CS 6350 Project Status Report

Image and Video Captioning with Deep Neural Networks

• Project Topic –

Image and video captioning with deep neural networks

Or

Recurrent Neural Networks (RNN) for time series prediction (e.g., stock market, weather, hurricane intensity data)

• Team Members:

Harichandana Neralla - HXN210036 | 2021659295

Mrinalini Raghavendran - MXR210111 | 2021599155

• Technique/Algorithm you plan to implement:

Image and Video Captioning with deep neural networks:

Recurrent Neural Networks (RNNs): RNNs are commonly used as the decoder network in image and video captioning. They can take the feature map generated by the encoder network and use it to generate a sequence of words that describe the image or video.

Long Short-Term Memory (LSTM) Networks: LSTMs are a type of RNN that are particularly well-suited to generating natural language descriptions. They can remember long-term dependencies in the sequence of words being generated and generate coherent and grammatically correct sentences.

Or

Recurrent Neural Networks (RNN) for time series prediction (e.g., stock market, weather, hurricane intensity data):

1. One popular technique for implementing RNNs for time series prediction is called "sliding window" or "windowing." This involves dividing the time series data into fixed-length windows, with each window consisting of a sequence of past observations and a target value to predict. The RNN model is then trained on these windows, with the goal of predicting the target value given the past observations.
2. Another technique is to use a "sequence-to-sequence" RNN model, which involves feeding the model a sequence of past observations and having it output a sequence of future predictions. This can be useful for predicting trends or patterns in the data over time.

• Dataset details, such as number of features, instances, data distribution:

Image and Video Captioning with deep neural networks:

MS COCO: This is a widely used dataset for image captioning, which contains over 330,000 images with 5 captions per image, resulting in a total of 1.6 million captions. The images cover a wide range of categories and the captions are written in natural language.

Flickr30k: This dataset contains 31,000 images with 5 captions per image, resulting in a total of 158,000 captions. The images are sourced from Flickr and cover a wide range of topics and categories.

YouTube2Text: This dataset contains a collection of YouTube videos with natural language descriptions. It contains over 1.2 million video segments with 15 words per segment on average.

ActivityNet Captions: This dataset contains videos of various activities, such as cooking, exercising, and dancing, with associated natural language captions. It consists of around 20,000 videos with a total of over 100,000 captions.

Or

Recurrent Neural Networks (RNN) for time series prediction (e.g., stock market, weather, hurricane intensity data):

1. Stock market data: This data set includes historical daily stock prices for a particular company or index, such as the S&P 500. The number of features in the dataset might include the opening price, closing price, high price, low price, and trading volume. The number of instances in the dataset will depend on the length of the time series, which could span several years or decades. The data distribution may exhibit non-stationarity and volatility, with sudden spikes and drops in value.
2. Weather data: This dataset includes historical weather measurements for a particular location, such as temperature, humidity, pressure, wind speed, and precipitation. The number of features in the dataset will depend on the specific measurements collected. The number of instances in the dataset will depend on the length of the time series, which could span several years or decades. The data distribution may exhibit seasonal patterns and trends.
3. Energy demand data: This dataset includes historical energy consumption data for a particular region, such as electricity or gas demand. The number of features in the dataset might include the time of day, day of the week, and month of the year. The number of instances in the dataset will depend on the length of the time series, which could span several years or decades. The data distribution may exhibit periodic patterns and trends.

• Coding language / technique to be used: PySpark

• Preliminary Results (if available)