A REVIEW ON RECENT ADVANCES IN DEEP LEARNING FOR SENTIMENT ANALYSIS: PERFORMANCES, CHALLENGES AND LIMITATIONS

Md Shofiqul Islam1,2, Ngahzaifa Ab Ghani1,2*, Md Manjur Ahmed3

1Faculty of Computing, Universiti Malaysia Pahang, 26300 Kuantan, Pahang, Malaysia
2IBM Centre of Excellence (Universiti Malaysia Pahang), Cybercentre, Pahang Technology Park, 26300 Kuantan, Pahang, Malaysia
3Faculty of Computer Science and Engineering, University of Barisal, Barisal 8200, Bangladesh.
shafiqcseiu07@gmail.com, zaifa@ump.edu.my, manjur_39@yahoo.com

Abstract: Now days the horizons of social online media keep expanding, the impacts they have on people are huge. For example, many businesses are taking advantage of the input from social media to advertise to specific target market. This is done by detecting and analyzing the sentiment (emotions, feelings, opinions) in social media about any topic or product from the texts. There are numerous machine learning as well as natural language processing methods used to examine public opinions with low time complexity. Deep learning techniques, however, have become widely popular in recent times because of their high efficiency and accuracy. This paper provides a complete overview of the common deep learning frameworks used in sentiment analysis in recent time. We offer a taxonomical study of text representations, learning model, evaluation, metrics and implications of recent advances in deep learning architectures. We also added a special emphasis on deep learning methods; the key findings and limitations of different authors are discussed. This will hopefully help other researchers to do further development of deep learning methods in text processing especially for sentiment analysis. The research also presents the quick summaries of the most popular datasets, lexicons with their related research, performance and main features of the datasets. The aim of this survey is to emphasize the ability to solve text-based sentiment analysis challenges in deep learning architectures with successful achievement for accuracy, speed with context, syntactic and semantic meaning. This review paper analyzes uniquely with the progress and recent advances in sentiment analysis based on existing advanced methods and approach based on deep learning with their findings, performance comparisons and the limitations.

Keywords: Sentiment Analysis(SA); Text; Deep learning; Emotion Recognition (ER); Classifiers; Neural Network (NN).

I. INTRODUCTION
Social media is leading a great role in content sharing. People express their feelings towards any particular topics in social media in order to alert other users, gain attention, spread information, or mostly just sharing their opinions whether it is positive or negative. This is why users’ opinions are very useful in predicting the interest of users on any particular topics of interest. As useful as it is, it is also nearly impossible to manually and systematically analyze opinions in social media from thousands of comments. This is where data mining and machine learning come into the picture. Sentiment analysis is a part of data mining that analyzes the text using the natural language processing and also the computational linguistics to extract the people’s opinions & feelings from subjective information obtained from online platforms like social media, shopping websites or applications. It also analyzes in-depth to the strength of the feelings, from parameters known as sentiment score. There is a close connection
between emotions recognition and opinion mining. It also covers machine linguistics, text mining and the storage of natural languages. It can help society to explore an individual in psychology through researching feelings, but it involves word abstraction. [1]. Sentiment analysis is done based on the reviews or comments from humans on social media or other sources. Nevertheless, feeling research are sometimes difficult because text can be tactical. The sentiment analysis has categorized mainly into three levels [2], which are document level, sentence level and the last one is aspect levels. Sentiment analysis in document-level is the simplest way to analyze public sentiments. It finds overall polarity and cannot find individual emotion for different entity. It aims to categorize the entire opinion document into one object unit, such as a book, business or hotel [1] where the whole file is used to assess whether it is positive or negative. On the other hand, Sentence-level sentiment analysis handles sentiment at sentence level with better performance for subjectivity and objectivity but it is not suitable for complex sentence [3]. The last type of sentiment analysis is aspect level sentiment analysis that is able to handle negation for simple and short sentence but weak in performance for negation in long and complex [3]. Aspect does not consider the language structure such as articles, sentences and clauses as compared to the other two levels. It helps to determine a sentence by separating feelings (negative or positive) as well as its objective (object).

