Evaluation of Impact of Neural Networks in Text Classification

DOI: 10.51201/JUSST/21/07257

Meghana S*
PG Student,
Department of ISE,
M. S. Ramaiah Institute of Technology, Bangalore, India.
1ms19sse10@gmail.com

Jagadeesh D Sai
Assistant Professor,
Department of ISE,
M. S. Ramaiah Institute of Technology, Bangalore, India.
djsai@msrit.edu

Dr. Krishna Raj P. M.
Associate Professor,
Department of ISE,
M. S. Ramaiah Institute of Technology, Bangalore, India.
krishnaraj@msrit.edu

Abstract: *One of the most trending and major areas of research in Natural Language Processing (NLP) is the classification of text data. This necessarily means that the category that the text belongs to is determined by the content of the text. Various algorithms such as Recurrent Neural Network along with its variation which is Long Short-Term Memory, Hierarchical Attention Networks and also Convolutional Neural Network have been used to analyse how the context of the text can be determined from the text data which in available in terms of datasets. These algorithms each have a special characteristic of their own. While Recurrent Neural Network maintains the structural sequence of the contexts, the Convolutional Neural Network manages to obtain the n-gram feature and the Hierarchical Attention Network manages the hierarchy of the documents or data. The above said algorithms have been implemented on the British Broadcasting Corporation News datasets. Various parameters such as recall, precision, accuracy etc. have been considered along with standards such as F1-score, confusion matrix etc. to deduce the impact.*

Keywords: *Natural Language Processing, Text Classification, Recurrent Neural Network, Hierarchical Attention Network, Convolutional Neural Network, Deep Learning.*

1. INTRODUCTION

It can be observed that nowadays, the research outcomes in the field of Artificial Intelligence are such that it is mimicking human like cognition on man-made machines. Natural language processing is an area where humans try to make man-made machines understand the common man language. Understanding of a common man language not only means the literal meaning of the word but also the contextual meaning. This requires common sense from the learning machines to fully understand the perspective of humans through their words. Although understanding various languages is easy for human beings, it is a very complicated process for a man-made machine to achieve this level of accuracy. We can see that a very huge volume of data in the form of text is being produced on a daily basis on the digital platforms. This is also unstructured data which needs to be pre-processed to further perform classification. This is what Natural Language Processing exactly does. It also makes sure that it can obtain the actual meaning of the contexts obtained. This is done to classify the data for operations such as classification of topics, translation of languages, analysis of sentiments, etc. Text classification can be made use of in different areas such as web surfing, classification of mails, filtering of spam mails, document organization, etc.

Machine learning is an approach which is both mathematical as well as statistical. It is used to find solutions for processes which were once thought of to be very difficult. Originally, the machine learning process involves two stages. The first is the feature extraction and the second is to prepare a model for performing the further classification process. One of the divisions of machine learning is deep learning. It is where the machine behaves or replicates the behaviour of the human brain. Deep neural networks are those where there are multiple layers of working for making a decision closest to required output. Deep learning has various advantages such as ability to learn from unlabelled data as well as unstructured data. Scalability, which is a feature where we can construct bigger neural network layers and then eventually give many more data as input for training is another accomplishment that deep learning can achieve. The problem with these original models is that after one point they become stagnant. Recent years have seen a development in algorithms such as recurrent neural network and its modified version being long short-term memory, hierarchical attention network and also convolutional neural network. In this work, these models are used to determine a comparison using various parameters to determine which works more efficiently.

2. LITERATURE SURVEY

In [1], for forecasting the series of times, an architecture of hierarchical neural network has been proposed. There are two levels in this. They make use of maximum likelihood learning competitively. In the first level, there are three experts which make use of the back propagation along with a gating networking in order to separate the inputs so that we can map the vectors of input to the vectors of output. In the second level of this hierarchy, there is the use of fuzzy technology to produce correct trends incoming from the previous level. Predictions show the capability to forecast the changes by classifying the trends correctly.

One of the most used algorithms in today's world is the text classification. Some of the reasons why this algorithm is used include spam detection in emails, classification of various categories in news, retrieval of information or data, analysis of emotions, etc. The earlier classifiers of text were based on machine learning algorithms. They had a lot of limitations such as not enough information, high dimensionality incompatibility, etc. Where as in neural network algorithms, these problems are either reduced to bare minimum or they are nullified. In [2] convolutional neural network has been used in order to classify the datasets.

A language model which is a mixture of convolutional neural network as well as bi-directional recurrent neural network has been proposed in [3] to classify the data of text sources into the character level. In traditional models, there is classifications at the word level. Here in [3] there is avoidance of problems caused by words that are not registered and this in turn helps in improving the strength of representation of texts in character level. The convolutional neural network, by using various filters augments the data whereas the bi-directional neural network gets the contextual data in to and fro directions to divide the text. It has been shown in [3] that this method works better than earlier models.

The earlier classification of texts was majorly done using the machine learning algorithms. This process includes a lot human intervention along with artificially induced data for the sake of training. The main problem in doing this is the scarcity of data and explosion of data in unnecessary dimensions. With the evolution of deep neural learning, text classifying is also being explored. In [4] the natural language processing in classifying text along with bi-directional long short-term memory based convolutional neural network has been proposed. A comprehensive expression has been introduced in order to maintain the semantics and meaning of the word classifying. In news data, [4] has proved to be very helpful.

