which is a sub field of NLP that focuses on the development of algorithms that can translate text from one language to another. The process typically involves training a machine learning model on a large corpus of parallel French-English texts, which allows the model to learn patterns and relationships between words and phrases in both languages. Once the model is trained, it can be used to translate new French text into English text automatically. [7]

2 Motivation

2.1 Problem Tackling

- What kind of pipeline to use to get the best outcome for translating. Moreover, can we get data for free or we need to use some API's. We are motivated to tackle all this issues using different resources like research papers, kaggle for data acquiring and more.
- Machine translation can significantly reduce the time and cost required for translating documents, websites, and other materials. This is especially true for large volumes of content that would be difficult or impossible for humans to translate within a reasonable timeframe. It is a rapidly evolving field, with new developments and advancements happening all the time. There is still much to be learned and discovered in this area, making it an exciting and intellectually stimulating field of study and research.

2.2 Application or Theoretical Study

- This an application that interest us and will try to apply learning algorithms so it's an **Applied project**. [1]

2.3 Why this problem?

- The most important part of the text translation process is accuracy. The goal of translation is to convey the original meaning of the French text to English language. Therefore, it is crucial that the translated text accurately represents the meaning of the original text. A translator must have an in-depth understanding of the source and target languages, cultural nuances, and subject matter to produce an accurate translation.

- must be able to communicate effectively with clients or other stakeholders to ensure that the translation meets their expectations and requirements. Effective communication can help prevent misunderstandings and ensure that the final product meets the needs of the intended audience.

3 Methods

1. Data Collection: A corpus of parallel English French sentences will be collected for training and testing purposes.
2. Pre-processing: The collected data will be pre-processed by removing irrelevant information such as numbers and punctuations and converting the text to lower case. The pre-processed text will be tokenized into sentences and words.
3. Model Development: A Neural Network-based Machine Translation model will be developed using Deep Learning techniques such as Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) [8] [3], Recurrent Neural Networks (RNN) maybe Encoder decoder . The model will be trained on the pre-processed parallel corpus using Backpropagation algorithm.
4. Model Evaluation: The developed model will be evaluated using standard evaluation metrics such as BLEU and METEOR. The performance of the model will be compared with existing Machine Translation models such as statistical machine translation and rule-based machine translation. [9] [4]

4 Intended experiments

- We are planning to do evolution matrix such as: Metric for Evaluation of Translation with Explicit Ordering (METEOR), BLEU (Bilingual Evaluation Understudy) this are widely used metric for evaluating machine translation systems. We will also use F1-score, accuracy Precision and recall. [2]

References

- [1] Peter Brown, John Cocke, Stephen A Della Pietra, Vincent J Della Pietra, Frederick Jelinek, Robert L Mercer, and Paul Roossin. A statistical approach to french/english translation. In *Proceedings of the Second Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, 1988.
- [2] Jian Liu, Qian Cui, Hongwei Cao, Tianyuan Shi, and Min Zhou. Auto-conversion from natural language to structured query language using neural networks embedded with pre-training and fine-tuning mechanism. In *2020 Chinese Automation Congress (CAC)*, pages 6651–6654. IEEE, 2020.
- [3] KM Tahsin Hassan Rahit, Rashidul Hasan Nabil, and Md Hasibul Huq. Machine translation from natural language to code using long-short term memory. In *Proceedings of the Future Technologies Conference (FTC) 2019: Volume 1*, pages 56–63. Springer, 2020.
- [4] Sarita G Rathod. Machine translation of natural language using different approaches. *International journal of computer applications*, 102(15), 2014.
- [5] Middi Venkata Sai Rishita, Middi Appala Raju, and Tanvir Ahmed Harris. Machine translation using natural language processing. In *MATEC Web of Conferences*, volume 277, page 02004. EDP Sciences, 2019.
- [6] Holger Schwenk, Marta R Costa-Jussa, and José AR Fonollosa. Smooth bilingual n-gram translation-gram translation. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 430–438, 2007.

- [7] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.
- [8] Lirong Yao and Yazhuo Guan. An improved lstm structure for natural language processing. In *2018 IEEE International Conference of Safety Produce Informatization (IICSPI)*, pages 565–569. IEEE, 2018.
- [9] Will Y Zou, Richard Socher, Daniel Cer, and Christopher D Manning. Bilingual word embeddings for phrase-based machine translation. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1393–1398, 2013.