Role of Deep Recurrent Neural Networks in Natural Language Processing

S. T. Shenbagavalli , D. Shanthi ,S. Naganandhini, R. Karthikeyan

***Abstract*— Deep learning methods are used to study hierarchical representations of data. Natural Language Processing is a group of computing methodologies used for analyzing and illustrating of Natural Language (NL). Natural Language is used to collect and present information in numerous fields. NLP can be to extract and process information in human language automatically. This paper is to highlight vital research contributions in text analysis, classification and extracting useful information using NLP.**

Keywords: Deep Learning, NLP, Applications

I. INTRODUCTION

A greater number of natural language-based processes are performed using only NLP in computers. NLP processes involves tagging, naming entities, chunking and labelling by semantics. Chunking is stamping the words in a whole sentence can be parsed by verifying noun and verbs. Tagging of the terns is carried out based on uniqueness of the words in a sentence. It can be done by verifying the syntactic role. NER categorizes individual elements into different classes. Parse tree construction and node analysis are carried out to give semantic roles to terms of a sentence in SRL. Deep Learning methods are used in recent NLP researches. Dense vector representations based Neural Networks gives great results in different NLP processes [1].

II.STATISTICAL NLP

Distributional vectors follow the distributional hypothesis. These vectors are used to find similar words by capturing the characteristics of the neighbors of a word. To obtain the similarity among the words, Cosine similarity method is used. Word embedding process is trained at the initial layer called as data processing layer. NLP models incorporated with deep learning methods are well performed in representing the targets such as words and phrases using distributional vectors. Character embedding deals with out-of-vocabulary issue which is common in languages that have large vocabularies. Here word is considered as composition of individual characters. Character level systems are useful in avoiding word segmentation problem.

III. NEURAL NETWORKS MODELS

A. *Convolutional Neural Networks (CNN) Models*

To extract features from constituting words which can be used in NLP processes like sentiment analysis, machine translation and summarization CNN models are considered as effective choice. [2]. Words are transformed into series of vectors of user-defined dimensions using look-up tables. CNNs have the ability to extract as much as possible number of features available in the input sentences to classify the words based on their semantic representations. Max pooling of CNN maps variable length input to fixed dimension output.

One of the important process is word embedding, where it can be done randomly and applying pre-training method. Training is applied on the unlabeled data. The pre training method is preferable when the labeled data is lesser. To create deep CNN networks Combination of convolution layer followed by max pooling is frequently used. The number of convolution layers are always more than two in any CNN architecture. It is because, to learn the input data thoroughly and deeply to extract all the features from the input data. It increases the mining accuracy of the NLP task.

These sequential convolutions help in improved mining of the sentence where it helps to observe the rich semantic based data. The existing kernels in the CNN model extracts the large portion of the kernel's presence in the CNN model. Finally, the fully connected layer classifies the sentences by summarizing the features extracted from the input data. Sentiment, subjectivity and question type classification, semantic modeling of sentences is some of the applications of CNNs.

B. *Recurrent Neural Networks (RNN) Models*

RNNs apply the idea of dealing out chronological data. Same computation is performed on each word presence in the sequence. Various processes carried out in each layer is depending on the previous result and process. RNN can get sequential nature of a language. Variable length text like character, word and a long sentence cab ne modeled using RNNs. RNN provides time distributed joint processing. Word level classification, sentence level classification, language generation are important applications of RNNs

Recursive Neural Networks, Deep Reinforced Models and Deep Unsupervised Learning models are other models used in NLP tasks.

Revised Version Manuscript Received on 16 September, 2019.

S.T.Shenbagavalli, Department of Computer science and Enginnering, PSNA College of Engineering and Technology, Dindigul, Tamilnadu,India, E-Mail: mailatshenbagavalli@gmail.com.

Dr.D.Shanthi, Department of Computer science and Enginnering, PSNA College of Engineering and Technology, Dindigul, Tamilnadu,India, E-Mail: dshan71@gmail.com

S.Naganandhini Department of Computer Science and Engineering and Technology,PSNA College of Engineering and Technology, Dindigul, Tamilnadu,India, E-Mail:nandhu.be2010a@gmail.com

Dr.R.Karthikeyan, Department of Computer Science and Engineering,, PSNA College of Engineering and Technology, Dindigul, Tamilnadu,India, E-Mail: karthic.bit@gmail.com

Retrieval Number: B15970982S1119/2019©BEIESP
DOI: 10.35940/ijrte.B1597.0982S1119

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

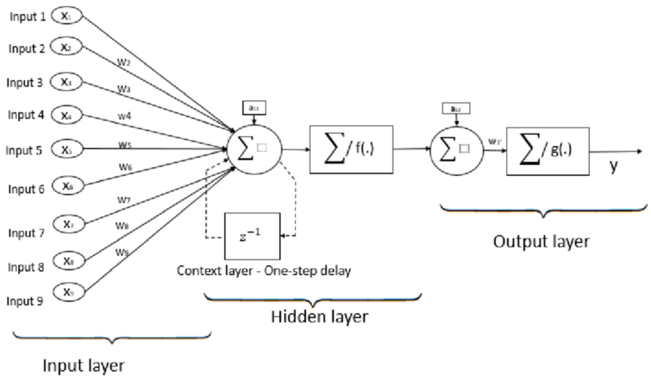

Fig. 1. Proposed RNN Architecture

IV.BIOLOGICAL NLP

Bioinformatics combines genetics, molecular biology and computer science to address various biological problems from a computational point of view. Biological NLP (BNLP) has more important in bioinformatics. BNLP approaches is applied to fetch important information on biological-texts for predictions such as gene-disease association, protein functions.