There is a fast development in using the evolving neural networks. Convolutional neural network and long short-term memory have proved to have better solution in classifying texts. In [5], the concept of convolutional neural network as well as long short-term memory or any of its variant has been used. A classifying model named non applicable convolutional neural network with long short-term memory has been proposed in [5] which does not involve activating functions. This method has proved greater efficiency when compared to the traditional and standard models.

The application of recurrent neural networks has shown amazing results in the domain of natural language processing. This is because of constant repetition in multiplying the recurrent weight matrix. The gradient and vanishing problem along with explosion problem are a common trouble in using recurrent neural network. To overcome these problems, [6] makes use of independently recurrent neural networks which makes all the neurons independent of each other. Along with this, the long short-term memory model is used for obtaining text classification.

Most of the people outside Thailand do not know the Thai language. They also don't provide English subtitles which further complicates the understanding. In [7] a system has been proposed in order to translate this language to English. The combination of MSER and convolutional neural network has been used to achieve this. After localizing the required context in Thai, the result is fed to a seven layered convolutional neural network which has shown to have achieved 98% accuracy. The natural images that are put up across is taken or captured, and the data is localized. In [7], there is also the addition of image classifying.

3. SYSTEM ARCHITECTURE

System architectures gives a conceptual design which further gives us the structure of the design, behaviour of the system and also a deep insight into the design. It gives us an idea as to how the project has been organized. It will contain components that work together to get a desired outcome.

Recurrent neural network is a model that is mainly used for data that is sequential and also there must be a dependency among the different inputs. It is general structure that a hierarchy is formed when we put many words that form sentences and many sentences that form documents and also that they all are interdependent in order to obtain the actual meaning of the text. The recurrent neural network takes the words from each sentence is a sequential manner and it then encodes the detailed information of semantic data by taking into consideration, all the previously accepted words to gain the actual meaning of the context. Various vectors are obtained in this process. The output vector that is obtained at the end of the model in recurrent neural network is basically under the influence of applied weights on the inputs or input vectors. There is also the hidden vector which is a state vector that stores the previous input's contextual meaning. This vector is constantly made for every input and it is later used in the next input's process. The model or algorithm is known as recurrent due to the similar change that is applied over and again to all the input vectors. Even though the same input vector has been passed through the model, there may be various other outputs that are produced owing to the various contextual meaning that the hidden vector may hold. Due to the problem of vanishing and exploding gradients, the traditional recurrent neural network can deal with only a restricted volume of contextual data. Due to this reason, long short-term memory is used which also an improvised version of recurrent neural network to overcome the traditional method's problems. The data that is required for processing the input for classification is selectively remembered. The flow is controlled by the input, forget and output gate which are in the long short-term memory model. This model's capability to hold long term dependencies can be made use of in categories such as generation of texts, recognition of handwritten texts etc.

Although convolutional neural networks are originally created for computer vision-based applications, it has shown great results even in the field of natural language processing. The model of convolutional neural network is similar to that of a feed forward network in neural networks. Various kernels are made use of in order to obtain n-gram features and hence produce feature maps of various dimensions depending upon the whole volume of inputs and dimensions of kernels. Another feature that is performed in this layer is the pooling. Or also the down-sampling. Here, output vectors obtained from various kernels of various dimensions are combined by making cautious choice if maximum values of every feature map. This is done and a single column of vectors are obtained. The dimension of the vectors is derived by the whole count of them in the layers of convolution. The n-grams are obtained by these convolution layers which is an extraction feature. The obtained outputs are given to the following layers in order to perform further prediction in the classification.

As a common following, the words together form sentences and sentences together form documents. Due to this presence of hierarchical patterns, recently the hierarchical attention networks were obtained. An important point to consider is the fact that not all words are of equal importance in order to obtain the meaning of the document. In earlier approaches, a token keyword was assigned to each of the document. Later classification was made of this basis. The huge disadvantage here is that, the sequential formation of files will be overlooked and therefore lost. As of the recurrent neural network, it maintains the sequential formation but overlooks the main token keyword meaning of the document. The hierarchical attention network consists of the following two divisions. A bidirectional recurrent neural network which gets an idea as to what could be the meaning of the sentence that has been given as input by making use of the input, output and hidden vectors. Next is the attention network itself. It is where the weight of all the words is worked upon and decided. This is on the basis of importance that each word has been assigned. A similar process is conducted to obtain the sentence vectors. On a whole the document vector is thus obtained and further passed on for classification.

3.1. RECURRENT NEURAL NETWORK ARCHITECTURE

The recurrent neural network architecture along with long short-term memory has four layers. They are: input layer, embedding layer, long short-term memory layer, output layer. The architecture is shown in Fig. 1.
INPUT LAYER: In this layer the input is given. It consists of unique words and is depicted by w1, w2, …… , wT, where T is the total count of the different words in the dataset. These are converted to the integer format by making use of tokenizers. Then they are put onto a dimension say d by making using of embedding.

EMBEDDING LAYER: In this layer, the weights fine tune the inputs. The order of integers is given to this layer and there is representation of words through this input.

LONG SHORT-TERM MEMORY LAYER: In this layer, the data that is stored in the input is encoded. The contextual data is represented in this step. Recurrently connected storage cells are present which contain the related data from the context and is able to store the dependability criteria by using various gates which are input gates, output gates and forget gates.

