ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 10, October 2017

Application of Deep learning Techniques to Natural Language Processing

[1] Sonam Gandotra, [2] Bhavna Arora,
[1] Research Scholar [2] Assistant Professor
Department of Computer Science & IT
Central University of Jammu,Jammu

Abstract **- Deep learning refers to artificial neural networks comprising of multiple layers. The deep learning algorithms automate the representation of abstract features by composing simple representations from raw data at one level to complex representations at the higher. These algorithms have improved the current state of art results and bought new insights to the current data. In this paper, the deep neural architectures and their application on NLP have been discussed. The paper also evaluates the various approaches to train the data. Various classifications of deep neural nets are also an integral part of the paper. On the basis of architecture and transfer of information from input to output layers via hidden layers, deep neural nets have been broadly classified and elaborated. Brief comparisons of the various techniques used in deep neural networks on various parameters are evaluated and have been presented..**

Index Terms: **--- Deep learning, Natural Language processing, CNN, RNN.**

1. INTRODUCTION

The human language is a system used for communication and natural language processing (NLP) focuses on making the computers analyse, understand and derive meanings from the human language in an efficient manner[1].Technically as defines, it is a combination of computer science, artificial intelligence and computational linguistics. The input and output of a NLP system can either be text or speech. Major challenge in NLP revolves around: Natural Language Understanding (NLU) and Natural Language Generation(NLG). The natural language understanding involves mapping the given input in natural language into useful representations and analysing different aspects of the language. The natural language generation is the process of producing meaningful phrases and sentences in the form of natural language from some internal representation. Developers can organize and structure knowledge based on these challenges to perform tasks such as automatic text summarization, machine translation, named entity recognition, relationship extraction, sentiment analysis, speech recognition, topic segmentation, question answering system, optical character recognition(OCR)etc.
NLP applications usually perform tasks using two families of approaches, the symbolic approach and the statistical approach as defined under:-

- Symbolic approach- This approach is based on a set of rules, often hand-written but sometimes automatically learned, that models different language phenomena.
- Statistical approach- This approach typically uses machine learning algorithms to learn the language phenomena.

This paper is organized in 5 sections. The first section introduces the concept of natural language processing. Section II gives brief introduction on deep learning concept and classification of deep learning neural net models. In Section III, a brief review of literature is described along with elaboration on how deep learning is applied to the field of NLP. Section IV analyses the performance of deep neural net models on various parameters like classification accuracy, F1 measure and mean accuracy. The paper concludes in Section V with future scope and conclusion.

II. THE DEEP LEARNING

A. Introduction to deep learning

The field of natural language processing is shifting from statistical methods to neural network methods. Many of the hard NLP tasks have been improved by Deep Learning. In this the representation of data has been shifted from word vector representation to word2vec and consequently significant results have been achieved. In the recent years, deep learning has gained immense success in almost all areas of AI due to its representation. The representation of the data is of significant importance for the system to extract most appropriate features from it. Two approaches are mostly used for training the data –

- Supervised approach-In supervised deep learning approach, algorithms learn by experience of the dataset containing many features but each data is associated with a label or tag. e.g:- If we want to classify an image as containing a car, a tree or a person, we have to feed images of those objects along with tags or labels to the model so that the model can learn from these labels and classify the given set of input images into the one that are required as output. These approaches mainly work on weight vectors that define the input-output function of the machine model.[2]

• Unsupervised approach- In unsupervised deep learning approach the system develops and organizes the data, searching common characteristics among them, and changing based on internal knowledge as the dataset provided to the system is not labelled with tags.e.g this algorithm is based on clustering similar objects based on a similarity measure being defined by the particular feature.[3]

B. Classification of Deep Neural Nets

On the basis of architecture and transfer of information from input to output layers via hidden layers deep neural nets are broadly classified as:

• Fully connected feed –forward neural nets:

Feed forward networks are artificial neural nets(ANN) inspired by the biological neuron consisting of a number of units connected by weighted links. The units are organized in several layers namely input layer, one or more hidden layers and output layer.[4] Depending on the number of hidden layers, they are classified as single-layer and multi-layer perceptron networks. Every unit in a layer is connected with all the units in the previous layer, each connection may have a different strength or weight. They are called feed-forward because information only travels forward in the network (no loops) from the input nodes through the hidden nodes (if present), and then finally through the output nodes.[5]. e.g multilayer perceptron.

