Proceedings of the Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)
IEEE Xplore Part Number: CFP21OSV-ART; ISBN: 978-1-6654-2642-8

Analysis on Sentiment Analytics Using Deep Learning Techniques

*Dr M Anusha[1], R Leelavathi[2]

[1]Assistant Professor, PG & Research Department of Computer Science, National College (Autonomous), Bharathidasan University, Tamilnadu, India
[1]Email id - anusha260505@gmail.com

[2] Research Scholar, PG & Research Department of Computer Science, National College (Autonomous), Bharathidasan University, Tamilnadu, India
[2]Email id - leelavathi4794@gmail.com

* Corresponding Author – R. Leelavathi

2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) | 978-1-6654-2642-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/I-SMAC52330.2021.9640790

Abstract-Sentiment analytics is the process of applying natural language processing and methods for text-based information to define and extract subjective knowledge of the text. Natural language processing and text classifications can deal with limited corpus data and more attention has been gained by semantic texts and word embedding methods. Deep learning is a powerful method that learns different layers of representations or qualities of information and produces state-of-the-art prediction results. In different applications of sentiment analytics, deep learning methods are used at the sentence, document, and aspect levels. This review paper is based on the main difficulties in the sentiment assessment stage that significantly affect sentiment score, pooling, and polarity detection. The most popular deep learning methods are a Convolution Neural Network and Recurrent Neural Network. Finally, a comparative study is made with a vast literature survey using deep learning models.

Keywords: Sentiment Analytics (SA), Natural Language Processing (NLP), Deep Learning (DL), Recurrent Neural Network (RNN), Convolution Neural Network (CNN).

I Introduction

Data is a prevalent source of social media communication. Data in the form of sentences and sentiments, individuals share their views [1]. Web 2.0 has contributed to the spread of blogs, forums, and social networks online, allowing people to discuss any topic and express their ideas. A social network for instance peeps or Facebook has become a common way to learn about user views and has a broad range of applications [2]. The way to recognize the individual polarity of the text is sentiment. It decides that positive, negative, or neutral is the given document. Sentiment analytics is the process of using techniques for natural language processing to extract individual knowledge from the text [3]. Sentiment analytics can be classified at various levels. Sentiments are classified into various forms, namely feature level, sentence level, and document level [4]. A sense dependent upon the position polarity for each element is defined by the identification of the character level. On the other hand, based on their intensity of meaning about certain topics, categorization at the level of the sentence in each word. It is mutual opposition of this text created in the class aside document level. The entire file is labeled as positive or negative and neutral [5]. In order to acquire the natural language processing (NLP) concerned with the processing and interpretation of human languages in computational terms. Natural Language Processing using linguistic techniques is sentence segmentation, word tokenization, removal of stop words, part-of-speech tagging stemming. The current present-day sentiment analytics has become more mainstream as a meaningful step forward in deep learning in the field of common language preparation [6]. The key regions of fundamental issues are language demonstrating, the measurement associations between words that naturally occur morphological handling, the segmentation of meaningful words, and recognize the real piece of word stream, grammar preparing, or parsing that create sentence structures as potential catalysts of semantic process. The fields of application include topics such as extracting useful knowledge, translating text within and between languages, comprehending written works, and answering questions automatically by referencing answers and classifying documents [7]. For sentiment analytics of texts are important, and Bag of Words (BoW), Word Embedding is a widely-used model. In this sequel, artificial intelligence is introduced to design a model that could perform. Sentiment analytics using automatic feature extraction. The feature extraction can make systemized with deep learning techniques. [8]. It is the subset of machine learning. This system learns many layers of data interpretations or features and delivers a predictive output that is state-of-the-art. The neural network model of sentiment examination is arranged into three classes, in particular CNN, RNN, and hybrid models. Also the achievement of deep learning in numerous application fields and has also been used in sentiment analytics [9]. This review paper focuses on sentiment analytics that combines natural language processing and deep learning methods. The sentiment rank is based on the entities, pooling, sentiment polarity, and themes categories within a sentence or phrase. This research uses deep learning models such as recurrent neural networks and convolution neural networks to help solve the issue of sentiment analytics [1]. This paper is organized as follows section 2 briefly introduces the related work section 3 comparative study section 4 is extended with results and the discussion ends with a conclusion.