OUTPUT LAYER: The output of the long short-term memory is given as input to the output layer where the activation function softmax is used to obtain the normalized values. The objective here is to reduce the cross-entropy or the not similar estimated values.

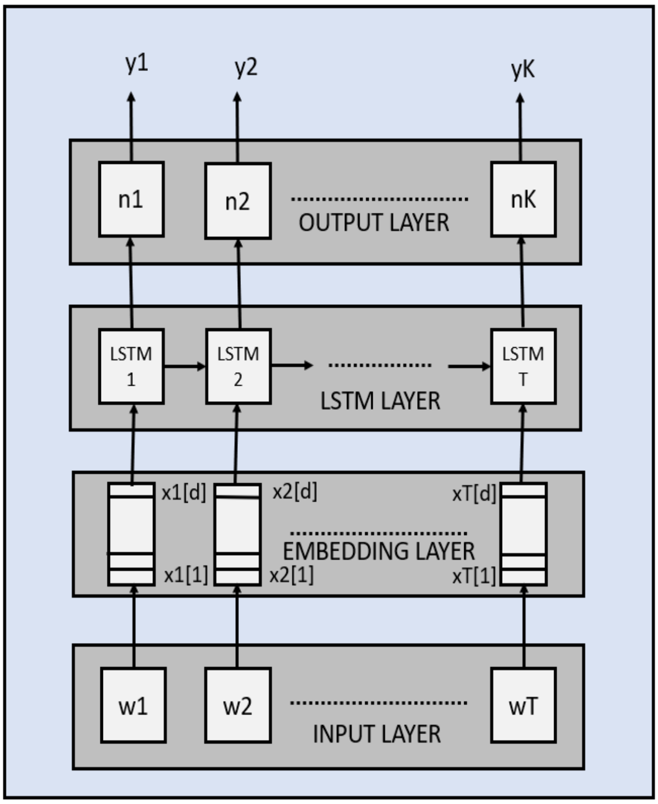


Fig. 1. Recurrent neural network along with long short-term memory architecture

3.2. CONVOLUTIONAL NEURAL NETWORK ARCHITECTURE

A single layer convolutional method is used. Just as in the long short-term method, the vector representation of the words that are fed as input will be generated using a previously defined embedding called GloVe. The convolutional operation is done by using various different kernels or filters which are differentiated based on various different mathematical equations for classifying. This procedure generates the attribute vectors of various dimensions for the input. This layer of input will be given to the output layer in order to perform the further process of classification. The architecture is shown in Fig. 2.

CONVOLUTIONAL LAYER: the vectors that come in as input and the various filters of the convolutional model have a height and width that is same as the length of the output of the embedding layer. In the convolutional layer, a convolutional operation in which a filter which is of a certain length and certain height with back-to-back works will be performed. Functions such as ReLu, sigmoid etc. can be used here. The feature detector makes sure to capture a feature of the input text which is nothing but the n-gram.

MAXPOOLING: the height of the filter that has been used in the convolutional layer will be the deciding factor in the dimension of the attribute vectors. In this operation of maxpooling, a maximum instance of some fixed number will be finalized from each of the attribute vector that has been given out as an output from the previous layer. An n-dimension attribute vector will be produced after it goes through various filters in this layer. This is later given out as an input for the output layer.

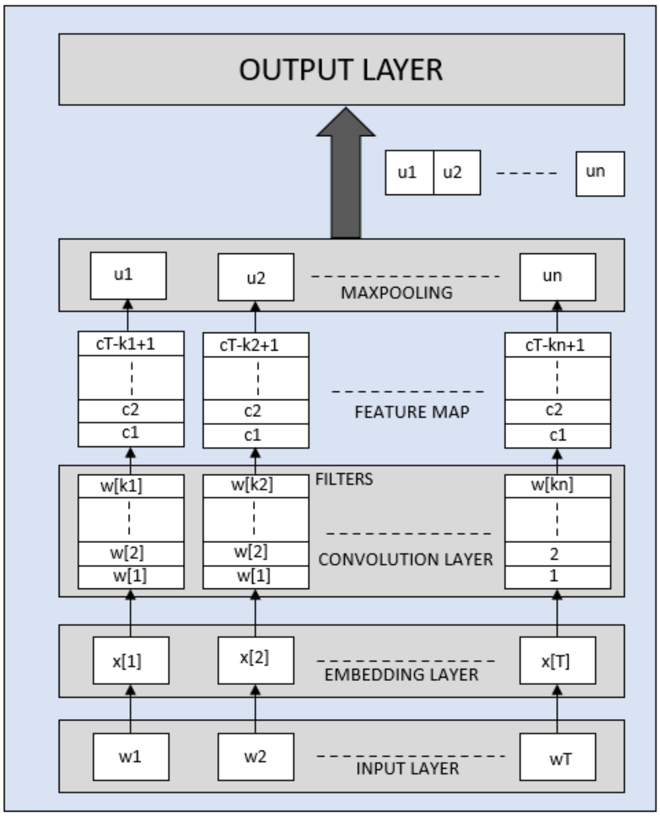


Fig. 2. Convolutional neural network architecture

3.3 HIERARCHICAL ATTENTION NETWORK ARCHITECTURE

The mechanism of attention has been approached in two different levels. In earlier models. The words in sequence were as an input. In this, the hierarchy of the document is taken. Which means it takes the order that the words form a sentence and a sentence forms a paragraph and so on. The representation in the vector form is obtained using the embedding GloVe technique. The hierarchical attention network has different layers in its architecture. They are: input layer, embedding layer, attention for sentences, attention for words, sequence encoder for the word level, sequence encoder for the sentence level and then in the end is the output layer. The architecture is shown in Fig. 3.

