



## Review on Text Sequence Processing with use of different Deep Neural Network Model

Sheetal S. Pandya<sup>1</sup>, Nilesh B. Kalani<sup>2</sup>

<sup>1</sup>Computer Engineering, RK University, School of Engineering, India, Sheetal.pandya@rku.ac.in

<sup>2</sup>Electronics and Communication, RK University, School of Engineering, India, Nilesh.kalani@rku.ac.in

### ABSTRACT

This paper underlines the importance of Deep Neural Network model in different learning tasks for Natural language processing. Deep learning techniques carry a lot of support to handle a large amount of trained or labeled data sets. Deep Neural Network model involves multiple processing layers to learn Sequential representations of Text data to achieve excellent performance in many domains. In this paper, we review significant deep learning models with end to end approach for Natural language processing that makes a more accurate assumption on the word sequence as well as represents a different experimental analysis for NLP tasks in terms of text and words with the help of CNN, RNN, LSTM, and GRU bidirectional Encode-Decoder. Numerous NLP tasks are involved in semantic execution to understand and generate a complete sentence where different Deep learning models represent awareness about numerous NLP tasks including sequential information processing. That's a reason why Deep Neural Network Model gradually became more famous for different NLP applications. Even we have endorsed reinforcement learning as an extension to the Deep neural model which is frequently used in Natural Language Processing. This paper differentiates the Deep Neural Model in various language processing tasks and directs the selection of Deep Neural Model in different NLP tasks.

**Key words:** Bidirectional Encoder-Decoder Techniques, CNN, GRU-RNN, LSTM-RNN, Natural Language Processing (NLP), Sequence Learning, Text Processing.

### 1. INTRODUCTION

Deep Neural Networks are very powerful machine learning models that derive an excellent performance on parallel computation in a large collection of data sets. Additionally, Large size Deep Neural Network model trained with a supervised backpropagation method with some labeled data sets has sufficient information to designate a new parameter. That's a reason why parameter found by supervised backpropagation technique can answer a number of sequence problems. Nowadays, in most of the problems related to the

sequence problem we cannot predict the length of the sequence on prior basis. During the sentence translation, voice to text conversion, voice recognition, etc. we observe that sequence-related issues arise. Also, during the question answer mapping between two words or text also create the same issue. Here, we discuss different Deep Neural Network Model which provide a solution to this kind of text sequence problem. CNN and RNN which are two main architectures of the Deep Neural Network model have the capability to understand the sequence of the text and preserve it for a long time [1].

For any text processing, CNN follows hierarchical structure and RNN follows sequential structure. Even many NLP tasks are involved in the semantic process to understand and generate complete words of sequence in terms of a sentence with the help of RNN model. On the other hand, CNN helps in text classification and Language Modeling. There is no specific selection on DNN model for any NLP issues. RNN shows awareness about sequential information processing. That's why nowadays Recurrent Neural Networks have gradually become more famous for different NLP applications. As per the prediction, each semantic input value is given by pre-processing method during text transformation, speech to text generation or dialogue creation. Based on this information, we are concerned about the procedure of expressing the current semantic input value which has a well-organized structure and is well-executed in terms of each sequence of words. Some approaches for this kind of representation have a rule-based, grammar-based, supervised learning and reinforcement learning [2].

A very generic issue common to all these mentioned approaches is that a very significant value in terms of human effort is demanded to transform a language generation to the new destination language [3]. Deep Neural Network-based Sequence to Sequence learning language model has proven to be a very precise and powerful model for many assumptions in terms of sequences of characters and words and provides a very robust methodology for automatic translation of text in any language. Execution of the sequence model starts with encoding the source text to the desired vector space and then decoding that text to a new target text sequence. With the help of a large amount of trained data sets the model will be trained to learn a prediction of correct text in sequence. In this paper, we differentiate different Deep Neural Network models like CNN, RNN including LSTM and GRU for various NLP tasks.

