

Neural network techniques for NLP tasks

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

Abstract:-- Increase in the number of Internet users have also make people from different communities to interact with each other and hence, a need to provide resources for communicating with each other has given rise to the idea of Natural Language Processing(NLP).NLP is a combination of computer science, artificial intelligence and computational linguistics. In this paper, various NLP tasks are discussed. This paper also describes the various neural net models and their classification on the basis of their architecture and transfer of information from input to output layers via hidden layers. A brief comparison of popular NLP tasks using neural architectures is also done.

Keywords:- Natural Language Processing, Neural Nets, Text Classification, Summarization

I. INTRODUCTION

Natural Language Processing(NLP) focuses on making the computers analyse, understand and derive meanings from the human language in an efficient manner[1]. A NLP system can take either text or speech as its input and output. Natural Language processing tasks can further be categorized as:

- a) Natural Language Generation (NLG): It is the process of producing meaningful phrases and sentences in the form of natural language from some internal representation.
- b) Natural Language Understanding (NLU): It involves around mapping the given input in natural language into useful representations and analysing different aspects of the language.

The two most common learning approaches used to solve NLP tasks are the rule-based approach and the statistical approach as defined under:

- **Rule-based learning Approach:** This approach is based on a set of rules, often hand-written but sometimes automatically learned, that *models* different language phenomena.
- **Statistical learning Approach:** This approach typically uses machine learning algorithms to *learn* the language phenomena. Most of NLP research has focused on statistical tools since the statistical revolution[2] as they yield better results as compared to the symbolic rules.

With the advancement in technology, both these approaches are being replaced by neural net models commonly referred to as deep learning approach which is providing state-of-art results in almost every task of NLP. Although, deep learning has provided state-of-art results in field of image processing but its application to NLP has also proven competent to the existing system.

This paper is organized in 5 sections. The first section introduces the concept of natural language processing and various approaches used for solving it. Section II gives brief introduction on the most common NLP tasks based on syntax and semantics. Introduction to neural nets and their classification is given in Section III. In Section IV, a brief review of literature on text classification, document summarization and question-answering using neural net models in recent years is presented along with the comparative analysis of the techniques used is also evaluated on various parameters. The paper concludes in Section V with future scope and conclusion.

II. MOST COMMON NLP TASKS

Natural Language Processing is a very wide field of study and considered as a hard task. Most of the tasks in NLP are inter-related, but for convenience, they are divided into several categories. Some of these tasks have direct implementation in solving real-world problem while others serve as a sub-task or precursor step to other bigger tasks.

Following are some of the most commonly researched task in NLP:

A. Based on Syntax

These tasks are considered as the pre-processing techniques applied to text for the purpose of data cleaning and removing ambiguous structure from the data. The processed result is then applied to other bigger tasks. Most commonly researched syntax based tasks are as follows:

- **Stemming:** It is a technique to remove suffices and prefixes from the word in order to find the stem. Search engines commonly apply stemming for indexing words as it increases the retrieval accuracy and reduces the size of the index(only stem has to be stored instead of all forms of the word)[3].
- **Tokenization:** It is the process of breaking the given text into individual units called tokens. A token is a piece of a whole, so a word is a token in a sentence, and a sentence is a token in a paragraph. It is also referred to as word segmentation[4].
- **Part of Speech Tagging (POS):** It is the process of assigning parts of speech i.e., noun, verb, adjective etc. to each word in the given document or sentence. Its application is mostly found in information retrieval[5].
- **Parsing:** It is a technique of analyzing the given text and determining its constituent parts from its structure. It resolves structural ambiguity of given text.

Table 1 summarizes each of these techniques with the help of examples.

Table 1 Syntax based Tasks

Task	Example	Output
Stemming	Walking, Walk, Walked	Walk
Tokenization	This sentence contains 6 tokens.	[This][sentence] [contain][6][tok ens][.]

Part-of-speech Tagging	Heat water in a large vessel.	Heat(VP) water(NP) in(PRP) a(DET) large(ADJ) vessel(NP)
Parsing	Bank raises interest rates.	(S(NP(N Bank))(VP(V raises)(NP(N interest)(N rates)))

B. Based on Semantics

These tasks define the literal meaning of the text considering the structure of the text given and mostly use the output of the syntax based task as its input. Commonly researched tasks in this category are as follows:

- **Named Entity Recognition (NER):** NER also known as entity identification, entity chunking 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[6].
- **Sentiment Analysis:** It is the process of determining emotions in a given text and mostly used to understand the attitudes, opinion and emotions expressed within an online mention. It is also sometimes referred to as opinion mining. The output of analysis is determined in positive, negative and neutral sense[7].
- **Summarization:** It is the task of producing a concise summary of given text while preserving the key information and overall meaning of the text. The main idea of summarization is to find a subset of data which contains the information of the whole document[8].
- **Machine Translation:** It is the task of automatically converting text from one source language to another,

preserving the meaning of text. It is considered one of the hardest tasks in NLP due to the variation of language spoken around the world. Each language is entirely different from other with its own set of complexities.

- **Question-Answering System:** It 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.

Table 2 summarizes each of these techniques with the help of examples

Table 2 Semantic based Tasks

Task	Example	Output
NER	Kunal went to Delhi on Wednesday.	Kunal/PER went to Delhi/LOC on Wednesday/DAY
Sentiment Analysis	Going to party was fun.	Positive
	He is not feeling good.	Negative
	Where are you going?	Neutral
Summarization	Rahul and Sam took the bus to visit the museum. They saw ancient sculptures, remains of extinct animals and various models.	Rahul and Sam visit museum and saw ancient things.

Machine Translation	He is going to market.	वह बाजार में जा रहा है।
Question-Answering	Who is the prime minister of Israel?	Steven Speilberg!

III. NEURAL NETS AND THEIR CLASSIFICATION

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 use of neural net techniques commonly referred as 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 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.[9]
- **Unsupervised approach-** In unsupervised 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. This algorithm is based on clustering similar objects based on a similarity measure being defined by the particular feature.[10]

A. 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.[11] 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.[12]. E.g. multilayer perceptron.

- *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].

- *Recurrent Neural Networks*

Recurrent networks are feed-forward 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[13].Over the years researchers have developed more sophisticated

types of RNNs such as Bidirectional RNN, Deep RNN, LSTM networks.

- *Recursive Neural Networks*

Recursive neural networks were introduced for processing data from structured domains and are suited for both classification and regression problems[14].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].

IV. NLP TASKS USING NEURAL NET TECHNIQUES

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 mentioned along with the comparisons of techniques used in the form of table for each task. The following are the NLP tasks which are implemented with neural net:

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 [15], 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 [16], 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 semantic cliques are used to supervise the selection stage.

Table 3 shows the text classification accuracy of the models as described above 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.

Table3. Text Classification Analysis

	Model		TREC-Dataset
Kim,2014[15]	CNN-rand		91.2
	CNN-static		92.8
	CNN-non-static		93.6
	CNN-multichannel		92.2
Wang,2015[16]	Semantic-CNN	Senna	96.4
		GloVe	97.2
		Word2Vec	95.6

From Table 3 it is evident that model of Semantic-CNN proposed in [16] has better accuracy as compared to the CNN models proposed in [15].

B. Document Summarization: Document summarization aims at creating a representative summary or abstract of the entire document, by finding the most informative sentences. In [17], 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[18], has developed a 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.

Table 4 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.

Table 4. Document Summarization

	Dataset	ROGUE-1
Yao, 2015[17]	DUC2006	33.20
	DUC2007	37.37
Wei,2016[18]	DUC2001	36.45
	DUC2002	36.90
	DUC2004	38.27

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 [19], 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 for learning an optimal representation of question and answer sentences by using relational information given by the matches between words from the two members of the pair is proposed in [20]. This CNN allows for better capturing interactions between questions and answers, resulting in a significant boost in accuracy.

Table 5 analyses the performance of question-answering model proposed by the authors[19],[20]. Both model's performance are parameterized on mean accuracy.

Table 5. Question-Answering

	Dataset	Mean Accuracy
Kumar, 2016[19]	Facebook's bAbI	93.6

Severyn,2016 [20]	WikiQA	69.51
----------------------	--------	-------

The comparative analysis as presented for each NLP task 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.

V. CONCLUSION AND FUTURE WORK

This paper presents state-of-the-art deep learning techniques for Natural Language Processing along with the description of various neural net 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 performance of 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 processing field. The researchers should make 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] P. Resnik, "Language Log » Four revolutions," 2011. [Online]. Available: <http://languagelog.lidc.upenn.edu/nll/?p=2946>. [Accessed: 02-Nov-2017].
- [3] N. Hardeniya, J. Perkins, and D. Chopra, "Stemming words - Natural Language Processing: Python and NLTK." [Online]. Available: https://www.packtpub.com/mapt/book/big_data_and_business_intelligence/9781787285101/12/ch02lv11sec019/stemming-words. [Accessed: 02-Nov-2017].
- [4] Robin, "Tokenization - Natural Language Processing," 2009. [Online]. Available: <http://language.worldofcomputing.net/category/tokenization>. [Accessed: 02-Nov-2017].
- [5] L. Speech, Processing, Daniel, H. James, and Martin, "Part-of-Speech Tagging."
- [6] R. Sharnagat, "Named Entity Recognition: A Literature Survey," pp. 1–27, 2014.
- [7] "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, Dec. 2014.
- [8] M. Gambhir and V. Gupta, "Recent automatic text summarization techniques:," *Artif. Intell. Rev.*, vol. 47, no. 1, pp. 1–66, 2017.
- [9] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," 2015.
- [10] 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].
- [11] M. Riedmiller, "Advanced Supervised Learning in Multi-layer Perceptrons - From Backpropagation to Adaptive Learning Algorithms."
- [12] J. McGonagle, "Feedforward Neural Networks | Brilliant Math & Science Wiki." [Online]. Available: <https://brilliant.org/wiki/feedforward-neural-networks/>. [Accessed: 02-Oct-2017].
- [13] "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].
- [14] A. Chinae, "Understanding the Principles of Recursive Neural Networks: A Generative Approach to Tackle Model Complexity."

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

- [15] Y. Kim, "Convolutional Neural Networks for Sentence Classification," pp. 1746–1751.
- [16] P. Wang *et al.*, "Semantic Clustering and Convolutional Neural Network for Short Text Categorization," pp. 352–357.
- [17] 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.
- [18] Z. Cao, W. Li, S. Li, and F. Wei, "Improving Multi-Document Summarization via Text Classification," pp. 3053–3059, 2016.
- [19] A. Kumar *et al.*, "Ask Me Anything: Dynamic Memory Networks for Natural Language Processing."
- [20] A. Severyn and A. Moschitti, "Modeling Relational Information in Question-Answer Pairs with Convolutional Neural Networks."

