Text Generation in Natural Language Processing

Sugam Mankad

U00305388

Contents

- Introduction
- Project Aim
- Background
  - LSTM
  - Text Representations
  - Transformer
  - BERT
  - ELMo
  - OpenAI's GPT-2

Contents

- Design
- Technology Used
- Conclusion & Future Work

Introduction

- Language modeling or natural language generation involves probabilistic analysis that results in predicting next words or sentences given a sequence of text.
- The text data is highly unstructured. Natural language in terms of text or speech is matter of contextualizing information.
- Communication between humans is interpretable by human beings easily. While to make machines learn the natural language, requires specific training of machine learning models.

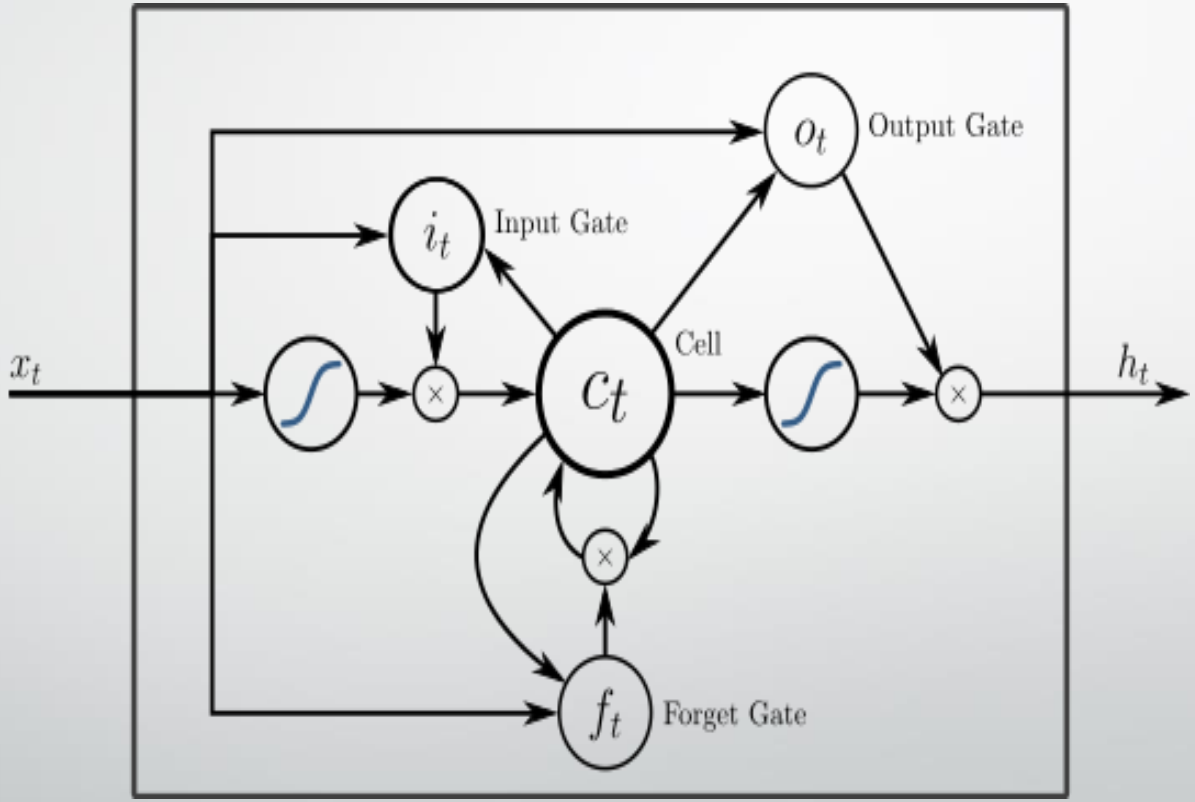
Project Aim

- The project demonstrated aims at generating natural language by using the recurrent neural networks, specifically LSTM – Long Short Term Memory.
- I have created machine learning models for article headlines, novel-stories and poem generation.

Background

- Natural language generation involves text prediction, where text is a continuous or sequential data.
- Since, next word prediction requires dependence in occurrences of a certain words in the sequence, text data training requires a machine learning model which involves previous data being fed as input at every step.
- Recurrent neural network called LSTM – Long Short Term Memory network, is specially used to train a sequential data.

LSTM – Long Short Term Memory



LSTM – Long Short Term Memory

- An LSTM unit consists of a cell, an input gate, an output gate and a forget gate components.
- The cell acts a memory unit to store arbitrary data at every time step while the gates regulate the flow of data.
- The gates perform activation of the weights and perform element wise multiplication and sum.
- The figure shows peephole connections between the components.

Text Data Representations

- Machine learning algorithms can only be applied to numerical data.
- As text is not numerical, training of text data requires pre-processing in terms of representation.
- Over all the years of natural language processing a technique to represent words in form of numbers or vectors was proposed. Commonly known as Word Embeddings, the technique converts words into numbers or sequence of numbers – vectors.

Text Data Representations

- There are numerous algorithms and models that given a text data converts it to vector representations.
- Word2Vec is such a model that was introduced by Google which works on conditional log-likelihood. It generates similar meaning words close to each other in the 2D space.
- The famous example:
  Man – Woman + Queen = King
- Keras' Embedding layer, GLoVe etc. are such examples of such word representation techniques.

Transformer Architecture

- A novel neural network architecture proposed by Google called the transformer works on self-attention mechanism.
- The attention mechanism tries to relate words in a sequence to each other and develop relationships.
- While, a sequential model usually follows either a left-to-right or right-to-left direction of text processing, the transformer compares a word to all the words, in both directions to generate contextual understanding.

