CSE 582 (NLP) Homework – 2

Name: Sri Krishna Chaitanya Velamakanni

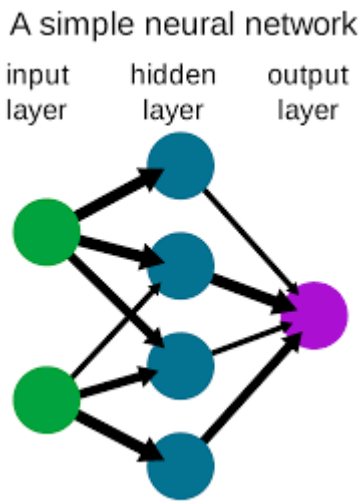
PSU Id: 971110320 (vzs5369)

Date: 04-09-2023

This document describes my understanding of Neural Networks and more precisely convolutional neural networks (**CNN**), long short-term memory networks (**LSTM**) and what I have done in the homework 2 regarding MY attempt in implementing CNN and LSTM and explains the details, challenges faced and learnings.

Firstly, what is a neural network?

A neural network is a type of machine learning algorithm that uses interconnected nodes, or "neurons," to process information. Each neuron receives input signals, which are combined and processed to produce an output signal. The connections between neurons are weighted, and these weights are adjusted during the learning process to optimize the performance of the network. Neural networks are useful for tasks where traditional programming is difficult or impossible, such as pattern recognition in complex data sets.



[1]

There are many different types of neural networks, each with their own unique architecture and characteristics. Here are a few examples:

- **Feedforward neural networks:** This is the most basic type of neural network, where information flows only in one direction, from input to output, without any feedback loops.
- **Recurrent neural networks:** These networks have feedback loops that allow information to flow in both directions, allowing them to process sequential data, such as text or speech.
- **Convolutional neural networks:** These networks are designed to process grid-like data, such as images or videos, and use convolutional layers to extract features from the input data.
- **Autoencoders:** These networks are used for unsupervised learning and can be used for tasks such as data compression or feature extraction.
- **Generative adversarial networks:** These networks consist of two models, a generator and a discriminator, that are trained together to generate new data that resembles a training dataset.
- **Long Short-Term Memory (LSTM) networks:** These networks are a type of recurrent neural network that are designed to overcome the problem of vanishing gradients in standard recurrent networks, allowing them to process long-term dependencies in sequential data.

These are just a few examples of the many different types of neural networks that are used for various tasks in machine learning and artificial intelligence.

How are neural networks used in natural language processing?