In some of the applications two DNN model are used like encoder and decoder where two LSTM-RNN models are simultaneously trained to study a sequence of each text word with their semantic meaning. The first model works as an encoder to study the semantic meaning of each word in an input sequence. The second model works as a decoder that studies how to decode a text representation in terms of each word sequence without changing the understating of the semantic input text [3]. That's a reason why Deep Neural Network Model has gradually become more famous for different NLP applications. We have endorsed Reinforcement learning as an extension to the Deep neural model which is frequently used in Natural Language Processing.

## 2. RELATED WORKS

With a recent experiment on Deep Neural Models, we got a very remarkable output in terms of text and images, especially in text processing such as word embedding, text translation, text classification, speech recognition, and text sentiment analysis. So, as per the large space of Natural language processing tasks, we do not have any specific criteria for selection in terms of the model.

Ngoc Thang Vu [4] examined the CNN model and without gate techniques, the RNN model works more on classification techniques. Observation is made in terms of performance. The performance of the CNN model is better in classification than RNN and generated results deliver balanced evidence hovering in between CNN and RNN. On the other hand, RNN provides subjective combination in terms of each word which is used in the sentence. Even CNN focuses on meaningful extraction of Ngrams for the better generation of results. Heike Adel & Wen-tau Yih[5][3] prefer CNN as the first choice in Text classification technique instead of RNN even when the sentence is long. Wenpeng Yin[6] represents an effective solution on answer selection. CNN performs better in an attention-based selection of LSTM. Yann N Dauphin [7] maintains that a fine-tuned Gated RNN and LSTM can also perform on long situation dependency and can get a new state-of-the-art Natural language model with a comparison of all DNN models. Besides, K. Arkhipenko[8] compares CNN, GRU/LSTM and word2vec in different sentiment analysis in Russian tweets and justifies that GRU/LSTM get better execution than CNN and RNN.

As a part of Empirical execution, Rafal Jozefowicz [9] states that there has been no specific comparison between GRU and LSTM execution in terms of many NLP tasks. Best Model selection is also focused on the length of the statement and text sequence for language generation and sentimental analysis. Related to Classification tasks like Sentimental Analysis, Text Categorization, and Span Detection CNN is a more preferable model in DNN [10]. Even CNN works on pooling operations which always follow traditional arrangements or says basic order of each text. According to this arrangement, Part of Speech tagging (POS) and Entity Extraction have posed a difficulty in execution with CNN architecture. Tom Young [11] represents CNNs as effective

because it can mine semantical representation of each text clues in entire sentence; but in terms of sequence arrangements of words in the sentences, it faces a lot of difficulties. RNN/LSTM models are better suited for such types of Sequence arrangements.

According to Fang Xu1, Guo Yi1[12] Recall-Oriented Understudy for Gisting Evaluation (ROUGE) evaluation tool is widely used for text summarization. Sequence Model used with LSTM is constructed to extract character level and word level features to create small text summary. Yong Zhang and Weidong [13] represent predicting a key phrase based on the source text entry model. They predict the key phrase with the use of three mechanisms. The first mechanism observes and predicts the meaning of the text. The second mechanism creates the key phrase based on an understanding of the entered text as input. The third mechanism observes, understands and removes some redundant assumptions from the generated key phrase. According to Omar AbdelWahab, [14] Text Sequence processing and creation technique using LSTM-RNN and GRU-RNN contrast with Markov Chain based text processing which focuses more on classification. The resultant measurements of each text generation will be task-specific. That means the accuracy of each generated text is based on the end classifier. Min-Yuh Day, Chao-Yu Chen [15] are more focused on feature extraction techniques in terms of words with the use of three models namely, obtain the Data, Data Preprocessing and Model Development in terms of short text summarization technique using Naïve-Bayes, hidden Markov models, log-linear models.

Hai-Tao Zheng [16] focused on Text Generation with the use of an encoder-decoder technique where the existing news is encoded, and the generated comments are decoded which summarize all the news in a short form.

In this paper, we review different Deep Neural Network models like CNN, RNN including LSTM and GRU for various NLP tasks. In some of the applications two DNN model are used as encoder and decoder where two LSTM-RNN models are simultaneously trained to study a sequence of each text word with its semantic meaning.

## 3. DEEP NEURAL NETWORK MODEL

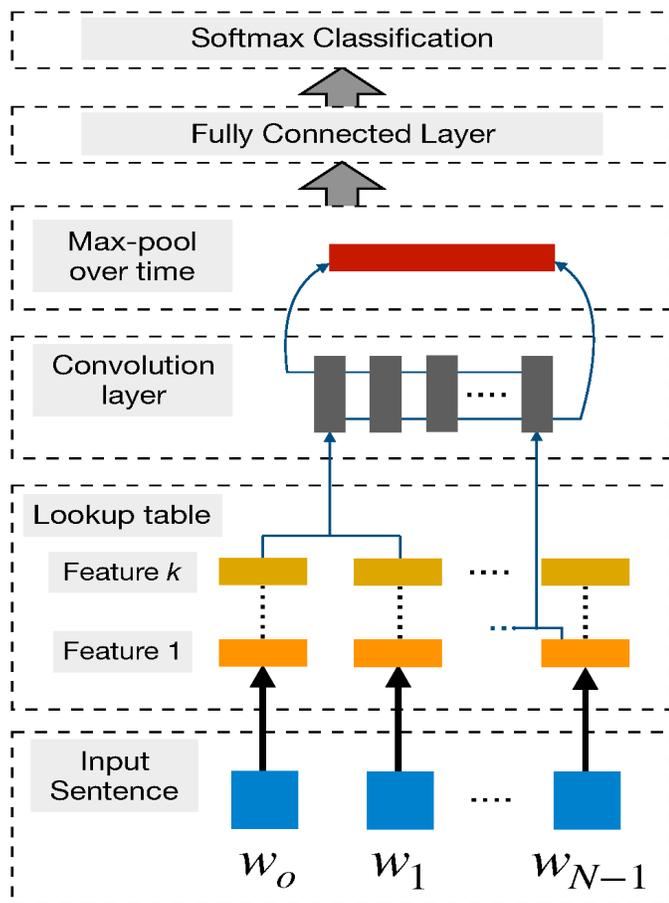
In this section, we give you brief information about each different type of DNN with its application and what the challenges are in terms of the NLP task.

### 3.1. Convolution Neural Network (CNN)

When working with any NLP task with a combination of the DNN model, we require each word in form of high dimensional divided vectors which play a major role in retrieving complete and semantic information of each word in the entire sentence. Now suppose consider  $d$  is the text length for word embedding and  $l$  is representing the entire sentence length than embedding vector-matrix is represented by encoded technique is  $C \in \mathbb{R} (d * l)$  where each  $C_i \in \mathbb{R}$ ,  $d$  represents a word embedding for an  $i$ th word in a given sentence.

As application of Convolutional Neural Network in terms of sentence modelling and processing use multitask learning approach to predict the output of multiple observation, prediction in NLP task like Part of Speech tagging (POS) tags, chunks, (NER) Named entity recognition tags, similarities in words, Language model and extracting meaningful information from text etc. [17]. Lookup table structure is used to convert each word in a sentence to form a high dimensional vector. Input sequence like  $[s_1, s_2, s_3, s_4, \dots, s_n]$  in  $n$  number of words converted in form of sequence vector  $[w_1, w_2, w_3, w_4, \dots, w_n]$  for each word in the sentence by lookup table each categories shown in given Figure-1[18].

From the given input statement CNN has the capability to derive  $n$ -gram input features for the creation of some meaningful and dormant information to perform some NLP tasks. N. Kalchbrenner [19] and R. Collobert [20] represents an application of CNN that is directed to a large amount of creation in sentence processing based on CNN architecture.



**Figure 1:** Sentence Modeling using CNN Architecture introduced by R. Collobert and J. Weston [18]

R. Johnson [21] represents CNN sentence modeling as completely dependent on the length of the sentence where efficient results are obtained in terms of long sentence processing instead of short length text. So, the Model wants to take some extra effort for sentence modeling to perform with short length texts. Even P. Wang, [22] represented a semantic

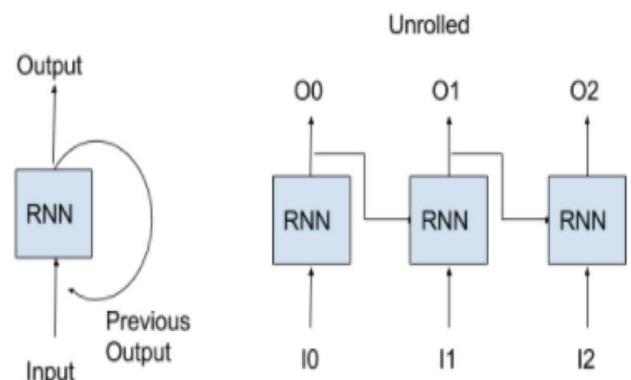
cluster model in which he mentioned larger scaled semantic factors that used exterior information for desired short text.

So, Overall the CNN model is very effective in terms of analysis semantic suspicions in relative text windows. The model covered very huge numbers of training parameters from large training datasets. Another tenacious matter with CNN Architecture is that it is unable to perform on large scale information to maintain the sequential arrangements of the texts [23],[24]. Even CNN Architecture is a very efficient in  $n$ -gram features extraction which is mainly required in sentence modelling and classification tasks of NLP. Their understanding of text sequence is delimited locally, and large-scale dependency is typically discounted.

### 3.2. Recurrent Neural Network (RNN)

The Choice behind RNN architecture works for sequential information processing. The term “Recurrent” mentions that the current computational task execution is correlated with the previous computational task. In general, a fixed dimensional vector value is generated for representing each text sequence by feeding token value turn by turn to the use of the recurrent model. We can say like RNN architecture has memory portion for computational task and mapping between previous to next text. Recent RNN model has gradually become more famous for such kind of Natural language processing tasks.

Even that an RNN model performs on sequential modelling where the model has the capability to detect intrinsic sequential performance in each text value, where all the tokens are in forms of either characters, words or entire sentences shown in Figure-2. Each word in a sentence creates its semantic or meaningful information which depends on earlier words in the sentence. As an example, any noun starts with the adjective meaning of both the words is different from each other like “orange” where the previous word of orange decided it's about color or fruit. RNN works for sequence processing task in language model which offers a very efficient role for researchers to use RNN architecture instead of CNN [25].



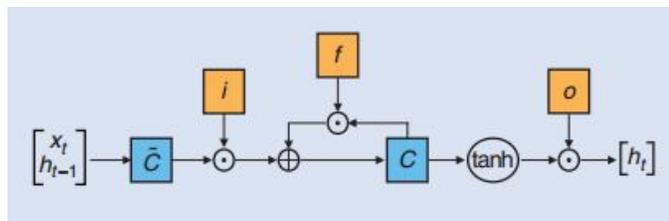
**Figure 2:** RNN Architecture

Additionally, RNN is more suitable for sequence modelling in NLP tasks in terms of each word length, involving any length of the texts, sentences and paragraphs including whole documents [26]. Unlike CNN, the RNN model has flexible evaluation parameters that have more capability to create an efficient semantic representation of each text. RNN has summarization capability through which it can understand complete sentence meaning which gets enhances its execution in machine translation. This characteristic caused RNN to become more popular in terms of sequence modelling [27]. Nowadays, numerous researches provide directions on the selection of CNN architecture over RNN architecture. RNN architecture is preferred in terms of language modelling task where CNN model proved to be very fast and competitive execution in comparison to RNN selection [28].

CNN and RNN, both the models, have different perspectives at the time of language modelling. RNN model attempts to create a configuration of an excessively long sentence along with the unrestrained situation. On the other side, CNN is powerful to abstract the valuable n-grams from the entire sentence. W. Yin, K. Kann [29] very nicely represented thought-provoking perceptions on the practical performance between CNN and RNN model. Numerous NLP tasks execution like Question-Answering task, Sentimental Analysis task, Text Classification and Part of Speech tagging, resolved that both models have equal performance level. Even the performance level of each architecture depends on the overall text’s semantics decided by NLP tasks itself. In the work of language modelling, Simple RNN architecture faces the biggest challenge in terms of vanishing gradient which makes it very complex to understand the text pattern and examine the relation between current text to previous text in the RNN architecture. This constraint was overwared by Long Short-Term Memory-LSTM, Gated Recurrent Units -GRU.