A. Applications of sentiment analysis with the model of deep learning

There are a lot of existing research and applications on sentiment analysis. In most cases, sentiment analysis is used for review analysis on social media data, financial market prediction, event prediction etc. Table shown gives the recent application of sentiment analysis in different fields.

<table>
<thead>
<tr>
<th>Name of Applications of Sentiment analysis</th>
<th>Authors and Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road Traffic Congestion</td>
<td>[9] -2020</td>
</tr>
<tr>
<td>Event prediction</td>
<td>[10] -2018</td>
</tr>
<tr>
<td>Restaurant Review</td>
<td>[16] -2020</td>
</tr>
<tr>
<td>Stock Market prediction</td>
<td>[17] -2018</td>
</tr>
</tbody>
</table>

Table-1: Application of Sentiment Analysis.

B. Sentiment analysis architecture with model of deep learning

In sentiment analysis there are some basic steps. This section deals with the description of the basic architecture for sentiment analysis with model of deep learning as shown in Figure-1.

Figure-1: Deep learning architecture for SA

The first step is data collection then followed by data pre-processing using NLP tools then tokenization of each word to convert into integer. The tokenized integers will be converted into real valued vector before forwarded to deep learning framework where tuning of different parameters to predict the sentiment class is done.

C. Text Embedding

The inputs cannot be taken directly from the deep learning framework, which is why data must be embedded. Several embedding strategies are possible. The embedding layer includes a vector representation based on the semantic connection to the relevant word with each incoming text word. Tabel-2 shows the developer and the related research, with some common text encoding briefly discussed.

<table>
<thead>
<tr>
<th>Used Text Embedding with author</th>
<th>Features of text embedding</th>
<th>One related Reference and Year</th>
<th>Details of using embedding and Deep learning model on related research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word2vec [20]</td>
<td>It used the distributed representation of each words</td>
<td>[21]-2019</td>
<td>Uses Semanitic and emotion embedding for BiLSTM and CNN model</td>
</tr>
<tr>
<td>Bag of words[22]</td>
<td>Words(multiset) of a text is represented as bag of words (BOW)</td>
<td>[21]-2016</td>
<td>Denoising auto encoder model uses target-oriented features of BOW with loss.</td>
</tr>
<tr>
<td>Concatenated embeddings with SSWE and \textit{word2vec}</td>
<td>SSWE provided a very strict learning by the analysis of the positive and also negative n-gram</td>
<td>[23]-2016</td>
<td>Uses Bidirectional and tri-way gated neural network with target oriented interaction of entity</td>
</tr>
</tbody>
</table>
Word embeddings with pre-trained word2vec and the Skip-gram

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>FastText[20]</td>
<td>The word representation is taught effectively when utilizing the details at the character level then this tool fastText is also works for all of the rear words</td>
<td>[25]-2018</td>
</tr>
<tr>
<td>GloVe[26]</td>
<td>Glove vectors is unsupervised, and the terms are represented by the vector for each words. The terms are identified by word similarity distance as well as semantic space.</td>
<td>[16]-2020</td>
</tr>
<tr>
<td>BERT[27]</td>
<td>BERT pretrains dual-directional representations of unlabelled data in each of these layers</td>
<td>(Gao, Feng, Song, &amp; Wu, 2019)</td>
</tr>
</tbody>
</table>

**Table-2: Summary of text embedding in deep learning based SA.**

### D. Deep Learning Model

There are three basic steps in the deep learning-based model which are application, architecture and preparation of inputted text using different embedding approach then feed forwarded to the deep learning model named CNN, RNN based model and finally predict NLP application. This section gives overall information about the tools and methods used in deep learning. Figure-2 shows the overall deep learning model for sentiment analysis.