The proposed model given in [3] used to tokenizes all the words and create a tree for parsing the input sentences. Nodes under a parent has a semantic relationship and it can be obtained by proposition. For example, constructing a path from node 1 to node 2 in the tree is unique. Dependency between candidate term, T_c, and input term T_i is characterized by a value that is resolute based on co-occurrence frequency of T_i and T_c, and semantic relationships between T_i and T_c in texts. $Score(T_i,T_c)$ is calculated using

$$Score(T_i,T_c) = f(T_i,T_c)-f'(T_i,T_c).$$

$F(T_i,T_c)$ is the frequency of T_i and T_c being semantically related and $f'(T_i,T_c)$ is unrelated. This model enhances biological text mining.

V.MEDICAL NLP

Medical NLP offers method to extract information from medical notes. A fame work to analyze patient readmission prediction by Ankur Agarwal et al., [4]. This framework consists of three subsystems. Feature extraction subsystem uses Bag of Words method and cTAKES annotation to fetch each separate measurable property. Feature selection subsystem is used to select useful features extracted by feature extraction subsystem. It employs feature relevance ranking, testing of all possible combinations of features and classification algorithm. Classification subsystem uses methods such as Naïve Bayes, Random Forest, K-Nearest Neighbors and Support Vector Machine to classify selected features. The system is able to predict hospital readmissions and provides easy integration with electronic health records.

VI. TCC NLP

Templates are important tools to increase precision of NL (Natural Language) requirements and to avoid ambiguity. NLP can be applied to automate Template Conformance Checking (TCC) [5]. To identify sentence segments text chunking approach is used. This method can identify chunks without expensive analysis. The system uses NLP parsing combined with text chunking to determine template compliance.

A tool named Requirements Template Analyzer is designed for automated TCC. Text chunking is sequence of processes such as tokenizing, sentence splitting, part of speech tagging, named entity recognizing and classifying phrases. Templates are expressed as grammars and conformance checking is done using pattern matching.

VII. QUALITY NLP

Software domain rules the world. Software project requirements such as delivery in time, quality and completion within the specified budget lead developers to use workarounds – Technical Depts. These tradeoffs lead to negative impact on software quality. Technical Debts can be deliberate or inadvertent. NLP can be applied to detect deliberate technical debts automatically [debt]. In order to find debts automatically open source projects are considered for analyzing the source code comments.

Numbers of classes, number of developers, total lines of source code, total number of comments are extracted for analysis. Source code is parsed to fetch details of comments such as type, location and context. Extracted comments are filtered to get most applicable and insight comments. Stanford Classifier is used to train maximum entropy classifier with classified dataset. This classifier extracts most important feature for each class automatically and finds features that contribute negatively. This method effectively detects design and requirement self admitted technical debt.

VIII. NLP FOR RE

Requirement Engineering (RE) is the process of collecting and defining the services provided by a system. Requirements are written in NL. Requirements tend to be redundant due to several concurrent projects and large number of requirements of those projects. Requirements written by several analysts of different projects or different departments of same project are in huge number [6].

Requirement engineering research includes finding quality issues and ambiguity, distinguishing and grouping of requirements, extraction of key abstractions, creating new models and traceability between NL requirements and code. NLP techniques can be applied to analyze requirements-related documents automatically.

IX. CONCLUSION

Basic tasks of Natural Language Processing are explained in brief in this article. Various Neural Network models applied in NLP tasks with its applications are mentioned to give basic understanding. Recent research contributions and important NLP applications are elucidated.

REFERENCES

1.Tom Young, Devamanyu Hazarika, Soujanya Poria, Eric Cambria, "Recent Trends in Deep Learning Based Natural Language Processing", IEEE Computational Intelligence, 1556-1603, August 2018.

2.R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," in Proc. 25th Int. Conf. Machine Learning, 2008, pp. 160–167.

3.Kamal Tasha, "Extracting Various Classes of Data From Biological Text Using the Concept of Existence Dependency", IEEE Journal of Biomedical and Health Informatics, Vol. 19, NO. 6, November 2015

4.Ankur Agarwal, Christopher Baechle , Ravi Behara, and Xingquan Zhu, "A Natural Language Processing Framework for Assessing Hospital Readmissions for Patients With COPD" IEEE Journal Of Biomedical And Health Informatics, Vol. 22, No. 2, March 2018

5.Chetan Arora, Mehrdad Sabetzadeh, Lionel Briand, Frank Zimmer, "Automated Checking of Conformance to Requirements Templates Using Natural Language Processing" IEEE Transactions On Software Engineering, Vol. 41, No. 10, October 2015

6.Fabiano Dalpiaz, Alessio Ferrari, Xavier Franch, and Cristina Palomares, "Natural Language Processing for Requirements Engineering" IEEE Software, September/October 2018

Retrieval Number: B15970982S1119/2019©BEIESP
DOI: 10.35940/ijrte.B1597.0982S1119

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication