Abstractive Summarization of Text

NLP Project

- Reuel Samuel Sam (2018103053)
- Sayf Zakir Hussain (2018103059)
- Karthik Srikanth (2018103035)

Introduction

Automatic text summarization is the task of producing a concise and fluent summary while preserving key information content and overall meaning.

Text Summarization follows 2 methodologies:

- Extractive Summarization
  - The important phrases and sentences are extracted as is from the text
- Abstractive Summarization
  - New sentences are generated as the summary for this given text

This project attempts to provide text summarization for news headlines taken from the InShort dataset

Objective and Dataset

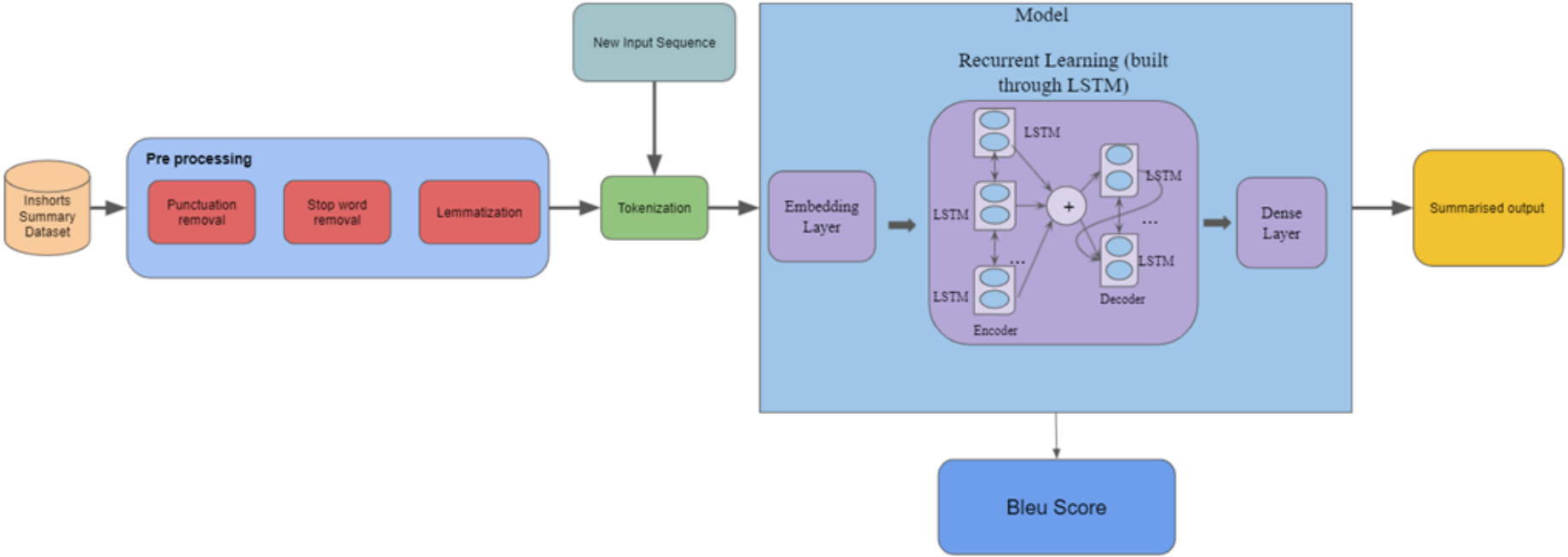
The main aim of this project is to provide Abstractive Text Summarization for news articles through the use of LSTM based Encoder and Decoder structure.

For training purposes, the InShorts News Article Dataset has been identified. Inshorts is a news service that provides short summaries of news from around the web. This dataset contains headlines and summary of news items along with its source.

The dataset contains 55104 Training Examples with 5 Columns (Headline, Short, Source, Time, Published Date)

Dataset Link: https://www.kaggle.com/shashichander009/inshorts-news-data

Methodology Diagram



Performance

- Bleu Score:
  - BLEU, or the Bilingual Evaluation Understudy, is a score for comparing a candidate translation of text to one or more reference translations.
  - Although developed for translation, it can be used to evaluate text generated for a suite of natural language processing tasks.
  - Perfect BLEU Score = 1
- This project achieved a BLEU Score of 0.6854

Performance

Model Type	BLEU Score
Standard LSTM (Encoder-Decoder Structure)	0.56877
Standard LSTM (Encoder-Decoder Structure) with GLoVe Embedding	0.60166
Standard LSTM (Encoder-Decoder Structure) and Attention Layer with GLoVe Embedding	0.68542