BERT – Bidirectional Encoder Representations from Transformers

- It's Google again!
- Google researches used the transformer's attention technique to generate a translator using bidirectional encoding.
- As the name suggests, BERT model process text in left-to-right and right-to-left directions using the self-attention mechanism.
- BERT turned out as a paramount technique for language translation.

BERT – Bidirectional Encoder Representations from Transformers

Input: The man went to the $[MASK]_1$. He bought a $[MASK]_2$ of milk .
Labels: $[MASK]_1$ = store; $[MASK]_2$ = gallon

Sentence A = The man went to the store.
Sentence B = He bought a gallon of milk.
Label = IsNextSentence

Sentence A = The man went to the store.
Sentence B = Penguins are flightless.
Label = NotNextSentence

- BERT uses masked language modeling MLM technique to achieve bidirectional efficiency for contextual understanding of a word.

ELMo – Deep Conceptualized Word Representations

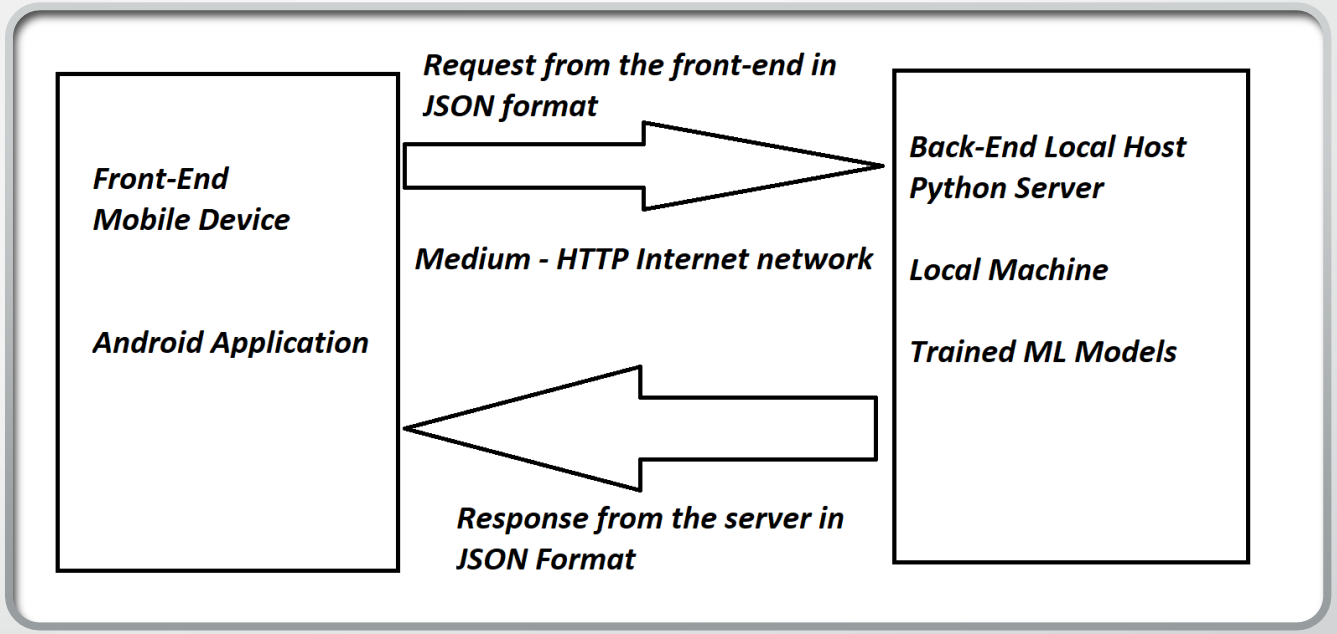
- Similar to the BERT ELMo is a bidirectional network that contextualizes word representations using deep bidirectional language models.
- Unlike BERT, ELMo uses recurrent neural network and not the transformer architecture to generate word embeddings.
- ELMo uses two-layer biLM. A forward LM generates a context-dependent representations in the forward direction, and a reverse LM generates in the reverse direction respectively.

OpenAI's GPT-2

- Unlike both BERT and ELMo, OpenAI's GPT-2 model was created with an aim to simply predict the next word token in a massive text corpus.
- OpenAI created and trained a model by scraping the internet pages, over 40GB of text, and produced a very powerful "weapon".
- In terms of similarity with BERT, GPT-2 also uses transformer architecture as it's backbone.
- GPT-2 model improvises as it goes, however also faces some unusual predictions 50% of the time as explained by OpenAI themselves.
- So, why not create a language model from scratch!

Design

- The project trains 3 different models based on 3 different datasets of article headlines, novel-stories and poems respectively.
- The trained models are saved for reuse and offline prediction.



Technology Used

- I have used various libraries/frameworks as a part of this project.
- For backend and front-end bases, I have created a python API using Flask framework and an Android mobile application in Android Studio respectively.
- Python API calls the offline saved models and predicts next word tokens based on some given input.
- Machine learning model created uses Keras with Tensorflow backend and LSTM layers in model compilation.
- Majority help received from Google Colaboratory!

Conclusion & Future Work

- Bidirectional training of text data and developing contextual knowledge allows machines to understand the natural language. Context-dependent modeling allows preserving the meaning of sentences in a comprehension.
- While there are pros of technology reaching it's peak, OpenAI stats in their blog about misuse of such a large model and violation human ethics.
- Using digital signal processing, continuous speech can be used to create expressions from a single word audio by applying certain learned parameters of several expressions.

Thank You!