BI-DIRECTIONAL LONG SHORT-TERM MEMORY FOR WORD LEVEL LAYER: Two long short-term memory layers are made use of. One goes in the forward way and the second goes in the backward way.

ATTENTION LAYER FOR THE WORD LEVEL: For every word, a weight will be calculated. This will represent the value of that word in a particular sentence. The vector for sentences will be calculated by making use of these weights.

BI-DIRECTIONAL LONG SHORT-TERM MEMORY FOR SENTENCE LEVEL LAYER: a sequence is entered in the document and it will be sent to the bi- directional long short-term memory layer. A combination of representations of contexts are obtained for all the sentences. This is so done by taking in to account all the sentences that are around which are being sent back and forth.

ATTENTION LAYER FOR THE SENTENCE LEVEL: From the vector of sentences, the vectors of the documents can be derived. A method which is same as in the attention for word level layer has been used to get the value of weights. Random initializing of the vectors of context will give the representation the documents. The normalization is done by using an activation function for example, the softmax function. This is further given to the output layer in order to carry on further classification.

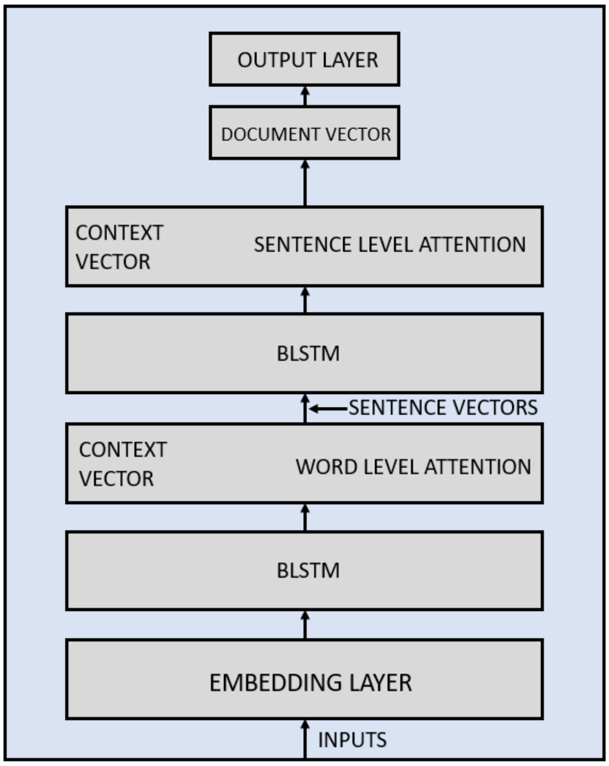


Fig. 3. Hierarchical attention neural network architecture

4. PERFORMANCE AND EVALUATION

4.1 DATASET

The recurrent neural network algorithm, convolutional neural network algorithm and hierarchical attention network are evaluated by projecting them on the British Broadcasting Company News dataset. The data set has 2225 collective news columns or articles. They have been further classified as business news, technology news, sports news, political news and entertainment news. The data is split into 10 percent for testing data and 90 percent for training data. The three algorithms work on the 90 percent of the data. Then the remaining 10 percent of the data will be used for the sake of validation. This experiment or project is done by making use of the python language along with the tensorflow framework of deep learning.

The dataset has been subjected to pre-processing in order to eliminate stop words and also some special symbols that do not convey any particular meaning to the sentence. Further this pre-processed data is tokenized by making use of the method of Stanford tokenizer. For the long short-term memory and the convolutional neural network algorithms, the inputs that are given to the model is accepted as a sequence of words or data. The upper limit of this is taken as 300 words. This is after the process of truncating and padding has taken place. The input that is given to the hierarchical attention network should be given in the form of sentences which is the sequence of data or words. The number of words in each sentence has been given a lower limit of 15 words and an upper limit of 100 words. Word vectors will be produced and they will be updated as and when the model is being trained. It further set true to all values of trainable.

4.2 EVALUATION MEASURES

4.2.1 LONG SHORT-TERM MEMORY

The validation dataset [8] is used to tune the hyper-parameters. A stochastic gradient descent algorithm has been made use of in order to enable continuous updating of the network parameters. A learning rate of 0.001 has been selected for the process. The main reason for making use of the cost function is to reduce the categorical cross entropy as much as possible. The regularization parameter for the l2 of weight vector is finalized for 10 to 5. Other way for regularization that has been used is the dropout mechanism. This essentially means that a part of the hidden units is under deletion or are being dropped out. This is done in order to avoid a coadaptation leading to overfitting problem regarding the hidden layers. For the embedding layer as well as the long short-term memory layer, a rate of about 0.2 is taken as the dropout. The hidden layer in long short-term memory has a size of about 200.

One of the metrics that has been used for the analysis of performance for the classification of data is the confusion matrix. This sums up a summary as to what are the labels that are actual and what are the labels that are predicted for every input vector. The Table- I shows the very same long short-term memory's confusion matrix. The 1st row implies that out of the 60 business labelled input vectors that are passed, 7 of them are predicted wrong and the rest 53 of them have been predicted right. In the five wrongly predicted inputs, 2 of them have been classified as technology, 2 of them as sports and the other 3 of them as political news. Going about the rest of the table, 35 technology labelled inputs have been predicted right, 50 sports labelled inputs have been predicted right, 29 politically labelled inputs have been predicted right and 36 entertainment labelled inputs have been predicted right.