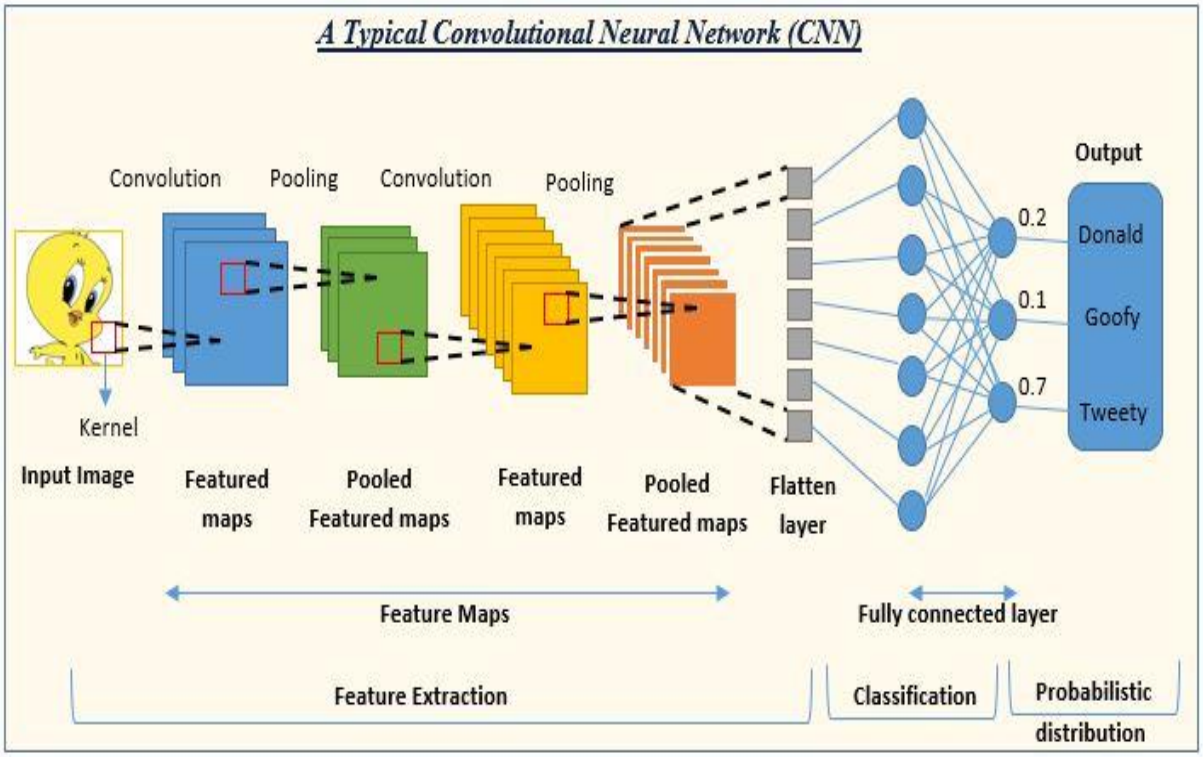
Neural networks are used in natural language processing to analyse large amounts of textual data. They can be trained on text datasets and learn to identify patterns and relationships between words and phrases. Neural networks are commonly used for sentiment analysis, language translation, text classification, and language generation. Text data is pre-processed and the network architecture is designed based on the task and data. The network is trained and weights and biases are adjusted to improve performance, and once trained it can be used to analyses new text data.

Let us confine our focus on convolutional neural networks and long short-term memory networks with respect to this homework.

What is a convolutional neural network?

Convolutional neural networks (CNNs) are a type of neural network used for image classification and recognition. They automatically learn hierarchical representations of input images through multiple layers of feature extraction and abstraction. A typical CNN architecture consists of convolutional layers that apply filters to the input image to extract features, pooling layers that down sample feature maps to reduce complexity, and fully connected layers that classify the input image. The weights and biases in each layer are adjusted during training using backpropagation.

CNNs are effective for image recognition because they can learn to recognize features at different levels of abstraction. They have been used for object detection, face recognition, and medical image analysis, among other applications. CNNs are particularly useful for processing large amounts of image data because they can learn to identify patterns and relationships between pixels in an image, allowing them to accurately classify and recognize images.



[2]

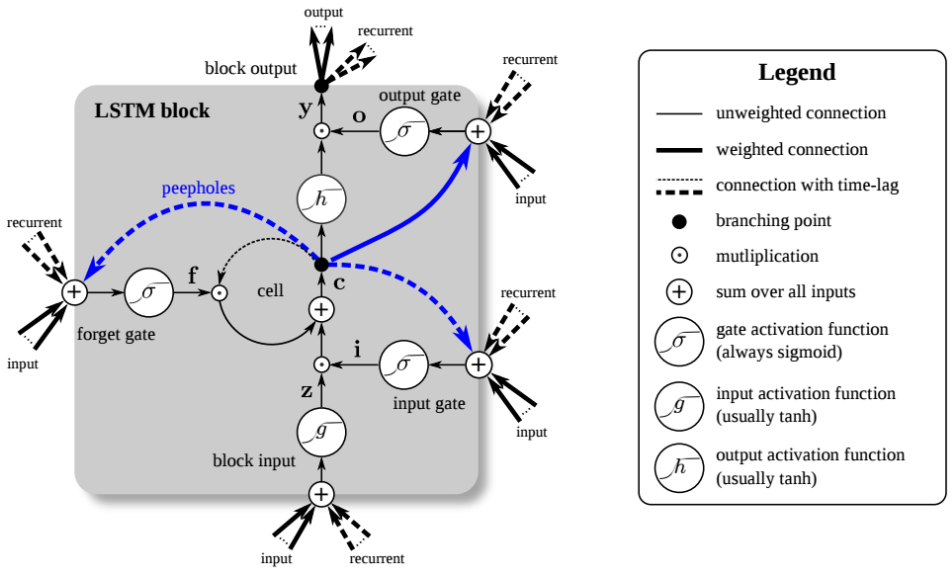
What is long short-term memory?