II Literature Survey

Kleenankandy, J. et al. [10] showed an improved LSTM architecture was implemented, called this relationship, which can show the connection between a sequence of two data sources utilizing a control input. These models are called Typed Dependency Tree that uses both forms of the term dependency parse and type of encoding the meaning of phrase into a dense vector. The DT-LSTM model was produced Precision of 88.4%, Recall of 84.8%, F1–score of 86.6%, and accuracy of 86.6%. Basiri, M.E. et al. [11] Attention-focused CNN-RNN Deep bidirectional model for temporal data flow in both information ABCDM can derive both past and future context with the use of two different models LSTM and GRU layers. The consideration meaning is often applied to outputs of functioning in two directions have more or less effect on individual terms. It incorporates complexity and pool mechanisms to minimize the dimensionality of characteristics and extract stance essential features. The efficacy is measured in determining the polarity of sentiment, which is the most common and important role of sentiment analytics. The proposed ABCDM model was performed Recall of 99.08%, Precision of 99.04%, F1-score of 96.94% and, the accuracy of 96.87%. Huang, M. et al. [12] showed that a neural network of context-dependent, lexicon-based convolution is used to successfully separate the attributes and intensity of focus from words and text. To create this convolution neural network, the model is prepared to utilize the polarization for each word with a co-event example of words and marks. The algorithm has produced sentiment strength of document for six different datasets accuracy of 83%, 77%, 65%,74%, 70%, 98%. Shuang, K. et al. [13] explored a function distillation network for noise reduction and an aspect-relevant view of the sentiment highlights for filtering. The aim of the Novel double gate mechanism is to designed includes the points of view and their respective contexts in fine granularity. In front of the double-gate mechanism add a relevant nonlinear prediction layer to deliver aspect-specific word representation, and helps the twofold entryway component to accurately separate between feeling qualities of a similar context word corresponding to the numerous viewpoints. The FDN model has produced three different gating mechanisms on ATSA accuracy of 76.8%, 82.3%, 73.7%, and macro F1 score of 72.5%, 75%, 72.2%. Senthil Kumar, N. K. et al. [14] proposed an enhanced richer presentation, not only binary and ternary classification it goes deeper in classifying the collection of text from the social network. To verify the multiple sentiment prediction classes checked for the short text in sentiment data based on polarity. The algorithm has produced an accuracy of 98.6%. Sharma, A.K et al. [15] the proposed model has cleaned the information and creates significance from the Word2Vec model used for the convolution layer to extract highlights for short sentences. The proposed model was built on training accuracy of 99.07% and testing accuracy of 82.19 %. Abid, F. et al. [16] projected a Convolutional Neural Network combined with a recurrent neural network based on weighted attentive pooling. The regular pooling activities contain a few layers that are just fit for capturing proper highlights. To enable sentiment analytics such as Word2vec, Fast Text, and Glove to produce a dense powerful linked representations following the long-term dependencies on a single RNN layer obtained explicitly by part of speech tagging with verbs, adverbs, and noun. The model has produced an accuracy of 89.67%. Zhou, M. et al. [17] proposed a text classification model based on the double word embedding techniques are Glove and Word2vec. Models combined with the content form a combinatory information double convolution neural network channel. The initial word vector is consistently trained and modified the classification model with a single input vector representation, based on the word vector fine-tuning technique. The glove and word2vec model has classified the accuracy of 94.8%. Samat, N.A. et al. [18] extract features from sentiment analytics used for the CNN algorithm. This algorithm used for three different pooling functions reduced the noisy data during the training process. Finally to capturing the semantic features. The three different pooling functions are produced the better accuracy of 97.73%, 93.05%, and 95.58%. Naderalvojoud, B. et al. [19] proposed two approaches for opinion to learn embedding techniques. The first approach encodes the word sentiment knowledge set up pre-trained importance of words, and the second approach semantic context creates synthetic sentiment context for