Table- I: Confusion matrix of long short-term memory

	BUSINESS	TECHNOLOGY	SPORTS	POLITICS	ENTERTAINMENT	
BUSINESS	53	2	2	3	0	PREDICTED
TECHNOLOGY	2	35	2	0	0	
SPORTS	0	0	50	2	2	
POLITICS	0	2	1	29	0	
ENTERTAINMENT	0	0	0	2	36	
ACTUAL						

Some of the other parameters that have made use of is the accuracy which essentially the number of right predictions made by the model. Loss is another parameter which gives the difference between the correct output and the predicted output. The rates of both accuracy and loss have been calculated.

Table- II shows the results that the other attributed have achieved while analyzing the long short-term memory neural network. The precision parameter means the ratio count of observations that have been predicted right to the count of total observational predictions which includes those that are both right and wrong. Recall is another parameter that has been used for the sake of evaluation that can be defined as all the true predictions against total count of samples in the dataset. F1 score parameter gives a calculation of the weight induced average of the previously used parameters which are recall as well as precision.

F1 score = (2*(recall*precision))/ (recall + precision)

Table- II: Classification report of long short-term memory

CATEGORY	PRECISION	RECALL	F1SCORE	SUPPORT
BUSINESS	0.97	0.91	0.94	61
TECHNOLOGY	0.92	0.97	0.96	37
SPORTS	0.97	0.93	0.94	56
POLITICS	0.89	0.97	0.96	31
ENTERTAINMENT	0.95	0.93	0.91	36

4.2.2 CONVOLUTIONAL NEURAL NETWORK

Some of the important hyper parameters that are required for the convolution neural network model are the total count of filters as well as the size of the kernel. In this project, one dimensional convolution along with one dimensional maxpooling has been used. A size of 32 has been taken for the filters with 8 being the size of the kernel. This gives the length of the window in convolution step. The network will be subjected to perform with a motive of minimizing as much as possible, the loss function. The dropout method has been used in various steps such as maxpooling step, convolution step, and also the embedding step. 0.2 is the rate of dropout that has been taken for regularizing. For the maxpooling layer, 2 has been taken as the size of the pool. Relu function has been used for fulfilling the need of activation function. The confusion matrix thus derived has been shown in Table- III.

	BUSINESS	TECHNOLOGY	SPORTS	POLITICS	ENTERTAINMENT	PREDICTED
BUSINESS	57	0	0	2	1	
TECHNOLOGY	0	36	0	0	3	
SPORTS	0	0	54	0	0	
POLITICS	0	0	0	32	0	
ENTERTAINMENT	0	0	0	0	38	
ACTUAL						

Table- III: Confusion matrix for convolutional neural network

It can be noticed that for the label sports, political and entertainment, every sample has been predicted right. Therefore, their result for recall is 1. For the label technology, sports, business again there have been no such wrong predictions and hence the calculated precision results in 1. Naturally, the F1 score will become 1 due to the precision and recall resulting in 1. These details have been provided in Table- IV. The accuracy as well as loss have been found out for these parameters.

Table- IV: Classification report of convolution neural network

CATEGORY	PRECISION	RECALL	F1SCORE	SUPPORT
BUSINESS	1	0.94	0.93	61
TECHNOLOGY	1	0.91	0.96	37
SPORTS	1	1	1	56
POLITICS	0.98	1	0.95	31
ENTERTAINMENT	0.91	1	0.94	36

4.2.3 HIERARCHICAL ATTENTION NETWORK

The output dimension coming from the bidirectional long short-term memory has been set to 300. Again 300 has been set for the context vector dimension as well which includes both the attention at the sentence level as well as the attention at the word level. The resulting output is forwarded to the attention layer in the form of encoded and sequential manner. This is done for every word in the sentence and hence focusing on the word level as well as the sentence level. To do all this, a time distributed method has been used. This method uses a dense layer which has been set to 200. The resulting confusion matrix has been depicted in Table-V.

	BUSINESS	TECHNOLOGY	SPORTS	POLITICS	ENTERTAINMENT	PREDICTED
BUSINESS	45	0	0	0	2	
TECHNOLOGY	0	40	0	0	0	
SPORTS	0	2	40	1	0	
POLITICS	0	0	0	52	0	
ENTERTAINMENT	0	2	0	0	38	
ACTUAL						

Table- V: Confusion matrix of hierarchical attention network

The results of the confusion matrix show that recall for technology as well as for politics has been 1, which means they all have been predicted right. The precision for the labels including data of business and sports is also 1 which means that none of them were predicted wrong. An elaborated depiction is shown in Table- VI.

Table- VI: Classification report of hierarchical attention network

CATEGORY	PRECISION	RECALL	F1SCORE	SUPPORT
BUSINESS	1	0.95	0.94	47
TECHNOLOGY	0.99	1	0.94	40
SPORTS	1	0.96	0.97	43
POLITICS	0.96	1	0.98	52
ENTERTAINMENT	0.97	0.96	0.97	40

5. RESULTS AND DISCUSSIONS

The experiment has been carried out three different datasets belonging to the British Broadcasting Corporation [8]. It was observed that when the dataset 1 was passed through all the three neural networks, all of them showed good accuracy although long short-term memory and hierarchical neural network were slightly better shown in Fig. 4. The time taken by the three algorithms however, were drastically different from each other. Hierarchical attention network took the most amount of time whereas convolutional neural network took the least amount of time as shown in Fig. 5.