**A. Long Short-Term Memory-LSTM**

In RNN architecture long dependency and vanishing gradient problem was resolved by LSTM (Figure 3).



**Figure 3: LSTM Architecture**

According to F. A. Gers [30], LSTM has one extra gate named “Forget Gate” as compared to a simple RNN structure which allows backpropagation with respect to any numbers of time-driven steps. LSTM represents three numbers of gates named Input Gate, Output Gate, and Forget Gate. The computation of hidden structures will be performed by these gate structures as per the instruction given here. Consider each word sequence in terms of  $X$  here.

$$I_t = \sigma (X_t U^i + h_{t-1} W^i + b_i) \tag{1}$$

$$F_t = \sigma (X_t U^f + h_{t-1} W^f + b_f) \tag{2}$$

$$O_t = \sigma (X_t U^o + h_{t-1} W^o + b_o) \tag{3}$$

$$Q_t = \tanh (X_t U^q + h_{t-1} W^q + b_q) \tag{4}$$

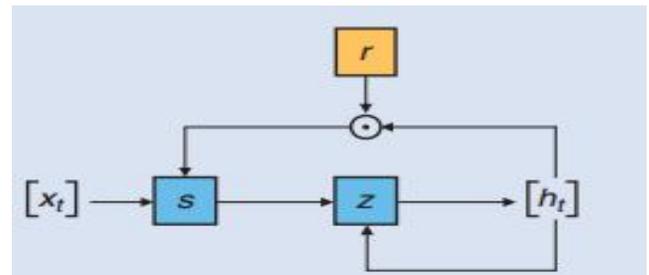
$$P_t = F_t * P_{t-1} + I_t * Q_t \tag{5}$$

$$H_t = O_t * \tanh (P_t) \tag{6}$$

All three gates are created by Sigmoid function. Here input value is  $X_t$  and hidden value earlier represented by  $h_{t-1}$  here. Each past value is represented by  $P_t$  and all the intermediate level answers are stored in  $Q_t$  by use of tanh function. Combination of intermediate level value with past history value  $P_{t-1}$  by input value  $I_t$  and Forget value  $F_t$  which get output gate value over the  $O_t$  in final hidden intermediate state value  $H_t$

**B. Gated Recurrent Unit-GRU**

In another RNN variant Gated recurrent unit of the smaller obstacle was conceived with the help of empirically comparable presentations to LSTM in different NLP tasks. Gated recurrent unit-GRU is covered by two gates named reset gate and an update gate which handle the complete flow of the information comparable with the LSTM structure without the use of memory unit. Consequently, its exposes the entire hidden information without focusing on any control mechanism. Through this mechanism we can say that Gated recurrent unit - GRU can be an additional efficient representation in RNN as compared to the LSTM structure. GRU structure is shown in Figure-4.



**Figure 4: GRU Architecture**

GRU model considers text  $X$  as input value and  $X_t \in R^d$  represents each token value  $X$  in respective position  $t$ .

$$Z = \sigma (X_t U^z + H_{t-1} W^z) \tag{7}$$

$$R = \sigma (X_t U^r + H_{t-1} W^r) \tag{8}$$

$$S_t = \tanh ( (X_t U^s + H_{t-1} * r) W^s ) \tag{9}$$

$$R = (1-Z) * S_t + Z * H_{t-1} \tag{10}$$

Here, each token value represents in their respective position with value  $t$  and their hidden state value  $H_t \in R^H$  in each text position like  $x_1, x_2, x_3 \dots x_t$  and  $R$  mentioned in two gates value. Every Sentence  $S \in R^{D * H}$ ,  $W \in R^{H * H}$  mentions token value in the given sentence.