inserting models. On the various sentiment classification tasks used for Skip-gram and Glove models. The proposed model based on two approaches has produced an accuracy of (87%,47.4%) and Macro F1 score of (91.3%, 69.4%). Jang, B. et al. [20] noted that Bi-LSTM+CNN hybrid models supporting an attention function to additional enhance accuracy and decrease the number of learnable parameters. It is used to extract specific characteristics in a sentence from various locations with reducing the number of input features. To remove the context-oriented information from the highlights acquired from the layer of convoluted. The length of the attention mechanism is improved the distribution process. The proposed hybrid model was built maximum accuracy of 91.41%, average accuracy of 90.26%, F1 score of 90.18%, Recall of 90.57%, and precision of 80.97%. Xu, J. et al. [21] planned that text characterization using CNN based on the neural network called DE-CNN. To coordinate with context-relevant ideas into a convolution neural network. First and foremost used for two layers are extracting their concepts and meaning individually. That point of short text classification is implemented into a text representation. The DE-CNN model was provided three different datasets have classified the accuracy of 94.6%, 84.6%, 88.9%. Hassan, A. et al. [25] Deep learning network pre-trained parameters were trained in word embedding then to activating the model. The new framework mixes accessible information with different feature maps are learned by a convolution layer, and the long-term dependencies learned by long-flitting memory at the last level. This technique achieves excellent results with minor hyperparameter tuning for static vectors using multiple sentiment analytics benchmarks dataset. The algorithm has performed the three different datasets, and the accuracy of 93.3%, 48.8%, and 89.2%. Liao, W. et al. [26] presented a new model for sequence-to-sequence learning for multi-label classification. The two unmistakable neural network modules are called encoder and decoder individually. Using the convolution neural network, the encoder removes the high-level local sequential semantics to joined with the word vector using the recurrent neural network and the attention mechanism creates the last content representation. Then totally initialized, the decoder used for the association layer randomly connecting the two labels. The algorithm has been produced a hamming loss of 78%, a micro-Recall of 85.1%, and a Micro-F1 score of 87.3%. Ray, P. et al. [27] proposed a novel deep convolutional neural network model is used for labeling every sentence. Deep learning approaches combined with previous rule-based techniques for improving the feature extraction based on sentiment scoring technique. The proposed method was performed with an overall accuracy of 87%. Chen, C. et al. [28] projected a gated sentiment relationship recurrent neural network model to get the sentiment information from the effects of text and their modifier. The inputs are updated multiplicatively by the existing encoded sentiment modifier proposed the sentiment polarity. The two subsets are dividing the sentiment relation model was performed the overall accuracy of 94.6% and 95.2%. Guo, X. et al. [4] designed have been analyzed three levels of granularity: text, sentence, and aspect level. An advanced att-RCNN model that joins the recurrent neural network and convolution neural network. Both models are extracting the features for word-level and sentence-level, improving the significant terms and functionality promoting the performance of the proposed model. The Att-RCNN model was built on the F1 Score of 86.6%. Umer, M. et al. [23] projected the dimensionality reduction techniques to classified the feature vectors. The principal component analysis and chi-square are taken care of to nonlinear characteristics offering more contextual features to detect fake news. The proposed CNN-LSTM with PCA performed an accuracy of 97.8%, precision of 97.4%, Recall of 98.2%, and f-score of 97.8%. Cekik, R. et al. [24] a novel filter feature selection tool, uses the rough set for a territorial distinction according to the term of a value set to classify reports that have a class. By multiplying a coefficient called documents that may belong to a class are penalized. Also, using a rough set, the impact of sparsity in the term vector space is determined. The current filter attribute selection techniques such as Gini index, data gain, feature selector distinguishing, max-min ratio methods of normalized difference calculation. The proposed model perform was six different datasets are F1 score 80.20%, 72.91%, 80.25%, 80.17%, 79.84%, and 81.84%.

