A literature survey on recurrent attention learning for text classification

Ayesha Mariyam¹, SK Althaf Hussain Basha², S Viswanadha Raju³

¹Department of Computer Science and Engineering, Jawaharlal Nehru Technological University, Hyderabad, India
²Department of Computer Science and Engineering, A1 Global Institute of Engineering and Technology, Markapur, India
³Department of Computer Science and Engineering, JNTUH College of Engineering Jagtial, Nachupally, Jagtial, India

¹ayesha.mariyam84@gmail.com
²althafbashacse@gmail.com
³svraju.jntu@gmail.com

Abstract: With a rapid rise of complex data every year needs more enrichment in machine learning methods to provide vigorous and accurate data classification. Deep learning models such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) have accomplished to obtain better results in the domain of computer vision, object recognition, speech recognition and natural language processing compared to traditional machine learning algorithms. This paper mainly discusses about the blending of attention mechanism with various deep learning models for text classification which improves the performance of text classification task.

1. Introduction

Text classification is a task of natural language processing which assigns predefined labels to text according to its contents. Text classification is an essential component in many applications such as sentiment analysis, document classification, email filtering etc. Traditional methods such as bag of words (BOW), n-grams, Term Frequency – Inverse Document Frequency (TF-IDF) use statistical methods to represent the text as feature vectors and then classify these features using Naïve Bayesian, SVM, KNN [2]. These methods are straightforward and robust but they often neglect the context-dependent data or word sequence in passage which influences the classification accuracy and also have the problem of sparsity [5].

However, in parallel with the emergence of deep learning, NLP has procured new facets which allows machine to analyze, realize and extract meaning out of texts. Deep learning models such as Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) based on Long Short-Term Memory (LSTM) outperforms standard methods of text classification[3].

RNN mainly handles sequence learning problems. One of the key ideas of RNN is to connect previous information with the present processing unit, that is to have a store for the previous
information. Conceptually, RNN can record any long-term information but practically it is not possible due to gradient exploding and vanishing problems. The solution of this problem is to alter the framework of the RNN, by including gated activation function to obtain the trade-off between the old and the new time information. Both Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) designs include gated activation functions which makes the network to learn long-term dependency data and reduce the gradient exploding and vanishing problems[1].

Attention mechanism was initially introduced for Machine Translation. But, now it has become a main concept in neural network studies. Attention is used extensively within the Artificial Intelligence (AI) community as an essential unit of neural frameworks for huge number of purposes such as Speech and Computer Vision, Natural Language Processing, Statistical Learning etc. With the help of human biological systems, the instinct behind attention can be well explained. For instance, our visual processing system focus its attention on particular segments of the image, while neglecting other irrelevant information in a way that helps in perceptron. Similarly, in many tasks such as speech, vision or language, some portion can be more relevant than other[6].

Attention mechanism improves the capacity of RNNs. Attention mechanism makes the network to seek information from inputs and proceed to succeeding state on the accordance of it. Word vector representation is another factor which influences the working capacity of NLP tasks in RNN[3]. In Natural language Processing, the attention mechanism is normally used with the Encoder-Decoder Model[6].

The rapid growth in casting attention in neural network is essentially due to three reasons: Firstly, these models are now revolutionary for different problems such as Sentiment Analysis, POS tagging, Question Answering, Sentiment Analysis etc. Secondly, they have been abundantly used for improving understanding of neural networks, which are in other respects reviewed as black-box models. Third, they surpass some issues with Recurrent Neural Networks (RNNs) such as performance degradation as input size increases and inefficient computation evolved from sequential processing of input[7]. Therefore, in this paper we intent to impart a brief, yet comprehensive study on attention modeling for text classification.

The paper is organized as, section II discusses about the literature survey of different deep neural networks with attention mechanism for text classification. Section III concludes with a future scope.

2. Literature Survey
The following is the literature survey conducted on various attention models for text classification:

Guogiang Zhong, et al [1], proposed a RNN model named as Recurrent Attention Unit (RAU), which absolutely combines the attention mechanism in the center of Gated Recurrent Unit (GRU) by adding an attention gate. This upgrades the GRU’s capacity to remember long-term memory and quickly get rid of unimportant matter. RAU has an ability of obtaining useful information from sequential data by extensively determining a sequence of locations and concentrate on those regions during learning.

The experiments performed on Recurrent Attention Unit shows that it is appropriate for language modeling, sentiment analysis, image classification. RAU performance is outstanding compared to both LSTM and GRU.
Jin Zheng et al [2], proposed a hybrid bidirectional recurrent convolution neural network attention-based model, named as BRCAN to address the challenges with the multi-class text classification and the fine-grained sentiment analysis. This model integrates the Bi-LSTM and CNN constructively with the benefit of the word2vec model and attention method for attenuated text classification.