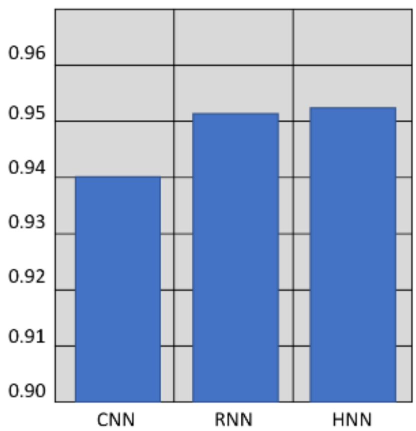


Fig. 4. Accuracy of the three algorithms on dataset 1

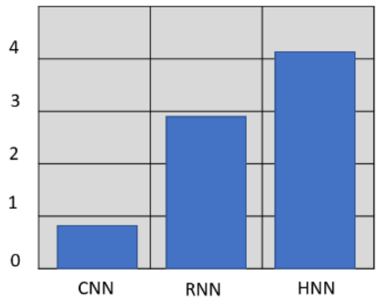
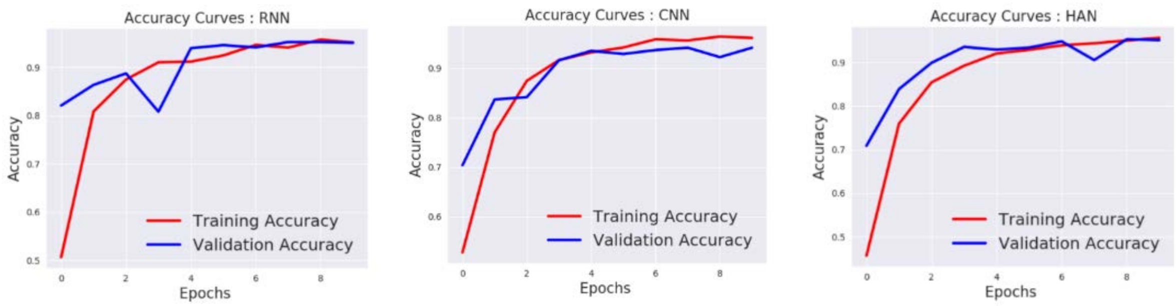


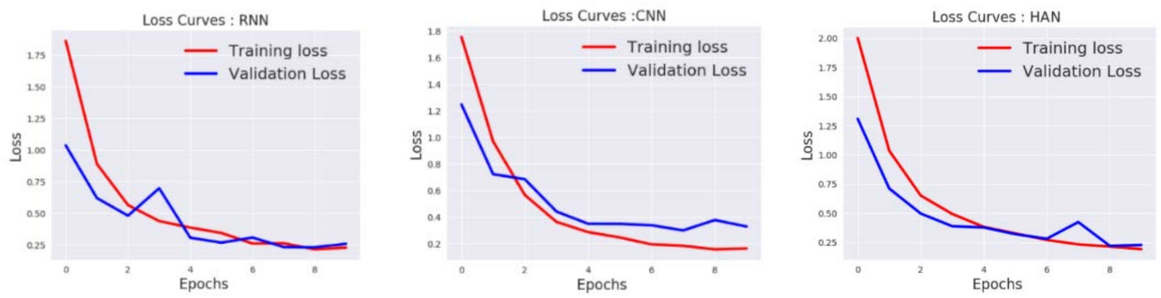
Fig. 5. Time taken(epoch) in minutes by the three algorithms for execution of dataset 1

The accuracy of dataset 1 has been depicted in Fig. 6(a), 6(b), 6(c) and loss data plots for the dataset 1 are depicted in Fig. 7(a), 7(b), 7(c).



(a) (b) (c)

Fig. 6. Accuracy curve for dataset 1 using (a) recurrent neural network, (b) convolutional neural network and (c) hierarchical attention network.



(a) (b) (c)

Fig. 7. Loss curve for dataset 1 using (a) recurrent neural network, (b) convolutional neural network and (c) hierarchical attention network.

It can be observed that when the dataset 2 was passed through all the three neural networks, the accuracy of each algorithm varied drastically. While hierarchical attention network performed best on this dataset, long short-term memory showed the least accuracy as shown in Fig. 8. The time taken by the three algorithms also varied highly from each other on this dataset. Long short-term memory took the most time in order to execute the model’s output whereas like for dataset 1, the convolutional neural network again took less time for dataset 2 as shown in Fig. 9.