III Comparative Study

In this section comparative study on deep learning models about datasets, merits, and demerits, to discuss below for the sentiment analytics process.

TABLE I. Comparative Study with Convolutional Neural Network Model and Long Short Term Memory

AUTHOR/ YEAR	**MODEL**	**DATASET**	**MERIT**	**DEMERIT**
Kleenankandy, J. et al.[10] / (2021)	CNN LSTM	Sick SST	Sentence representation	Sentence Dependency Parse tree
Basiri, M.E., et al.[11]/ (2021)	CNN LSTM	Twitter Dataset	Decrease the dimensionality	Rating prediction
Umer, M. et al. [23] / (2020)	CNN LSTM	FNC	Dimensionality reduction of the feature vector Reduce overfitting	Noise data Inconsistency
Senthil kumar, N.K., et al.[14] / (2020)	CNN LSTM	Sentence Polarity	Part-of-speech (PoS) tagging	Sarcasm class Sentiment classification
Jang, B., et al. [20] / (2020)	CNN LSTM	IMDB	Text classification Sentence classification	LongTerm dependency.
Shuang, K. et al .[22] / (2020)	CNN LSTM	English dataset	The fuses context and task specific information	Polysemous unaware Task – unaware

Fig.1 Overall Accuracy of CNN-LSTM

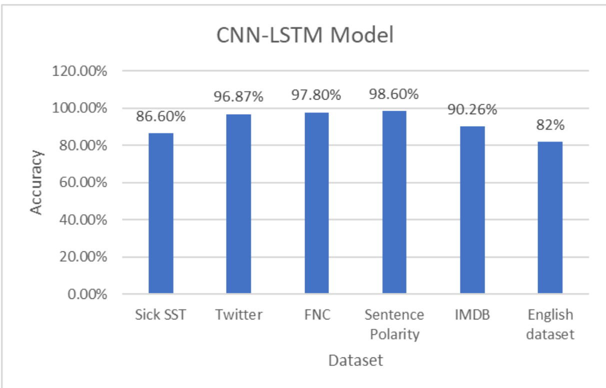

TABLE II. Comparative Study with Convolutional Neural Network Model

AUTHOR/ YEAR	MODEL	DATASET	MERIT	DEMERIT
Huang, M. et al .[12] / (2020)	CNN	Senti-Strength	Sentiment Strength	Context-dependent sentiment lexicon
Sharma, A.K., et al .[15] / (2020)	CNN	Sentence Polarity	Word embedding -text classification	Long sentence categorization
Zhou, M. et al. [17] / (2020)	CNN	Hotel Data	Single-channel representation	Non-static Word Vector
Samat, N.A. et al.[18] / (2020)	CNN	IMDB	Pooling Operation	Overfitting on the training set
Ray, P. et al.[27] / (2019))	CNN	SemEval	Sentence Score	Word Embedding

Fig.2 Overall accuracy of CNN

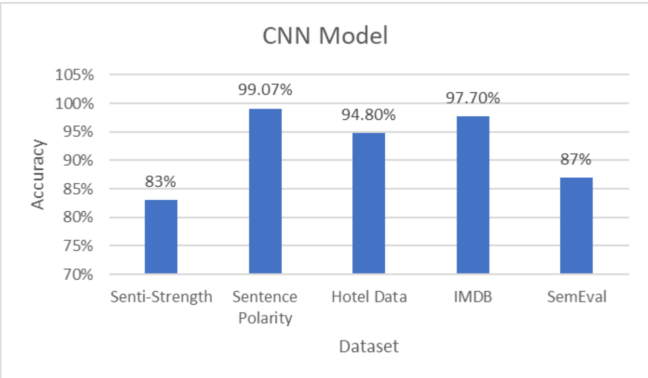

TABLE III. Comparative Study with Convolutional Neural Network and Recurrent Neural Network Model

AUTHOR/ YEAR	MODEL	DATASET	MERIT	DEMERIT
Abid, F. et al. [16] / (2020)	CNN RNN	IMDB	Word weighted Mechanism	Overfitting
Dang, N.C. et al.[2] / (2020)	CNN RNN	Sentiment 140	CPU runtime. Reduce the computational cost	Automatic Sentiment Analytics
Guo, X. et al .[4] / (2019)	CNN RNN	KBP37	Attention Mechanisms	Sentiment polarity Noise data based on SDP-word
Chen,C. et al. [28] / (2019)	RNN	SST, DM, CS	Sentiment modifier sentiment relations	Sentiment context Content word and Shifting negation