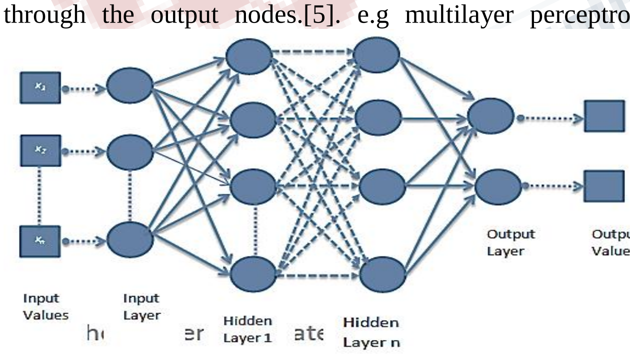

Figure 1:Fully-Connected Multi-layer feed-forward Neural network

- *Networks with convolutional and pooling layers*

  Convolutional Neural Networks are inspired from the multi-layer feed-forward network but instead of using connections, convolutions are used over the input layer to compute the output and then pooling(sub sampling) layers are applied after the convolutional layers. The pooling process provides

  a) fixed size output matrix, which is required for classification

  b) reduces the output dimensionality while keeping the most salient information.[1]

  The figure 2 below illustrates a full layer in a CNN consisting of convolutional and subsampling sub layers.

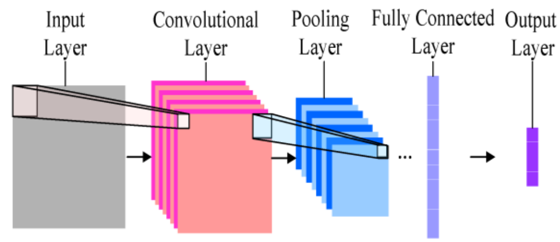

Figure 2: Architecture of CNN[6]

- *Recurrent Neural Networks*

  Recurrent networks are feedforward networks having memory. A feedback loop is connected to their past decisions, ingesting their own outputs moment and taking that moment as input. They make use of sequential information that feed forward networks can't. That sequential information is preserved in the recurrent network's hidden state, which manages to span many time steps as it cascades forward to affect the processing of each new example[7].Over the years researchers have developed more sophisticated types of RNNs such as Bidirectional RNN, Deep RNN, LSTM networks.Figure 3 describes the architecture of a typical RNN.

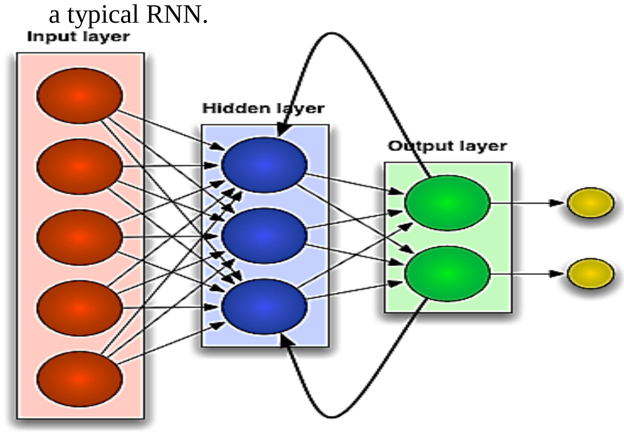

Figure 3: A typical RNN[8]

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 10, October 2017

- *Recursive Neural Networks*
  Recursive neural networks were introduced for processing data from structured domains and are suited for both classification and regression problems[9].These are created by applying the same set of weights recursively over a structure, in order to have structured/scalar prediction over the input, by traversing a given structure in topological order. It can be seen as a generalization of the recurrent neural network which are deep in time whereas recursive neural networks are deep in structure, because of the repeated application of recursive connections[1].The general structure of a recursive neural net is given in figure 4:

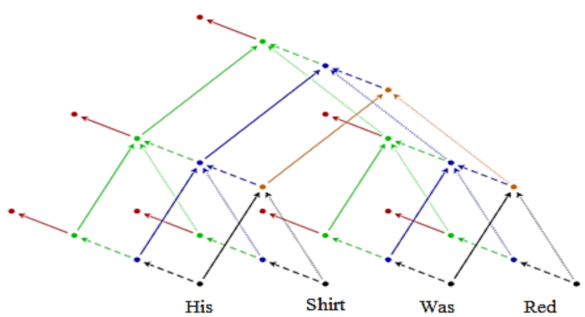

Figure 4: Structure of Recursive neural net

Due to the properties of having memory element and the recursive structure of the data, the recurrent and recursive neural nets are considered more apt for processing of natural language. CNNs are also used for NLP task but huge amount of training and testing data sets are required. So the RNN's are appropriate for the language generation and understanding tasks.

