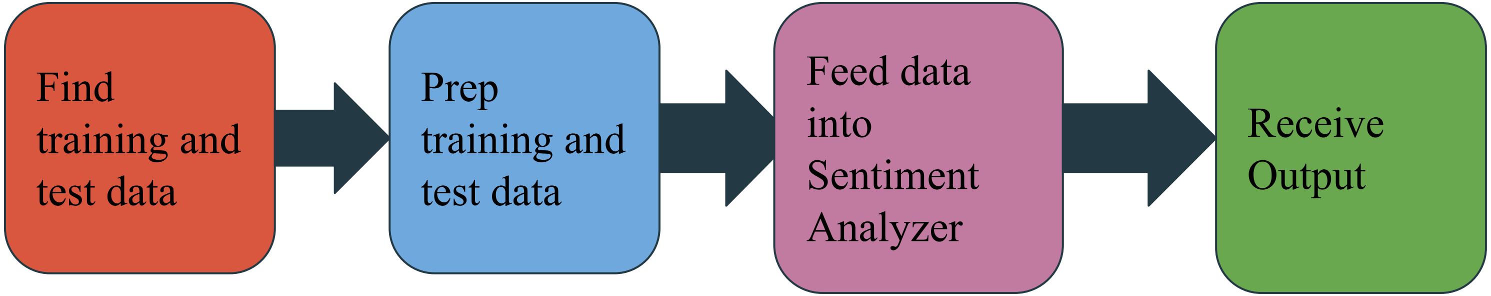
Senior Design Powerpoint Presentation

Anurag Neti

Project Overview

1. For my Senior Design, I decided to build an LSTM Neural Network to perform sentiment analysis.
2. The technology leveraged Python's IMDB movie review corpus to achieve this task.

Project Pipeline



Find training and test data

Prep training and test data

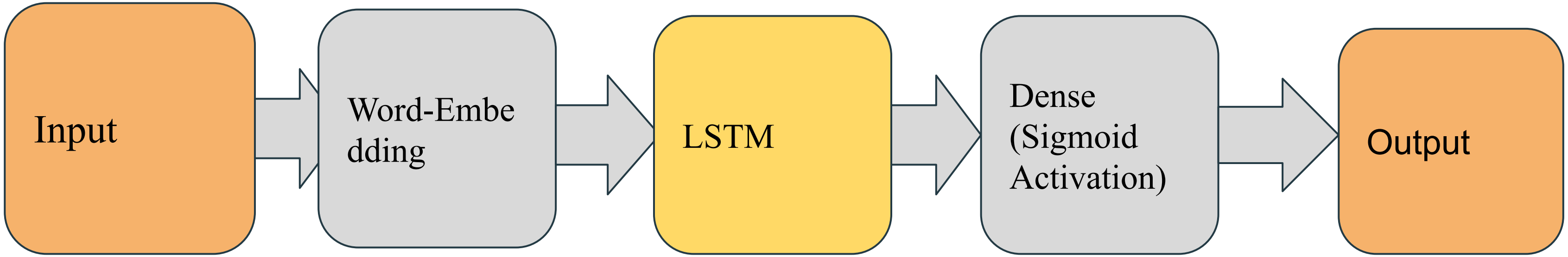
Feed data into Sentiment Analyzer

Receive Output

Choice of Neural Network Architecture

1. Recursive Neural Networks (RNN) work well on sequential data. They, however, fail to adequately perform sentiment analysis because they do not handle long-term dependencies well-as can be the case when performing sentiment analysis-and are susceptible to the vanishing gradient problem.
2. Long Short-Term Memory networks, which are based on the RNN architecture, address this issue because they can selectively remember important inputs-RNNs are incapable of distinguishing between important and unimportant bits of information. The internal structure of an LSTM allows it to remember how important words encountered in the distant past (relatively speaking, of course) affect the weighting of the current words. This detail makes LSTMs ideal for performing sentiment analysis; they can take into account the various nuances of the language they analyze.

Machine Translation Pipeline



The Input Layer

IMDB movie reviews served as Input to the Neural Network.

Word Embedding layer

1. Because Neural Networks only accept mathematical inputs, these movie reviews needed to be converted to a form the network could understand.
2. Word Embeddings are a learned form of vector representation.
3. Python’s Keras library handles the details of converting these reviews into word embeddings.

LSTM Layer

The LSTM layer takes as input the vector representation of words i.e the word embeddings. It learns to classify these reviews as positive or negative.

Dense Layer

1. Every neuron in the input layer is connected to every neuron in the Dense layer. This layer performs intermediate mathematical operations, and feeds the result of these mathematical operations to the output layer.

Output Layer

The output of this layer it a probability between 0 and 1. A probability near to 1 means the review is positive, while a probability close to 0 means that the review is negative.

Challenges faced

1. Research
2. Finding training data
3. Preparing the training data

Research

Primary Sources of research include

1. YouTube (tutorials)

2. Books

Finding Training Data

1. Python has an IMDB corpus containing numerous movie reviews.
2. These reviews served as input to the Neural Network.

Preparing Training Data

1. Since Neural Networks can only process mathematical inputs, these reviews needed to be converted from text to a mathematical form. Python’s Keras library handled this stage of the pipeline by performing word-embedding on the reviews.