**Figure-2: Deep learning based model for sentiment analysis**

### E. Convolutional Neural Network:

A convolution layer learns to convert small data parts into high-level characteristic vectors. In 2014, Kim has developed and applied CNN [28]. Figure-3 shows basic structure of a CNN model with its related layers. Convolutional neural networks (CNNs) are generally flexible with larger inputs and very broad scale. Unlike conventional ANNS, CNNs include inputs, completely connected as well as output layers, but they do have convolutional and pooling layers including additional layers which are essential to CNN effectiveness. Such layers involve the processing of filtered feature maps. A feature map is source data representations. In back propagation time, the filters used in both convolution and pooling layers.

**Figure-3: CNN model [29].**

### F. Long Term Short Term Memory(LSTM)

LSTM is generally an RNN prolongation that requires inputs to be stored for a long period. LSTM has an advanced memory, as opposed to RNN's basic internal memories. Figure-4 shows basic architecture of LSTM. The memory content can be read, written and erased. Therefore, it addresses RNN's drawback that suffered from vanishing point. LSTM will determine which knowledge to remember and what to forget. The memory may be gated in LSTM. It has three gates named as input, forget and the output gates. Basic equation of LSTM is given below.

**Figure-4: LSTM memory cell [30].**
i_t = \sigma(w_i[h_{t-1}, x_t]+b_i), f_t = \sigma(w_f[h_{t-1}, x_t]+b_f), o_t = \sigma(w_o[h_{t-1}, x_t]+b_o)

Here, i_t is for input gate, o_t for output gate and f_t denotes forget gate. \sigma represents the activation function, w_i is used to indicate weights of different gates(x), h_{t-1} is the output from the previous LSTM with time stamp t-1, x_t is the current time stamp, b used for bias value in different gates.

G. Attention mechanism with RNN

Attention processes have become popular in the NLP with a significant paper regarding machine translation[31]. RNN works with a single hidden layer in order to obtain an intuition for the attention process. The goal of the network for the attention function is to extract the meaning of each hidden state and to calculate the value of a weighted summation of features. Here Figure-5 is the schematic from Bahdanau’s Attention Method. An LSTM series for each input sentence is generated throughout the Bidirectional LSTM used in this (h1, h2, hTx). All the h1, h2, ..., etc. vectors used during their function simply are just the concatenation with hidden states as forward and backwards in the encoder. In basic terms, the Tx indicates words number in the input sentence is represented by the vectors variables h1, h2, h3, hTx. Only the last condition of the encoder, LSTM (hTx in this case), works as context vector on the basic encoder as well as decoder method.

![Figure-5: Attention based model](image)

the weights will be learned by a feed-forward NN with its equation given below. The context-based vector c_i for the produced output word is y_i that is generated by the weighted summation of the annotations and the weights a_ij is calculated using a softmax function with the given equation:

\[ a_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})} \]

\[ e_{ij} = a(s_{i-1}, h_j) \quad c_i = \sum_{j=1}^{T_x} a_{ij} \cdot h_j \]

Here e_{ij} is the product of a neural feed forward of network that is represented by the functions to track the alignment from input in j as well as output in i.

H. Gated Recurrent Unit(GRU)

Gated Recurrent Units (GRUs) was developed by Kyunghyun Cho in 2014 as a gating function with Recurrent Neural Network(CNN)[32]. There is a lot of similarity between a LSTM and GRU and GRU has limited parameters as compared with LSTM. Generally, GRU performs good enough than LSTM. GRU has use of gates for high performance namely two gates those are reset and update gate. Each reset gate defines how the new inputted data is merged with the prior information, while the update gate defined which prior memory will be maintained. GRUs has no background conditions (ct) like LSTMs. Update gate and reset gate makes forget gate and linked with past hidden layer. Therefore, in a GRU, the purpose of the LSTM reset gate is essentially separated into the reset and upgrades that shown in figure-6.

![Figure-6: Gated Recurrent Unit](image)

Here x is used for input and r for reset gate and h for hidden state and z for update and sigma for

\[ r_t = \sigma(W_r x_t + W_{hr} h_{t-1} + b_r) \]

\[ z_t = \sigma(W_z x_t + W_{hz} h_{t-1} + b_z) \]

\[ h_t = \tanh(W_{h} x_t + W_{h0}(r_t \odot h_{t-1}) + b_0) \]

\[ h_t = z_t \odot h_{t-1} + (1-z_t) \odot h_t \]

here W represents matrices, b represents model parameters, \sigma represents sigmoid function and the symbol \odot for the multiplication.

I. Capsule Network of Recurrent Neural Network(RNN)

The hierarchical relationship among local features, which can misclassify concepts due to their characteristics, could not be modeled in CNN. With max pooling in CNN, some important information would be lost, because the active neurons will just transfer to the next stage. Capsule networks have also been suggested to overcome these constraints. These networks tackle the spatial relations among entities by the use of dynamic routing with capsules. This is much better when compared to CNN's max pooling service. The dynamic routing is used to train the neuronal vectors of capsule networks; the dynamic routing is used to replace the conventional neural network cell node. The capsule networks allow it to be with relatively lower information than most of the other models of neural network. Main role of the capsule network to establish spatial relations and also the location directions focused on the conventional neural network, and by merging invariance with covarability to recognize objects. A first level capsule chooses to give its production towards
next level capsules which vectors provide a broad scalar element by lower level capsule projection. Figure 7 shows routing process in capsule network.

Figure-7: Routing process in Capsule network layer [33]

II. RELATED WORKS ON SENTIMENT ANALYSIS BASED ON DEEP LEARNING

Nowadays newer approaches using deep learning architecture performed with higher accuracy. Based on large-scale dataset, many deep learning approaches will systematically obtain the theoretically somaticized and syntactic features of texts that effectively tackle the shortage of artificial content engineering with improved intensity and precision. This section discusses and reviews the recent progress and development in deep learning-based sentiment analysis based on category with their analysis type, journals tasks, data, approach, sentiment analysis method, languages, advantages, and disadvantages. There are so much related works for sentiment analysis. Traditional methods are subdivided into following types: CNN, RNN. CNN have been used for sentence classification[28], in this method, the textual features can be extracted easily and relational research has made a lot of progress in sentiment analysis. Some CNN applications include use of character level for text classification [34] very deep CNN used for text classification [35], and Twitter abusive language detection by using CNN [36]. However, CNN is still not also fully satisfactory because it cannot capture long ranged feature and it does not recognize the important dependent features information of its function and spatial position details. The above limitations of handling long ranged dependent feature of information have been almost solved by the Recurrent Neural Network (RNN) [32] [37] [38] and capsule neural network [8] [39]. RNN has been widely used in different fields for classification purposes with better result.

By studying the previous knowledge or features, RNN may predict results with long-range features. Because there are already certain flaws in the initial RNN, like gradient absence and gradient dispersion, authors have suggested a variety of RNN variants like LSTM: The LSTM is able to monitor long-term dependencies in sequences with storage units with gate structures which determine how data in storage can be used and modified in order to get more data to improve model calculation advantages.[32, 40], Dai introduces sequence learning in recurrent neural network for pretrained LSTMs model [41]. Sayeed developed a deep neural networks with GRU and maximum pooling to classify the overlapping sentiments with high accuracy Bi-GRU [42]. Hinton developed the idea of the "capsule" at first with the vigorous creation of CNN and RNN[43]. Another works with capsule neural network with attention mechanism done by Du et al.[33].

From the analysis of this section and the table-3 below of related works it can be concluded that existing recent work in the field of sentiment analysis from text has good performance but still is very lacking in terms of coherence, context, semantic meaning handling, negation, modifiers, intensifier of the sentence. In recent deep learning-based method gives higher accuracy with handling independent textual features. There are also some limitations in deep learning such as handling of context and syntactic properly. Automatic emotion recognition remains an important research area to explore. However, most of the technique’s has limitations. Damaged data cannot produce good result in measuring exact sentiments. So data should be secure when it need to be transferred from a source to another, a multi-level security of the sensitive text is ensured with encryption and compression algorithm[44].