### 3.3. Recursive Neural Network

RNN architecture focuses on the sequential arrangement of each text. Recursive Neural network follows the Hierarchical Tree structure which is more suitable for text classification and regression techniques. Representation of Recursive NN is each text and subtext associated with different phrases in the tree structure. Tree representation of sentences is more suitable for semantic modelling to extract more useful information from the sentence [31].

It's a very simple representation of the node structure. Each node is combined at a root node using its respective weight matrix value shown in Figure-5.

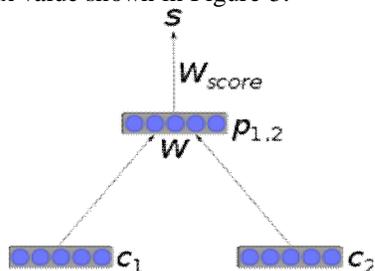


Figure 5: Recursive Neural Network

Here \$c\_1\$ and \$c\_2\$ both are \$n\$ numbers of dimensional vector value with the use of tanh function. The calculation is done with the help of \$P\_{1,2}\$ value.

$$P_{1,2} = \tanh [W [c_1 \rightarrow c_2]] \quad (11)$$

where \$W\$ represents the \$n\$ factors like \$n \rightarrow 2n\$ parameter value.

### 4. COMPARISON OF DNN MODEL

Here we consider Richard Socher[32] Sentimental Treebank training Dataset for Analysis. Even for CNN and RNN comparison with help of different NLP parameters we used some existing training dataset from Samuel R Bowman[33] and Ying Wen [34] to treat trained expressions that arise as sub-parts of trained judgments as independent training examples to measure accuracy.

Table 1: Comparison of DNN Model on NLP Tasks.

Sr.No	DNN on Different NLP Tasks		
	NLP Tasks	DNN model Name	Accuracy %
1	Text Classification	CNN	82.39
		GRU	86.38
		LSTM	84.58
2	Sentimental Classification	CNN	77.15
		GRU	78.79
		LSTM	77.84
3	Sequence to Sequence learning	CNN	54.43
		GRU	68.90
		LSTM	66.57
4	Part of Speech tagging	CNN	89.29
		GRU	90.25
		LSTM	91.20
5	Part of Question Answering	CNN	94.17
		GRU	93.11

Sr.No	DNN on Different NLP Tasks		
	NLP Tasks	DNN model Name	Accuracy %
		LSTM	93.10
		B-GRU	94.22
		B-LSTM	94.32

Table-1 mentions a small comparison of different Deep Neural Network Model in various NLP tasks. Specifically, Text Classification and Sentimental Classification techniques in NLP RNN-GRU achieves the best result when compared to CNN and RNN-LSTM. On the other side, CNN achieves the best performance in Part of Speech tagging and part of Question-Answering task even in the same tasks LSTM model represents very proficiently [35]. Sequence to Sequence learning part of text RNN-GRU and RNN-LSTM perform very efficiently in comparison to the other models. In some of the NLP tasks like part of speech tagging and Question-Answering part, the Bidirectional DNN model is used for better execution in terms of long text dependency. The execution of Sequential text performance and Context Deep Learning RNN model is a very appropriate selection for best output. In other direction CNN is well suitable for Text Classification task [36]. CNN is best for short text dependency whereas RNN for Long text range dependency. In terms of Sentimental Classification RNN model always retrieves entire sentence semantics where CNN always focuses on text to text semantics

### 5. CONCLUSION

In this paper, we underlined the importance of Deep Learning Models in different NLP tasks and comparison of CNN, RNN, RNN-GRU, and RNN-LSTM with representative examples of NLP parameters. The numerous NLP task's execution like Question-Answering task, Sentimental Analysis task, Text Classification and Part of Speech tagging, resolved that both the model have equal performance level. We conclude the RNN model has flexible evaluation parameters that have more capability to create an efficient semantic representation of each text and has summarization capability through which we can understand complete sentence meaning which gets better execution in machine translation. Additionally, CNN model is very suitable for text and sentimental classification technique and RNN is well suited for sequence modelling and large-scaled dependency tasks.

