Unlocking Communication: RNN Pipeline for English-to-French Translation

Kavya P M
Roll No: 33
Reg.No: KTE22MCA-2033

Guided By
Prof. Shilpa M Thomas
Department of Computer Applications
Rajiv Gandhi Institute Of Technology, Kottayam

Contents

Introduction

- Globalization demands effective cross-linguistic communication.
- Machine translation, driven by deep learning, facilitates this communication.
- This project constructs a neural network pipeline for English-to-French translation, aiming for optimal accuracy and advancing global linguistic accessibility.

Current state of art

- The current translation approach relies on traditional methods such as statistical models or rule-based systems.
- These methods often struggle to capture the real essence of language, especially in tricky situations.
- The system heavily relies on dictionaries and rules, making it hard to handle all the different ways people talk and the specific words they use.

Motivation

- Accuracy Enhancement: Traditional methods often struggle with nuances; This project aims to improve accuracy using deep neural networks.
- Adaptability: Language evolves; deep learning allows the system to adapt to changes and handle diverse linguistic variations.
- Global Communication: This project aims to enhance translation, fostering greater connectivity and understanding across languages.

Objectives

- To build a deep neural network that functions as part of a machine translation pipeline for English French translation.
- Achieve accurate translation with RNN-based models.
- Implement word embedding layers for relationships.
- Train RNN model for translation.
- Evaluate accuracy and refine iteratively.

Literature survey

Sl No.	Title	Author	Source	Findings
1	Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation	Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre	University of Montreal	Introduces an improved RNN model for English-to-French translation, surpassing basic systems with a score of 34.64.
2	Neural Machine Translation by Jointly Learning to Align and Translate	Dzmitry Bahdanau, KyungHyun Cho Yoshua Bengio	Published as a conference paper at ICLR 2015	Improves translation, especially with longer sentences, by dynamically aligning words without fixed-length encoding.
3	Sequence to Sequence Learning with Neural Networks	Ilya Sutskever, Oriol Vinyals, Quoc V. Le	Google	Superior performance in translation, but struggles with out-of-vocabulary words, affecting accuracy.

Table: Literature survey (Part 1)

Literature survey

Sl No.	Title	Author	Source	Findings
4	Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation	Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi	Google	GNMT implementation outlined with focus on techniques enhancing accuracy, speed, and robustness.
5	Achieving Open Vocabulary Neural Machine Translation with Hybrid Word-Character Models	Minh-Thang Luong and Christopher D. Manning	Stanford University	Introduces hybrid word-character approach for neural translation

Table: Literature survey (Part 2)

Proposed Methodology

- Prepare Data: Clean and organize English-French texts.
- Build Model: Create a smart neural network setup.
- Teach and Test: Train model with examples and evaluate performance.
- Check Accuracy: Translate sentences and compare to known translations.
- Refine and Experiment: Keep improving the model through trial and error.

Architecture

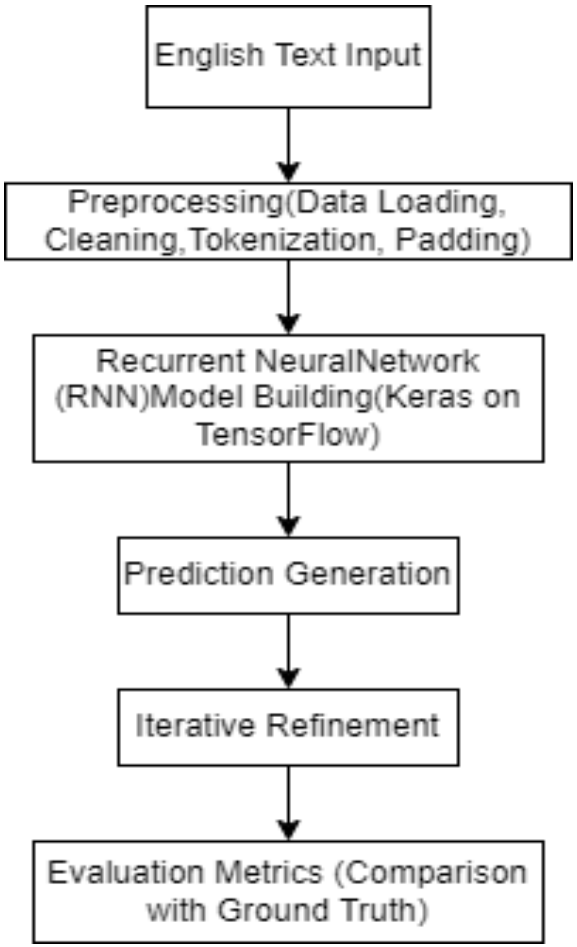

Figure: Architecture diagram

Implementation Status and Plan

Task	Status	Remarks
Data collection	Completed	
Data Cleaning	Completed	
Tokenization	Completed	
Padding	Completed	
Preprocess Pipeline	Completed	
Implementation of Model 1 - simple RNN	Ongoing	ETC March 10th 2024
Implementation of Model 2 - RNN with Embedding	Ongoing	ETC March 15th 2024
Implementation of Model 3 - Bidirectional RNN	Yet to start	ETC March 25th 2024
Implementation of Model 4 - optional Encoder-Decoder RNN	Yet to start	ETC April 1st 2024
Prediction	Yet to start	ETC April 15th 2024

Table: Implementation Status and Plan

Thank you!