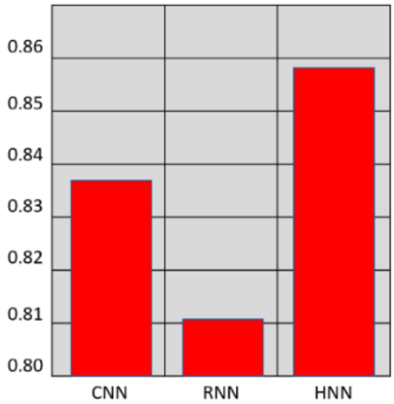


Fig. 8. Accuracy of the three algorithms on dataset 2

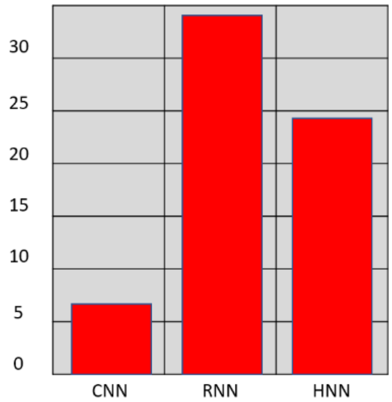
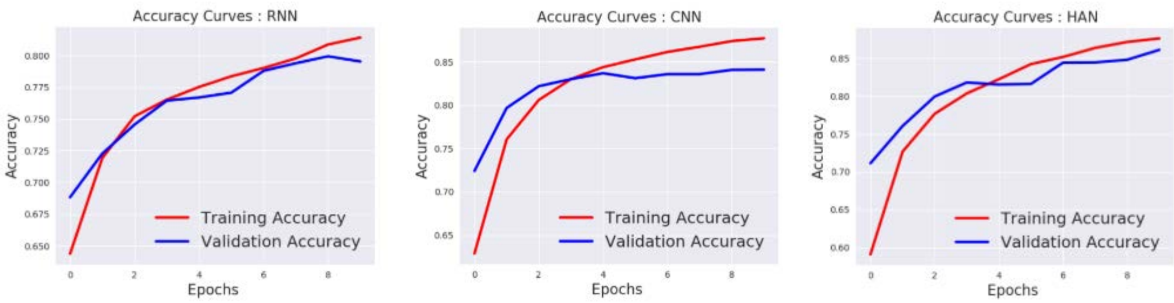


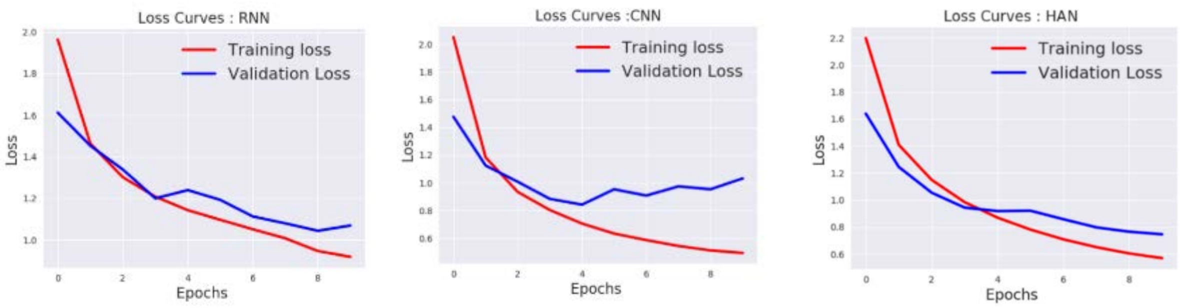
Fig. 9. Time taken(epoch) in minutes by the three algorithms for execution of dataset 2

The accuracy of dataset 2 has been depicted in Fig. 10(a), (b), (c) and loss data plots for the dataset 2 are depicted in Fig. 11(a), (b), (c).



(a) (b) (c)

Fig. 10. Accuracy curve for dataset 2 using (a) recurrent neural network, (b) convolutional neural network and (c) hierarchical attention network.



(a) (b) (c)

Fig. 11. Loss curve for dataset 2 using (a) recurrent neural network, (b) convolutional neural network and (c) hierarchical attention network.

It was observed that when the dataset 3 was passed through all the three neural networks, the convolutional neural network gave the highest accuracy whereas hierarchical attention network gave the least accuracy as shown in Fig. 12. The time taken by the three algorithms also varied extremely from each other on dataset 3. Long short-term memory took the most time in order to execute the model's output whereas the convolutional neural network hardly took time for this dataset which is shown in Fig. 13.

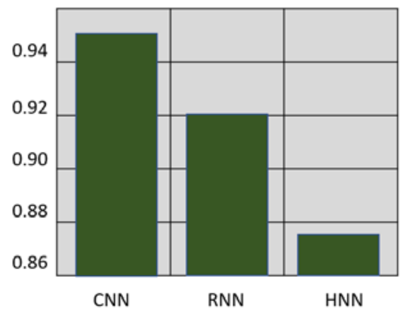


Fig. 12. Accuracy of the three algorithms on dataset 3

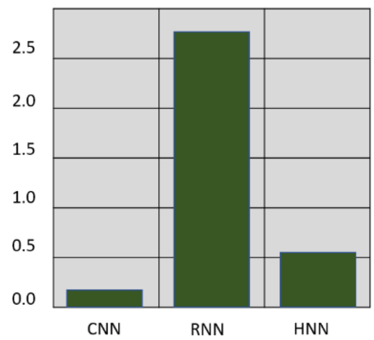
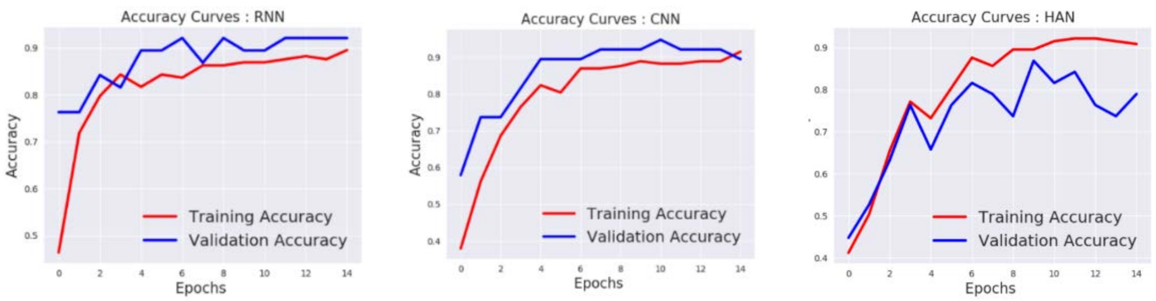