### REFERENCES

- [1] Yann LeCun, L'eon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. "Gradient-based learning applied to document recognition". Proceedings of the IEEE 86(11):2278-2324. <https://doi.org/10.1109/5.726791>
- [2] Jing Jiang, "Information Extraction from Text", Springer Science Business Media, DOI 10.1007/978-1-4614-3223-4\_2, pp 11-41,2012. [https://doi.org/10.1007/978-1-4614-3223-4\\_2](https://doi.org/10.1007/978-1-4614-3223-4_2)

- [3] Heike Adel and Hinrich Schütze. 2017. "Exploring different dimensions of attention for uncertainty detection". In Proceedings of EACL.
- [4] Ngoc Thang Vu, Heike Adel, Pankaj Gupta, and Hinrich Schütze. 2016. "Combining recurrent and convolutional neural networks for relation classification". In Proceedings of NAACL HLT pages 534–539.
- [5] Wen-tau Yih, Matthew Richardson, Chris Meek, MingWei Chang, and Jina Suh. 2016. "The value of semantic parse labeling for knowledge base question answering". In Proceedings of ACL. pages 201–206.
- [6] Wenpeng Yin, Hinrich Schütze, Bing Xiang, and Bowen Zhou. 2016. "ABCNN: attention-based convolutional neural network for modeling sentence pairs". *TACL* 4:259–272.  
[https://doi.org/10.1162/tacl\\_a\\_00097](https://doi.org/10.1162/tacl_a_00097)
- [7] Yann N Dauphin, Angela Fan, Michael Auli, and David Grangier. 2016. "Language modeling with gated convolutional networks". arXiv preprint arXiv:1612.08083.
- [8] K. Arkhipenko, I. Kozlov, J. Trofimovich, K. Skorniakov, A. Gomzin, and D. Turdakov. 2016. "Comparison of neural network architectures for sentiment analysis of russian tweets". In Proceedings of "Dialogue 2016".
- [9] Rafal Jozefowicz, Wojciech Zaremba, and Ilya Sutskever. 2015. "An empirical exploration of recurrent network architectures". In Proceedings of ICML. pages 2342–2350.
- [10] Duyu Tang, Bing Qin, and Ting Liu. 2015. "Document modeling with gated recurrent neural network for sentiment classification". In Proceedings of EMNLP. pages 1422–1432.
- [11] Tom Young, Devamanyu Hazarika, 2018. "Recent Trends in Deep Learning Based Natural Language Processing". In Proceeding of the IEEE 1556-603x/18©2018. Pages 55-75. Computational Intelligence Magazine.  
<https://doi.org/10.1109/MCI.2018.2840738>
- [12] Fang Xu1, Guo Yi1 (2018). "Research on Automatic Summary of Chinese Short Text Based on LSTM and Keywords correction". 978-1-5386-4362-4/18/\$31.00 © 2018 IEEE. International Conference on Advanced Computational Intelligence (ICACI).
- [13] Yong Zhang and Weidong Xiao, (2018). "Key phrase Generation Based on Deep Seq2seq Model". 2169-3536/2018 IEEE. Translations and content mining are permitted for academic research only.
- [14] Omar Abdelwahab and Adel S. Elmaghraby (2018). "Deep Learning based vs Markov Chain based Text Generation for Cross Domain Adaptation for Sentiment Classification" 978-1-5386-2659-7/18/\$31.00 ©2018 IEEE. International Conference on Information Reuse and Integration for Data Science.  
<https://doi.org/10.1109/IRI.2018.00046>
- [15] Min-Yuh Day, Chao-Yu Chen (2018). "Artificial Intelligence for Automatic Text Summarization". 978-1-5386-2659-7/18/\$31.00 ©2018 IEEE International Conference on Information Reuse and Integration for Data Science.
- [16] Hai-Tao Zheng1, Wei Wang (2018). "Automatic Generation of News Comments Based on Gated Attention Neural Networks". 2169-3536@ IEEE. Translations and content mining are permitted for academic research.
- [17] 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  
<https://doi.org/10.1145/1390156.1390177>