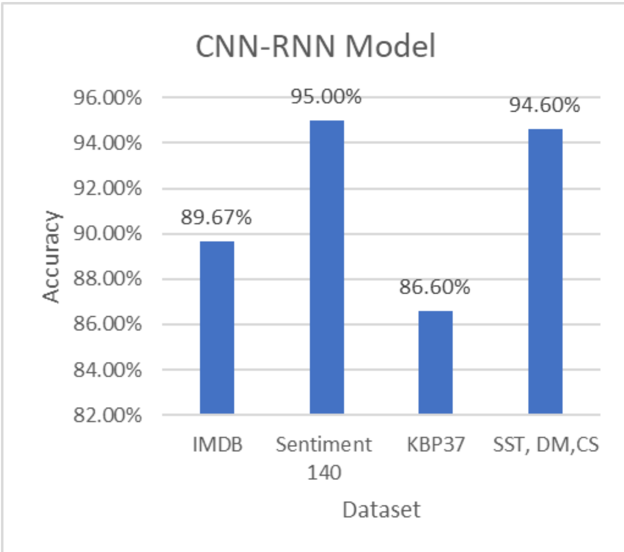

Fig.3 Overall accuracy of CNN-RNN

IV. Result and Discussion

In this survey, it is explored the applications of natural language processing and deep learning techniques. It is inferred that most of the research work is carried out in a feature extraction phase. The enormous literature survey related to large datasets, overfitting, number of classifiers, pooling, and dimensionality reduction were reviewed and compared with some viewpoints, such as datasets, methodology, merits, demerits, and overall accuracy of the existing models. With the help of comparative study, the core research is identified along with the related techniques. The datasets are chosen as unstructured datasets that have an impact on social media sectors. Hence, by selecting the social media dataset the research may move towards the problem in social media related identified the streaming news. Based on the comparative study, a research gap on feature extraction is noted for sentiment analytics.

V. Conclusion

The major revelation of the survey is based on real time problem available in social meadia which has more impacts on sentiment analytics. It is the kind of streamed text research that systematically determines the successful state of subjective knowledge at the sentence level by extracting quantifies on polarity detection. This paper addresses the problem of detecting sentiment polarity, pooling, Sentiment ranking, and sentiment extraction. Deep learning methods have been established to address the major obstacles of sentiment analytics. The merits and demerits are thoroughly identified. The limitations of the research work are aspect and word level. Hence, with the help of a study on sentiment analytics using deep learning, it is motivated to proceed with the research work is noise reduction, overfitting, dimensionality reduction and pooling operation using deep learning models are LSTM, CNN, and RNN for sentiment analytics.

References

[1] Chakraborty, K., Bhattacharyya, S., Senior Member, IEEE, and Bag, R.: A Survey of Sentiment Analytics from Social Media Data. Compute. Social. Syst. 1-15 IEEE Trans (2020)

[2] Dang, N.C., María, N., Moreno-Garcia and Prieta, F.D.: Sentiment analytics Based on Deep Learning: A Comparative Study. *Electronics*. 483. 9, 2-29 MDPI (2020)

[3] Mohey, D.E1, Hussein, D. M.: A survey on sentiment analytics challenges. 30. 4, 330-338 Elsevier (2018).

[4] Guo, X., Zhang, H., Yang, H., Xu, L., and Zhiwen.: A single Attention-based combination of CNN and RNN for Relation Classification. Proc. Comput. Science. 7.131, 12467-12475 IEEE Access (2019).

[5] Habimana1, O., Li, L., Li, R., Gu X., & Yu, G.: Sentiment analytics using deep learning approaches: an overview," Info. sci. 63, Springer (2020)

[6] Daniel W. O., Julian R. Medina, and Jugal K. Kalita.: A Survey of the Usages of Deep Learning for Natural Language Processing. Neural. Netws. Learning. Syst., IEEE. Trans (2019)

[7] Wafaa S. EI-Kassas, Cherif R. Salama, Ahmed A. Rafea, Hoda K. Mohamed, " Automatic text Summarization: A comprehensive survey," Expert. Syst. Applns. 165, 113679 Elsevier (2021)

[8] Liu, R., Shi, Y., Ji, C., and Jia, M.: A Survey of Sentiment Analytics Based on Transfer Learning. IEEE Access (2019).