I. Applications of NLP tasks using deep learning

This section discusses the various applications of natural language processing tasks using deep learning. A brief review of literature of each of the tasks is also mentioned. The following are the key applications that are used using deep learning:-

A. *Text Classification*: Text categorization or text classification is the task of assigning predefined categories to text documents. Its application can be seen in many tasks such as sentence classification, document classification, sentiment analysis, spam filtering, language identification, genre classification etc.

In the paper [10], the author has described the convolutional neural networks for sentence classification which discusses the sentiment analysis of movie reviews, classifying sentences as being subjective or objective, classifying question types, and sentiment of product reviews. In [11], the authors has proposed a novel method in which semantic cliques are discovered using fast clustering method based on searching density peaks. The author also proposes method for fine-tuning multi- scale semantic units and also the

B. *Document Summarization*: Document summarization aims at creating a representative summary or abstract of the entire document, by finding the most informative sentences. In [12], author has designed a framework of document summarization via Deep Neural Networks and tested this framework on the DUC2006 and DUC2007 dataset. The results outperform the other state of art methods. The author in paper[13], has developeda novel summarization system called TCSum, which leverages plentiful text classification data to improve the performance of multi-document summarization. TCSum projects documents onto distributed representations which act as a bridge between text classification and summarization.

C. *Question Answering:* Question answering is the problem where given a subject, such as a document of text, answers a specific question about the subject. It tries to answer a user query that is formulated in the form of a question by returning the appropriate noun phrase such as a location, a person, or a date.In paper [14], author proposes a dynamic memory network (DMN), a neural network architecture which processes input sequences and questions, forms episodic memories, and generates relevant answers. Questions trigger an iterative attention process which allows the model to condition its attention on the inputs and the result of previous iterations.A CNN forlearning an optimal representation of question and answersentences by using relational information given by the matchesbetween words from the two members of the pair is proposed in [15]. This CNN allows for better capturing interactions betweenquestions and answers, resulting in a significant boost in accuracy.

*Named Entity Recognition:*NER also known as entity identification, entitychunking and entity extraction concerns itself with detecting and classifying certain elements in a string of text contained within a natural language document. These elements may be identified as places, quantities, time expressions, or names of persons or locations. In the paper [16], author presented a new form semantic cliques are used to supervise the selection stage.

D. of learning word embeddingby extending the skip-gram model that can leverage information from relevant lexicons to improve the representations, and the first system to use neural word embeddings to achieve state-of-the-art results on named-entity recognition in both CoNLL and Onto Notes NER.The author in paper [17] presented a hybrid model of bi-directional LSTMs and CNN that learns both characters- and word-level features, presenting the first evaluation of such an architecture on well-established English language evaluation datasets and a new lexicon encoding scheme and matching algorithm that can make use of partial matches and compared it to the previous approaches.

I. Analysis of performance of deep neural model

The performance of series of deep learning methods on standard datasets developed in recent years on the above four NLP tasks as discussed in Section-III have been analysed in this section. A brief comparison of the techniques discussed above are compared on various parameters has been evaluated in this section.

- *Analysis-I-(Text Classification)*

Table 1 shows the text classification accuracy of the models as described in previous Section II A on the TREC-dataset.
TREC:It is a questions dataset containing 6 different question types. The training dataset consists of 5452 labelled questions whereas the test dataset consists of 500 questions

Table1. Text Classification Analysis

	Model		TREC-Dataset
Kim,2014[10]	CNN-rand		91.2
	CNN-static		92.8
	CNN-non-static		93.6
	CNN-multichannel		92.2
Wang,2015[11]	Semantic-CNN	Senna	96.4
		GloVe	97.2
		Word2Vec	95.6

From Table 1 it is evident that model of Semantic-CNN proposed in [11] has better accuracy as compared to the CNN models proposed in [10]

- *Analysis-II-(Document Summarization)*

Table-2 analyses the performance of the models of document summarization on ROGUE parameter which can generate three scores: recall, precision and F1 measure. Only F1 score is considered in this analysis. The dataset used is DUC which contains news articles.

Table2. Document Summarization

	Dataset	ROGUE-1
Yao, 2015[12]	DUC2006	33.20
	DUC2007	37.37
Wei,2016[13]	DUC2001	36.45
	DUC2002	36.90
	DUC2004	38.27

Table 3 analyses the performance of question-answering model proposed by the authors[14],[15] in Section-IIIC. Both model's performance are parameterized on mean accuracy.

