Machine Learning Models For Predicting And Modeling More Engaging Business Manuscripts

Aakarsh Sinha (asinha16@earlham.edu)
Department of Computer Science, Earlham College, Richmond, IN

Introduction

Given the high stakes and intense competition within all areas of the industry, intelligent business decisions are more important than ever. Data analysis plays an important role as a critical strategic weapon in business. Neural networks (NN) are one of the most powerful machine learning algorithms that can model and solve approximation, optimization, prediction and classification problems which can be extremely beneficial for businesses. This project presents a basic machine learning Long short-term memory (LSTM) encoder-decoder model that takes in text based advertisements as input, analyzes and restructures them to make them more engaging and readable to the target audience.

Contributions

The following are the major contributions of this research project:

A web scraping tool used to scrape advertisements using the Groupon API.

An analysis tool that uses Coleman-Liau Readability Formula, The Flesch Reading Ease formula and a dataset of active call-to-action words to analyze, score and create a weighted dataset of advertisements.

A weighted dataset of readable and engaging advertisements.

A web-based application that uses the proposed neural network model to restructure business sentences to be more engaging.

A machine learning model to restructure non-engaging advertisements to make them more engaging and readable.

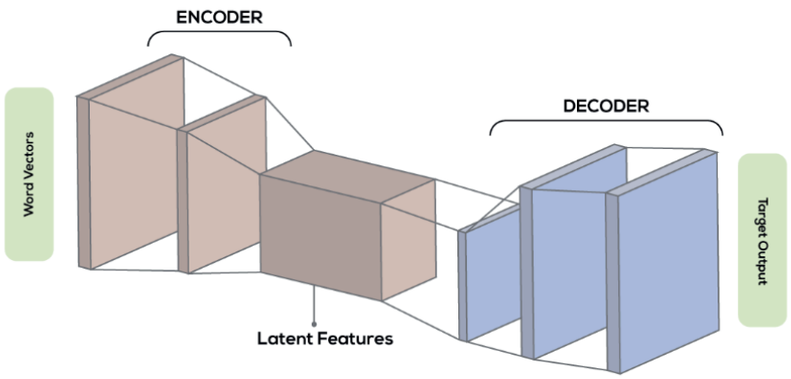


Figure 1: Encoder-Decoder Model

Dataset

The dataset used to train the model is created from advertisements scraped using APIs provided by Groupon.com. The scraped data is then analyzed and scored using Coleman-Liau Readability Formula, the Flesch Reading Ease formula and a dataset of active call-to-action words to create a new dataset of readable and engaging advertisements.

Design and Implementation

The framework of the project is divided into two interconnected elements - the frontend and the backend (See Figure 2). ReactJS is used to create the frontend of the web application that allows the user to upload a text based advertisement as input and get a more engaging and readable advertisement as output. The backend handles the creation, analysis and pre-processing of the dataset. It also contains the neural network which is created using Keras with a Tensorflow backend. The backend also contains the middleware that contains GET and POST APIs that handle http requests that take in advertisements from the frontend.

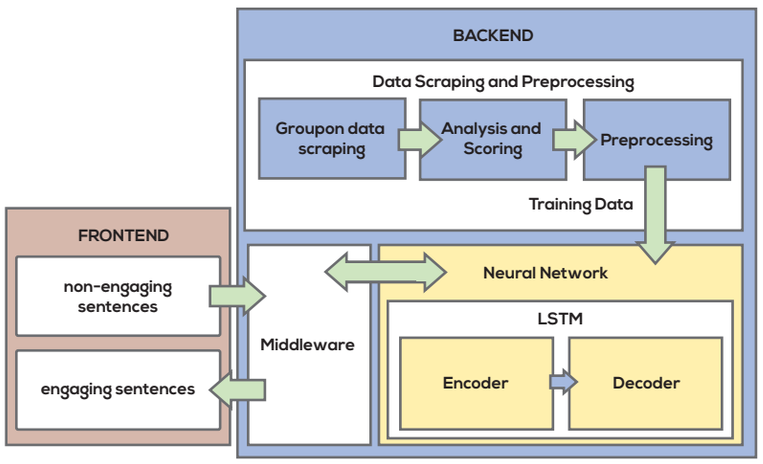


Figure 2: Framework Of The Project

The neural network is made up of an LSTM encoder-decoder model (See Figure 1). The encoder and the decoder are both neural networks, the encoder takes in a sequence of text as an input and outputs vectors and the decoder takes in the encoder's output and tries to reconstruct it. During the training phase, the neural network is fed sentences of advertisements from the engaging and readable dataset and uses a method called "teacher forcing" which forces the decoder to generate sequences and compare it to what the correct output should be.

Results

The accuracy of the model represents how well the model can reproduce an engaging and readable sentence and is calculated by comparing the the similarities between the input to the output of the model during training. The accuracy hovers around 90% during training (See Figure 3) and the model can accurately modify sentence structures it has seen during training (See Table 1) but is not able to convert and modify advertisements with sentence structures that it has not seen during training.

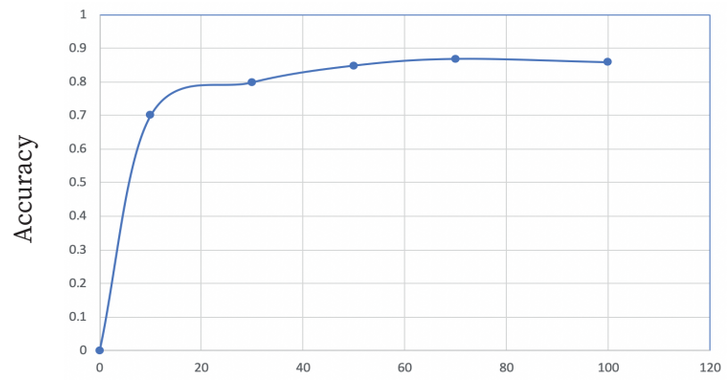


Figure 3: Training Accuracy

Input
Use this guide featuring carefully selected deals from Amazon, starting at just $9.
Output
Check out this guide featuring carefully handpicked deals from Amazon, starting at just $9.

Table 1: Example of an input and output that the machine learning model produces

Conclusion and Next Steps

The study presented a basic encoder-decoder LSTM model aimed at transforming text-based advertisements to make them more readable and engaging. The model achieved an accuracy of 90% while training and was successfully able to modify advertisements with sentence structures that the model had seen during training. Using other variants of LSTM models such as bi-directional LSTM or reduced LSTM present potential avenues for increasing the accuracy of the model.

This poster is based on work described in - *"Machine Learning Models For Predicting And Modeling More Engaging Business Manuscripts" by Aakarsh Sinha*, available at **https://portfolios.cs.earlham.edu/index.php/author/asinha16/**. This project was supported by the Department of Computer Science at Earlham College and would not have been possible without the help, constructive feedback and suggestions of Dr. Charles Peck, Dr. David Barbella, Dr. Igor Minevich and Dr. Xunfei Jiang.