[9] Liu, L. Z. W. B.: Deep learning for sentiment analytics: A survey. Springer (2020)

[10] Kleenankandy, J., Nazeer, K.A.: An enhanced Tree-LSTM architecture for sentence semantic modelling using typed dependencies. Info. Pro. Management 57(6), 102362 ACM(2020)

[11] Basiri, M, E., Nemati, S., Adar, M., Cambria, E.: ABCDM: An Attention-based Bidirectional CNN-RNN Deep Model for sentiment analytics. Future Gen. Compute. Syst. 279-294 Elsevier (2021)

[12] Huang, M., Xie, H., Jingrong, Y.R., Fen, Wang, F.L.: Sentiment strength detection with a context-dependent lexicon-based convolution neural network. Info. Science. 520,389-399 Elsevier(2020)

[13] Shuang, K., Yanga, Q., Loo, J., Li, R., Gua, M., "Feature distillation network for aspect-based sentiment analytics," Info. Fusion, vol. 61, pp: 13-23, Elsevier, 2020.

[14] Senthil Kumar, N, K., Malarvizhi, N.: "Bi-directional LSTM-CNN Combined method for sentiment analytics in Part of Speech Tagging (POS)," Int. Journal. Of. Speech. Tech., Springer, 2020.

[15] Sharma, A.K., Chaurasia, S., Srivastava, D. K.: Sentiment Short Sentences Classification by Using CNN Deep Learning Model with Fine Tuned Word2Vec. Proc. Compute. Science. 167, 1139-1147 Elsevier (2020)

[16] Abid, F., Li, C., Alam, M.: Multi-source social media data sentiment analytics using bidirectional recurrent convolution neural networks, Compute. Communes. 157, 102-115 Elsevier (2020)

[17] Zhou, M., Liu, D., Zheng, Y., Zhu, Q., Guo, P.: A text sentiment classification model using double word embedding methods, Multmed. Tools. Appln. Springer (2020)

[18] Samat, N.A., Mohd, N.S., Ali, H.: The Comparison of pooling functions in a conventional neural network for sentiment analytics task. Research Gate (2020).

[19] Naderalvoijoud, B., Sezer, E.A.: Sentiment Aware Word Embedding using Refinement and Senti-Contextualized Learning Approach. Neuro. Comput, 405, 149-160 Elsevier (2020).

[20] Jang, B., Kim, M., Harerimana, G., Sang-ug K., and Kim, J.W.: Bi-LSTM Model to Increase Accuracy in Text Classification: Combining Word2Vec CNN and Attention Mechanism, Appl. Sci, 10. 5841. 1-14 MDPI (2020).

[21] Xu, J., Cai, Y., Wu, X., Lei, X., Huang, Q., Ho-Fung L, Li, Q.: Incorporating Context-Relevant Concepts into Convolution Neural Networks for short Text Classification," Neuro Comput., 386, 42-53 Elsevier (2020)

[22] Shuang, K., Zhang, Z., Loo, J., Sen Su.: Convolution-deconvolution word embedding: An end-to-end multi-prototype fusion embedding method for natural language processing," Info.Fusion. 53. 112-122, Elsevier (2020)

[23] Umer, M., Imtiaz, Z., Ulla, S., Mehmood, A., Choi G.S, and Byung-Won On.: "Fake News Stance Detection Using Deep Learning Architecture (CNN-LSTM). 8, 156695 – 156706 IEEE Access (2020)

[24] Cekik, R., Uysal, A.K.: A novel filter feature selection method using a rough set for short text data," Expert. Syst. Applns. 160, 113691 Elsevier (2020)

[25] Hassan, A., Mahmood, A, A.: Convolutional Recurrent Deep Learning Model for Sentence Classification," MultiDisci. Rapid. View. 6, 13949 - 13957 IEEE Access, (2018)

[26] Liao, W., Wang, Y., Yin, Y., Zhang, X., Pan M.: Improved sequence generation model for multi-label classification via CNN and initialized fully connection. Neuro. Comput. Elsevier (2019)

[27] Ray, P., Chakrabarti, A.: A Mixed approach of Deep Learning method and Rule-Based method to improve Aspect Level Sentiment analytics. Appl. Compute. Info. Elsevier (2019)

[28] Chen, C., Zhuo, R., Ren, J.: Gated recurrent neural network with sentiment relations for sentiment classification. Info. Science. 502, 268-278 Elsevier(2020).