Table 3. Question-Answering

Paper	Dataset	Mean Accuracy
Kumar, 2016[14]	Facebook's bAbI	93.6
Severyn,2016[15]	WikiQA	69.51

- *Analysis-IV-(Named Entity Recognition)*

Table 4 analyses the performance of NER system as described in Section-IIID on F1 measure with CoNLL-2003 and OntoNotes 5.0 datasets.
CoNLL-2003: A subset of the Reuters dataset containing around 10k words, which has been manually re-annotated with 27 classes.
Onto Notes 5.0: An annotated corpus containing 1.5 million words of English, 800 K of Chinese, and 300 K of Arabic.

Table 4. Named Entity Recognition

	Model	CoNLL-2003	Onto Notes 5.0
Passos, 2014[16]	Lexicon infused phrase embedding	90.90	82.24
Chiu, 2016[17]	Bidirectional LSTM-CNN	91.62	86.28

The comparative analysis as presented in this section clearly reveals that new models are being developed and the performance of these models have shown significant results as compared to their predecessor models. Although deep learning is not very popular in the NLP community as it is in computer vision. But with the improvement of results in NLP task with deep learning techniques, it is surely going to create a benchmark in NLP.

III. CONCLUSION AND FUTURE WORK

This paper presents state-of-the-art deep learning techniques for Natural Language Processing along with the description of various deep learning models. In this paper, only a few sub-tasks of the natural processing as compared to a variety of available tasks in NLP are discussed. The recent advances in the area of Natural Language Processing using deep learning are taken into consideration and how these techniques provide break through to the performanceof the existing systems for solving these tasks is also described in this paper. After the advances made in digital image processing and computer vision using deep learning tools, researchers are keen to apply these techniques to NLP to make major breakthroughs. However, the results, for now, are only promising but there is still more scope for improvements in natural language processingfield.The researchers shouldmake a shift to this new technique for achieving better performance of the existing scenario.

REFERENCES

[1] M. M. Lopez, "Deep Learning applied to NLP."

[2] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," 2015.

[3] C. Masolo, "Supervised, unsupervised and deep learning – Towards Data Science – Medium," 2017. [Online]. Available: https://medium.com/towards-data-science/supervised-unsupervised-and-deep learning-aa61a0e5471c. [Accessed: 07-Oct-2017].

[4] M. Riedmiller, "Advanced Supervised Learning in Multi-layer Perceptrons - From Backpropagation to Adaptive Learning Algorithms."

[5] J. McGonagle, "Feedforward Neural Networks | Brilliant Math & Science Wiki." [Online]. Available: https://brilliant.org/wiki/feedforward-neural-networks/. [Accessed: 02-Oct-2017].

[6] M. Peng, C. Wang, T. Chen, and G. Liu, "NIRFaceNet: A convolutional neural network for near-infrared face identification," *Inf.*, vol. 7, no. 4, p. 61, Oct. 2016.

[7] "A Beginner's Guide to Recurrent Networks and LSTMs - Deeplearning4j: Open-source, Distributed Deep Learning for the JVM." [Online]. Available: https://deeplearning4j.org/lstm.html. [Accessed: 03-Oct-2017].

[8] "Recurrent neural networks in Ruby - Joseph Wilk." [Online]. Available: http://blog.josephwilk.net/ruby/recurrent-neural-networks-in-ruby.html. [Accessed: 03-Oct-2017].

[9] A. Chinea, "Understanding the Principles of Recursive Neural Networks: A Generative Approach to Tackle Model Complexity."

[10] Y. Kim, "Convolutional Neural Networks for Sentence Classification," pp. 1746–1751.

[11] P. Wang *et al.*, "Semantic Clustering and Convolutional Neural Network for Short Text Categorization," pp. 352–357.

[12] C. Yao, J. Shen, and G. Chen, "Automatic Document Summarization via Deep Neural Networks," in *2015 8th International Symposium on Computational Intelligence and Design (ISCID)*, 2015, pp. 291–296.

[13] Z. Cao, W. Li, S. Li, and F. Wei, "Improving Multi-Document Summarization via Text Classification," pp. 3053–3059, 2016.

[14] A. Kumar *et al.*, "Ask Me Anything: Dynamic Memory Networks for Natural Language Processing."

[15] A. Severyn and A. Moschitti, "Modeling Relational Information in Question-Answer Pairs with Convolutional Neural Networks."

[16] A. Passos, V. Kumar, and A. Mccallum, "Lexicon Infused Phrase Embeddings for Named Entity Resolution," pp. 78–86, 2014.

[17] J. P. C. Chiu and E. Nichols, "Named Entity Recognition with Bidirectional LSTM-CNNs," no. 2003, 2014.