In BRCAN, a Bi-directional recurrent construct acquires contextual information and vulnerability of sentences and a maximum pool layer of CNN examines which words are essential for text classification. Four text classification datasets, which are separated into topic classification and sentiment classification are used to assess the proposed model. The topic classification dataset has Yahoo! Answers and Sogou News. The sentiment classification dataset consists of Douban movie top250 short reviews full and Douban movie top250 short reviews polarity, Yelp Reviews full and Yelp Reviews polarity. The BRCAN model achieves outstanding performance in comparison with other alternative models on all tasks.

The proposed framework is usually not restricted to text classification problem but can also be implemented for other approaches [8] which is significant for future research.

Ankit Kumar et al [3], proposed a new mechanism of attention based recurrent neural network for sentence classification. This model is designed on attentional implementation of Bidirectional and Vanilla Recurrent Neural Network framework on top of two kinds of recurrent cells, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) for sentence classification. These are called as Attentional LSTM(ALSTM) and Attentional GRU(AGRU), and Attentional Bi-LSTM and Attentional Bi-GRU(ABGRU).

Word2Vec model is used and the resulting feature vectors are fed as input to this model. The words are initialized randomly which are not present in word2vec model.

To check the efficiency of this proposed model, its performance is analogized with the other methods on six datasets namely CR, MPQA, MR, Subj, SST-1 and SST-2 which confirms that this model has a capability to store long-term dependencies in the problem of sentence classification.

Jun He et al [4], proposed a hierarchical attention networks for document classification which has two distinguishing properties. Firstly, it has a hierarchical design which reflects the hierarchical structure of documents. Secondly, it has two stages where attention mechanisms are applied that is at the word-level and sentence-level.

This model gradually builds a document vector by merging important words into sentence vectors and then merging important sentences vectors into document vectors. Experiments done with this model on six enlarged datasets outperforms the previous models by considerable amount.

Zichao Yang et al [5], has proposed a new document classification model, which is based on recurrent attention learning which needs to glimpse only a few groups of words. This is preferable for long document because it does not make use of all detailed information. With respect to the coarse document text, only a few glimpses are utilized to make the selective feature for classification. This problem is addressed as a reinforcement learning task in which the actions are motivated for right classification and the actions are demotivated for wrong classification.

This method intent to prepare an agent which handles how to thereafter look at the next position in a given time steps to get correct classification.

This model is evaluated with the arXiv papers data set which excels in performance in comparison with CNN model which employs different length convolution kernels.

Motivated by [4], in future, there is a scope of improving the controller network by inspecting the hierarchical attention mechanism.
3. **Conclusion**

The main intention behind this study is to give brief idea on text classification using Attention Mechanism. The study shows that all of the models can be applied to increase the accuracy of text classification over different datasets but with a constraint of tuning hyper parameters of the model. The aim behind this survey is to develop an appropriate and effective model for text classification.

**References**

[1] Guogiang Zhong, Guohua Yue and Xiao Ling,2018, Recurrent Attention Unit, *arXiv: 1810.12754*, [Online]. Available: https://arxiv.org/abs/1810.12754.

[2] Jin Zheng, Limin Zheng,2019, A Hybrid Bidirectional Recurrent Convolutional Neural Network Attention-Based Model for Text Classification, *IEEE Access*, 7, 106673–106685.

[3] Ankit Kumar and Reshma Rastogi (nee Khemchandani),2018, Attentional Recurrent Neural Networks for Sentence Classification, *Innovations in Infrastructure, Advances in Intelligent Systems and Computing* 757, 549-559.

[4] Jun He, Liqun Wang, Liu Liu, Jiao Feng, Hao Wu,2019, Long Document Classification from Local Word Glimpses via Recurrent Attention Learning,” *IEEE Access*, 7, 40707-40718.

[5] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, Eduard Hovy, 2016, Hierarchical Attention Networks for Document Classification, *Proceedings of NAACL-HLT*, 1480–1489.

[6] Sneha Chaudhari, Gungor Polatkan, Rohan Ramanath, Varun Mithal, An Attentive Survey of Attention Models, *arXiv:1904.02874v1*, [Online]. Available: https://arxiv.org/abs/1904.02874.

[7] Dzmitry Bahdanau, Kyung Hyun Cho, Yoshua Bengio, 2016, Neural Machine Translation by Jointly Learning to Align and Translate, *arXiv:1409.0473v7*, 1-15.

[8] Xiaoyu Guo, Hui Zhang, Haijun Yang, Lianyuan Xu, Zhiwen Ye, A single attention-based combination of CNN and RNN for relation classification, *IEEE Access*, 4.