Deep Learning vs Machine Learning in Natural Language Processing

Table of Contents

Introduction

Natural Language Processing (NLP)

Natural Language Processing (NLP) is a domain in computer science which is heavily used in processing human language. It is extensively increasingly being deployed in various areas. NLP was first developed by the Cambridge linguistics department (1876), and today NLP is applied extensively in research and teaching. It is particularly used in natural language processing (Naive Bayes, Fast Algorithm Learning, Probabilistic Learning), machine translation (Nuance, Bayesian Algorithm), data analysis (Lets say for example, for linguistic classification). The reason why NLP has emerged with high popularity is that it is simple, fast and cost-effective. In this study we have investigated the efficiency and efficiency from the use of NLP in a large-scale natural language processing (NLP) task. Here we focus on two NLP tasks: (a) the classification of English grammar using human-generated language models , with respect to the word list, and (b) the understanding of language using natural language processing (NLP), with respect to lexical/semantic structure, syntactic patterns, syntax and prosody.

Deep Learning

Deep learning is a derived form of Machine Learning. In typical machine learning engineers try to dig features in datasets in order to develop highly accurate models. In deep leaning the machine does that for us. A deep learning model typically can be seen as a multi-layered artificial neural network. the speciality here is those Neural networks automatically learn the features without pre-defined rules or human domain knowledge. This eliminates the need for labelled data sets. Data set labelling is the biggest problem in Machine learning.

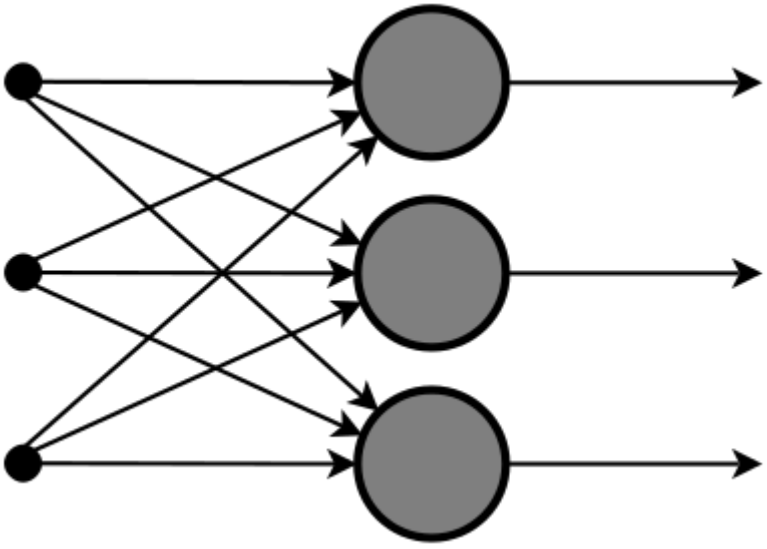
There are six types of Neural Networks currently in use,

1. Feed-forward Neural Network – Artificial Neuron:
2. Radial basis function Neural Network:
3. Kohonen Self Organizing Neural Network:
4. Recurrent Neural Network(RNN) – Long Short Term Memory:
5. Convolutional Neural Network:
6. Modular Neural Network:

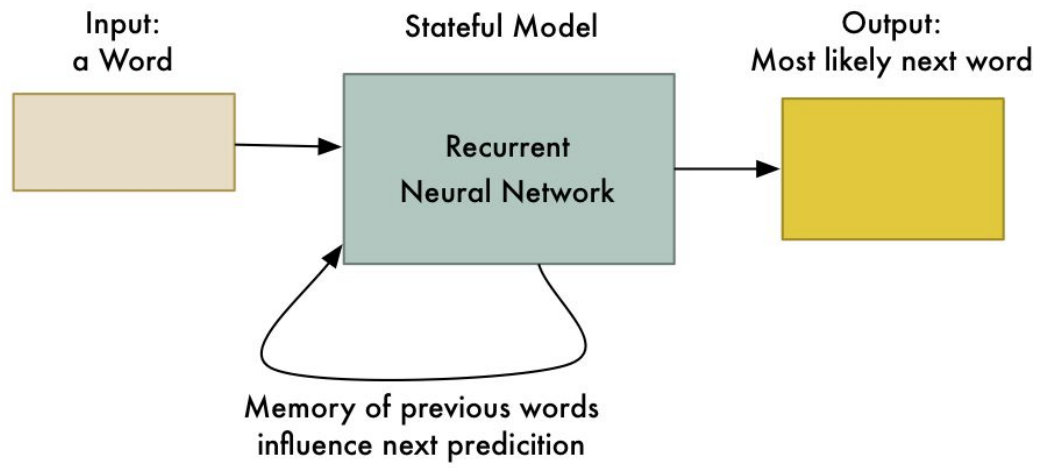
For the purpose of this study, we limit ourself to study the two most popular types of NNs, the Feed-forward type and RNN.

Feed-forward Neural networks are the simplest of all Neural Networks. In such networks all inputs follow the same direction layer-by-layer, node-by-node.

Given below is an image representing a Feed-forward Neural Network.



Recurrent Neural Networks are a type of Neural networks where output is fed back again as the input. This process is called back-propagation. Following GIF is a representation of a simple RNN.



Uses of Deep Learning in NLP

As I said before, there are limitations in ML. Particularly the dataset labelling problem. Deep Learning can be utilized in a various number of ways in NLP. In particular, they can provide an automated way to classify text precisely, without much training. The deep learning method provides significant advantages over traditional methods and its learning accuracy and speed have improved steadily over the years. In addition, because of its relatively small processing power, the deep learning method can easily be trained using a single machine with few processors because it has been trained using a large number of inputs that produce large output. Moreover to this, deep learning techniques can be used to generate datasets which can be useful to train ML models too.

Hate Speech Detection with Deep Learning Model

Hate speech detection in this context is a text classification problem. This text is generated by a human. Human language is a fairly complex thing. Hence its difficult to find features of it to classify it as hate or not. This is the ultimate reason why a deep learning model may perform better than an ML model. As mentioned above there are so many types of deep learning neural networks in use today. Out of those, Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN) are the most effective when it comes to text classification.

Comparing Deep Learning vs Machine Learning implementations

Remaining of this document focus on implementing deep learning for the purpose of predicting hate speech using given dataset and comparing it with a fairly good machine learning-based prediction model. Along with this document,

1. The Jupyter notebook to run the deep-learning python code.
2. An exported HTML page with outputs.

Are given.

Given below is a sample extract from the DL code.

In [17]:
```
from sklearn.feature_extraction.text import CountVectorizer

vectorizer = CountVectorizer()
vectorizer.fit(sentences_train)

X_train = vectorizer.transform(sentences_train)
X_test = vectorizer.transform(sentences_test)
```

Creating and training nueral network

In [18]:
```
from keras.models import Sequential
from keras import layers

input_dim  = X_train.shape[1]

model = Sequential()
model.add(layers.Dense( 10, input_dim=input_dim , activation="relu"))
model.add(layers.Dense(1, activation="sigmoid"))

model.compile(loss="binary_crossentropy" ,
              optimizer = "adam",
              metrics = ['accuracy']
              )
model.summary()
history  = model.fit(X_train, y_train,
                     epochs = 100,
                     verbose = False,
                     validation_data= (X_test, y_test),
                     batch_size = 10
                     )
```

```
Using TensorFlow backend.
```

Our Logistic regression model gave out accuracy of 69% (Section 23 of DeepLearning jupyter notebook.

Deep learning neural network model gave an accuracy of 64%. But there was this overfitting problem with that. (0.993). This could be due to the higher number of epochs.