Long Short-Term Memory (LSTM) is a type of recurrent neural network that is designed to address the problem of vanishing gradients in standard recurrent neural networks. LSTM networks are particularly effective for processing sequences of data, such as time series or natural language text, where long-term dependencies are important.

The key innovation of LSTM is the use of memory cells, which are special units that can maintain information over time. The LSTM architecture includes several gates that control the flow of information into and out of the memory cells. These gates include the input gate, which controls the input of new information into the memory cells, the forget gate, which controls the removal of old information from the memory cells, and the output gate, which controls the output of information from the memory cells to the rest of the network.

During training, the LSTM network adjusts the weights and biases of the gates and the memory cells using backpropagation through time, which allows it to learn to recognize and process patterns in the input sequence. LSTM networks have been used for a variety of tasks, including speech recognition, language translation, and music generation, and have been shown to outperform other recurrent neural network architectures on many sequence-processing tasks.

LSTM Architecture



[3]

CNN Implementation for the given homework

This code is an implementation of Convolutional Neural Network (CNN) for sentiment classification on Yelp movie review data. The following steps are carried out:

- Yelp dataset is loaded in chunks, and the top 100,000 rows are saved to a CSV file. The top 10,000 samples of each sentiment are selected for analysis.
- The stars column is mapped to sentiment using a function, and the sentiment distribution is plotted.
- A function is created to retrieve the top few samples of each sentiment.
- The text column is pre-processed using the genism library's pre-processing functions. Stopwords are removed and tokens are stemmed.
- A train-test split function is created.
- A CNN model is implemented using PyTorch. The model contains two convolutional layers, two max-pooling layers, and two linear layers. ReLU activation is used, and dropout is applied.
- The model is trained on the training dataset and tested on the test dataset.
- The loss and accuracy are plotted to evaluate the model's performance.
- I have used 30 epochs in this implementation and it took me around 45 minutes for training.

Challenges Faced:

- Debugging errors in the code was a challenge as it required a deep understanding of the code structure and how the different components interacted with each other.
- Training the CNN model was time-consuming and resource-intensive, as it required a large amount of training data and computation power.
- Finding an optimal set of hyperparameters for the model was a challenging task as it required experimenting with different combinations and assessing their impact on the model's performance.

Learnings: The overall learning curve was high as I have implemented the neural network on my own and understood the depth of implementation.

Conclusion

The primary objective of this homework was to gain a thorough understanding of neural networks, specifically convolutional neural networks (CNN) and long short-term memory networks (LSTM), and to apply this knowledge to perform sentiment classification on yelp movie review data. This involved exploring and experimenting with various hyperparameters and activation functions to observe the impact on model performance.

In addition to CNN, I also worked on LSTM, and a separate report will be created specifically to document my findings on this topic. Through this homework, I aimed to gain practical experience in implementing and fine-tuning deep learning models and to develop a deeper understanding of their underlying principles and workings.

References

[1] https://en.wikipedia.org/wiki/Neural_network

[2] https://www.analyticsvidhya.com/blog/2022/01/convolutional-neural-network-an-overview/

[3] https://developer.nvidia.com/discover/lstm

I would like to acknowledge that **some** (used only in certain places) of the text generated in this report was done with the help of generative AI technology. More precisely, just to paraphrase my own thoughts in a better wording and to bring it in a good form. **Ideas** are solely my **own**.