<table>
<thead>
<tr>
<th>Author and Year</th>
<th>Type</th>
<th>Journals</th>
<th>Task</th>
<th>Lexicon or dataset</th>
<th>Approach</th>
<th>Sentiment analysis features</th>
<th>Text</th>
<th>Performance</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>[16] - 2020</td>
<td>Hybrid</td>
<td>Neurocomputing, Elsevier</td>
<td>Opinion Mining</td>
<td>Works with five datasets: laptop2014, restaurant2014, restaurant2015, restaurant2016, and twitter</td>
<td>Glove, Vector, Encoder, Cosine Similarity</td>
<td>Attention Mechanism, done attention weights using mean value, Cosine similarity is used for position analysis for context and aspect</td>
<td>English</td>
<td>Accuracy, Macro-F1 percentage on data: Laptop2014(78.13, 73.20) Restaurant2014(84.38, 71.31) Restaurant2015(82.65, 69.10) Restaurant2016(85.87, 73.28)Twitte r(76.56, 72.19)</td>
<td>This method used Multi headed attention mechanism that increase the performance for sentiment classification</td>
<td>Its mean value-based tasks sometimes misses context, Uses cosine similarity that it cannot able to handle with the irrelevant portion of text</td>
</tr>
<tr>
<td>Year</td>
<td>Type</td>
<td>Journal/Conference</td>
<td>Corpora</td>
<td>Models</td>
<td>Tasks</td>
<td>Languages</td>
<td>Metrics</td>
<td>Comments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>------</td>
<td>---------------------</td>
<td>---------</td>
<td>--------</td>
<td>-------</td>
<td>-----------</td>
<td>---------</td>
<td>----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td>Hybrid</td>
<td>IEEE Access</td>
<td>Laptop review, Restaurant review and Twitter dataset</td>
<td>BERT with target dependency approach Glove Vector, Word2Vec</td>
<td>Uses Word Piece tokenizer, segment, and position embeddings and encoder layer for classification with BERT</td>
<td>English</td>
<td>Performance on Laptop data: Accuracy: 78.87 ±1.13, Micro F1:74.38 ±1.39 Restaurant data: Accuracy: 83.87 ±2.7 ,MicroF1:79.61±0 .79 Twitter data: Accuracy: 77.31 ±0.79, MicroF1: 75.56±0.93</td>
<td>Did not consider the whole sentence, but focuses on target terms instead</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td>Hybrid</td>
<td>IEEE Access</td>
<td>Movie Review data, NLPCC2014 dataset</td>
<td>BiGRU and CNN model and</td>
<td>CNN used for feature extraction BiGRU is used for and Capsule network can extract independent text features</td>
<td>English</td>
<td>82.55% accuracy for Movie review data 87.84% accuracy for NLPCC data</td>
<td>This method performs well but attention-CNN layer is not rich enough with its operations for better performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td>Semi-supervised</td>
<td>IEEE Access</td>
<td>ISEAR and other 9 datasets</td>
<td>Deep learning</td>
<td>Uses CNN, Bi-LSTM, Word embeddings, Word2Vec, GloVe, and FastText</td>
<td>English</td>
<td>Accuracy: 74.6% for ISEAR dataset, Good performance for other 9 datasets</td>
<td>Performance is limited with the data of word embedding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td>Supervised</td>
<td>Computation and Language, Cornell University Publisher</td>
<td>SemEval-2019 Task-3 data collections of labeled conversations</td>
<td>Deep learning: Bi-LSTM</td>
<td>Uses ASGD (Average Stochastic Gradient Descent) for training, uses attention based AWD-Bi-LSTM for classification.</td>
<td>English</td>
<td>Accuracy: 75.82% for SemVal2019 data and also shown its different model data</td>
<td>Extract sad, happy, angry emotions in conversion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td>Hybrid</td>
<td>Conference paper, ACL Anthology, ACL web</td>
<td>Kaggle toxicity detection dataset</td>
<td>Capsule Network with Dynamic Routing, LSTM</td>
<td>Word embedding, Focal Loss for better performance in Toxic comments classification</td>
<td>English</td>
<td>98.46 Accuracy on Kaggle data for toxic comment classification</td>
<td>It performs well for domain dependent dataset for Kaggle toxicity dataset</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td>Deep learning</td>
<td>Knowledge based system, Elsevier</td>
<td>Amazon multi-domain sentiment dataset, Sanders</td>
<td>Attention mechanism</td>
<td>Uses two level modules named as domain module and sentiment</td>
<td>English</td>
<td>Accuracy Amazon Books 87.75% DVD 86.58% Electronics 87.50% Kitchen</td>
<td>Domain dependent and do not perform for multi domain and multilingual text.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
From the review of deep learning in this article, it can be concluded that the deep learning architectures have shown outstanding results and important advances in sentiment analysis, there are still some disadvantage in using the following algorithms:

1. In order to ensure that a machine achieves the required output, most deep learning strategies allow several labelled data to be trained. Thus, for sentiment analysis research, a big set of dataset is necessary for training the deep learning architecture in order to predict the class labels correctly. Huge quantities of data can be exceedingly complicated and cumbersome to collect and label.

2. Unlike conventional machine learning or lexical approaches, which display what features are chosen to predict a certain feeling, it is difficult to find out what the real explanation for the neuronal network, by finding at weights in various stages, for predicting multilevel sentiment of text. It makes its challenging to achieve the result about the prediction analysis of the model of neural networks, as they function works looks like “black box.”

3. Deep learning approaches such as CNN need to be tuned on initial parameters. You see this in Stoyanovsky et al.[49]. The network's efficiency therefore relies on the value of the hyper parameters on the networks. This is also a difficult job to determine the optimum hyper parameter values.

4. The time it takes to train them is also really nice as there are a huge number of the parameters in deep learning. In addition, to increase performance[50], they need high performance based hardware such as GPUs and wide RAM.

III. CHALLENGES, LIMITATIONS AND FUTURE WORK IN SENTIMENT ANALYSIS WITH THE MODEL OF DEEP LEARNING

It is not possible for a machine to work like human to recognize sentiment. However, existing recent of sentiment analysis from text has generally performed with good accuracy. They are however still lacking in terms of lacking coherence, context, semantic meaning handling, negation, modifiers, and intensifier of the sentence. Context based task is giving some satisfaction for this problem. Lexicon or dictionary-based approaches can handle grammatical syntax but also have some limitations such as low accuracy, higher time complexity, dictionary, and domain dependency. Unsupervised based works give adaptability, simplicity, lower complexity but these methods also come with the limitations in time complexity and accuracy. But, Machine learning approaches like Tf-idf, Naïve Bayes (NB), Random Forest (RF), Support Vector (SVM), Logistic Regression (LR), Bayesian, k-means, Maximum entropy classification, Conditional Random Field (CRF) classifier work better for faster time but have limitation of handling semantics and dependency of words in the sentence. In recent deep learning-based approach CNN, LSTM, GRU, BERT, Capsule neural network gives higher accuracy by handling independent text features. There are also some limitations in deep learning such as handling of context and syntactic properly. Deep learning and quantum deep learning is the current trends in the area of sentiment analysis. Now live sentiment analysis is also the trends task for a game, product or other.

IV. CONCLUSION

This article gives a systematic review and analysis on deep learning methods of text sentiment analysis. It mainly introduces several different deep learning methods with textual data for different categories, and further summarizes and analyses their benefits, disadvantages, limitations and applicability etc. Sentiment analysis from test, image, speech, and video is very important in Human Computer Interaction. For social media, the text-based sentiment analysis plays a vital role. From this review paper we can conclude that deep learning method gives higher accuracy than all other methods. But in the case to
handle the coherence and semantic in sentence the knowledge-based approach is better but has the limitations of accuracy, time and space complexity. On the other hand, ontology-based sentiment analysis is good to handle text properly but it is time consuming as compared to all other approaches. Although machine learning based supervised technique is faster and more accurate but this type of method cannot handle negation, intensifier or modifier clause in the sentence. For this case the unsupervised knowledge-based approach and deep learning is good than all other methods.

REFERENCES