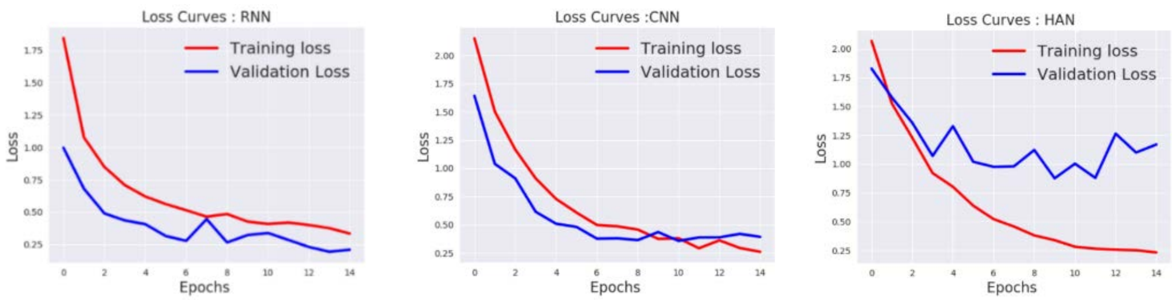
Fig. 13. Time taken(epoch) in minutes by the three algorithms for execution of dataset 2

The accuracy of dataset 3 has been depicted in Fig. 14(a), (b), (c) and loss data plots for the dataset 2 are depicted in Fig. 15(a), (b), (c).



(a) (b) (c)

Fig. 14. Accuracy curve for dataset 3 using (a) recurrent neural network, (b) convolutional neural network and (c) hierarchical attention network.



(a) (b) (c)

Fig. 15. Loss curve for dataset 3 using (a) recurrent neural network, (b) convolutional neural network and (c) hierarchical attention network.

6. CONCLUSION

There are ample number of algorithms invented as of today and more to come that are used for text classification under natural language processing. These algorithms are achieving man level of thought process. All these algorithms have their specific attribute that provides uniqueness in the classification. But, due to the vast number of algorithms, it is difficult for one to choose which algorithm is more suitable for his or her dataset's classification. This project is a step towards solving this problem. We have compared recurrent neural network along with long short-term memory, convolutional neural network and hierarchical attention network algorithms in this project. These algorithms have been applied on the British broadcasting company datasets. We use the sequential feature of recurrent neural network and its variant long short-term memory, the n-gram feature provided by convolutional neural network and the hierarchical feature provided by the hierarchical attention network on the datasets. Upon doing this, various parameters such as accuracy, recall, precision etc. are obtained and a detailed comparison of the same is done. This would help for future references as to which algorithm would provide what advantages and hence which would be best suitable for the particular dataset under classification process. In the future, many more algorithms and various measures of evaluation can be added to enhance the richness of the project.

7. ACKNOWLEDGMENT

I would like to express by sincere gratitude towards the Department of Information Technology, Ramaiah Institute of Technology and the management for providing me the environment to conduct my Masters project with world class facilities.

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

http://mlg.ucd.ie/datasets/bbc.html

REFERENCES

[1] M. A. Ozbayoglu, C. H. Dagli and B. Fulkerson, 1994, “A hierarchical neural networking implementation for forecast”, Proceedings of 1994 IEEE International Conference on Neural Networking, Orlando, USA, pages 3184 to 3189.

[2] J. Cai, J. Li, W. Li and J. Wang, 2018, “Deep learning model used in text classification”, 15th International Computer Conference on Wavelet Active Media Technology and Information Processing, Chengdu, China, pages 123 to 126.

[3] Q. Hua, S. Qundong, J. Dingchao, G. Lei, Z. Yanpeng and L. Pengkang, 2018, “A character-level method for text classification”, 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference, Xi'an, pages 402 to 406.

[4] Li, G. Zhan and Z. Li, 2018, “News text classification based on improved Bi-LSTM-CNN”, 9th International Conference on Information Technology in Medicine and Education, Hangzhou, pages 890 to 893.

[5] Y. Luan and S. Lin, 2019, “Research on text classification based on CNN and LSTM”, IEEE International Conference on Artificial Intelligence and Computer Applications, Dalian, China, pages 352 to 355.

[6] H. Hu, M. Liao, C. Zhang and Y. Jing, 2020, “Text classification based recurrent neural network”, IEEE fifth Information Technology and Mechatronics Engineering Conference, Chongqing, China, pages 652 to 655.

[7] S. Malakar and W. Chiracharit, 2020, “Thai text detection and classification using convolutional neural network”, fifty ninth Annual Conference of the Society of Instrument and Control Engineers of Japan, Chiang Mai, Thailand, pages 99 to 102.

[8] D. Greene and P. Cunningham., 2006, “Practical Solutions to the Problem of Diagonal Dominance in Kernel Document”.