- [18] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," in Proc. 52nd Annu. Meeting Association Computational Linguistics, 2014, vol. 1, pp. 655–665.  
<https://doi.org/10.3115/v1/P14-1062>
- [19] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *J. Mach. Learn. Res.*, vol. 12, pp. 2493–2537, Aug. 2011.
- [20] R. Johnson and T. Zhang, "Semi-supervised convolutional neural networks for text categorization via region embedding," in Proc. Advances Neural Information Processing Systems, 2015, pp. 919–927
- [21] P. Wang, J. Xu, B. Xu, C. Liu, H. Zhang, F. Wang, and H. Hao, "Semantic clustering and convolutional neural network for short text categorization," in Proc. Annu. Meeting Association Computational Linguistics, 2015, pp. 352–357.  
<https://doi.org/10.3115/v1/P15-2058>
- [22] N. Kalchbrenner, E. Grefenstette, and P. Blunsom. "A convolutional neural network for modelling sentences," in Proc. 52nd Annu. Meeting Association Computational Linguistics, 2014, vol. 1, pp. 655–665.  
<https://doi.org/10.3115/v1/P14-1062>
- [23] Z. Tu, B. Hu, Z. Lu, and H. Li, "Context-dependent translation selection using convolutional neural network," arXiv Preprint, arXiv:1503.02357, 2015
- [24] H. Sak, A. Senior, and F. Beaufays, "Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition," arXiv Preprint, arXiv:1402.1128, 2014
- [25] D. Tang, B. Qin, and T. Liu, "Document modeling with gated recurrent neural network for sentiment classification," in Proc. Conf. Empirical Methods Natural Language Processing, 2015, pp. 1422–1432.  
<https://doi.org/10.18653/v1/D15-1167>
- [26] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," arXiv Preprint, arXiv:1412.3555, 2014.
- [27] A Y. N. Dauphin, A. Fan, M. Auli, and D. Grangier, "Language modeling with gated convolutional networks," arXiv Preprint, ArXiv:1612.08083, 2016.
- [28] W. Yin, K. Kann, M. Yu, and H. Schütze, "Comparative study of CNN and RNN for natural language processing," arXiv Preprint, ArXiv:1702.01923, 2017.
- [29] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," in Proc. 9th Int. Conf. Artificial Neural Networks, pp. 850–855, 1999.  
<https://doi.org/10.1049/cp:19991218>
- [30] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in Proc. Conf. Empirical Methods Natural Language Processing, 2013, pp. 1631–1642.
- [31] Richard Socher, Alex Perelygin, Jean Y Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. 2013. "Recursive deep models for semantic compositionality over a sentiment treebank". In Proceedings EMNLP. volume 1631, page 1642.
- [32] Samuel R Bowman, Gabor Angeli, Christopher Potts, and Christopher D Manning, 2015. "A large annotated corpus for learning natural language inference." In Proceedings of EMNLP. pages 632–642.

<https://doi.org/10.18653/v1/D15-1075>

- [34] Ying Wen, Weinan Zhang, Rui Luo, and Jun Wang. 2016. "Learning text representation using recurrent convolutional neural network with highway layers". SIGIR Workshop on Neural Information Retrieval.
- [35] Priyanka Thakur, Dr. Rajiv Shrivastava, "A Review on Text Based Emotion Recognition System " in Proceeding International Journal of Advanced Trends in Computer Science and Engineering,(IJATCSE) Vol-7(5), September - October 2018, pp. 67- 71  
<https://doi.org/10.30534/ijatcse/2018/01752018>
- [36] Lavanya Maddiseti , Ranjan K. Senapati , JVR Ravindra, "Training Neural Network as Approximate 4:2 Compressor applying Machine Learning Algorithms for Accuracy Comparison " in Proceeding International Journal of Advanced Trends in Computer Science and Engineering,(IJATCSE) Vol-8(2), March - April 2019, pp. 211-215.  
<https://doi.org/10.30534/ijatcse/2019/17822019>