Sentiment Analysis On Text Using Neural Network

[1]Joshua Fernandes,[2]Eldridge Fernandes,[3]Ronak Sakhiya,[4]Sourabh Mayekar,[5]Prathibha Sudhakaran

[1]Student,[2]Student,[3]Student,[4]Student,[5]Assistant Professor
[1] Department of Electronics and Telecommunication,
[1] Xavier Institute Of Engineering, Mahim, Mumbai, India.

Abstract : Sentiment Analysis is the analysis of the sentiment of the sentence based on the type of the word used. In this digital era, there are huge datasets of reviews available from customers across the globe. It is tedious to analyze the large datasets of reviews manually one by one. In this paper, we proposed to use a bunch of models for the analysis of sentiments. The models we employed are Support Vector Machine (SVM), Logistic Regression (LR), Multilevel Perceptron (MLP), Convolutional Neural Networks (CNN) and Long Short Term Memory (LSTM) Network. We used the standard IMDB datasets for all these five models. IMDB dataset is one of the most reputable dataset and its well-labeled data makes it easy to carry out operations which you intend to quiet feasibly. These datasets consist of reviews from different movies. Using this dataset, we tried to compare the accuracy obtained by each of the models. As per our results obtained we can certainly state that deep learning models are able to outperform the other models as they are able to analyze a deeper relationship within the sentence which drastically improves its classification with every epoch.

***IndexTerms* - Sentiments analysis, IMDB dataset, SVM, LR, MLP, CNN, and LSTM.**

I. Introduction

The research on sentiment analysis has been going on through ages and the main reason for that is that data is readily available in the form of reviews, feedback comments, etc. But this information can be harnessed by a machine using a deep learning neural network. A deep learning code combined with a neural network makes the system more dynamic and adaptable. This helps the neural network in analyzing various convoluted patterns and also classifying it in the presence of noise. A neural network was designed based on the observation of complexity present in the neurons of biological organisms. This neuron mostly consisted of 3 parts, i.e. the dendrites, stoma, and terminal. There are various types of neural networks available and based on their structure and adaptability they can be used for a plethora of activities. For example, a CNN is mostly used for pattern detection and classification in an image, whereas an RNN is used for text due to its exhaustive iterations and high adaptability in the presence of noise. Deep learning is very useful in unsupervised and supervised learning as many researchers are handling sentiment analysis by using deep learning. It consists of numerous effective and popular models as these models are used to solve it efficiently providing a variable accuracy of the working model.

Multilayer Perceptron is a feed-forward neural network which is able to map the input to the output to a process of weight update A Convolution Neural network takes in an input layer and then filters it to classify the value between 0 to 1. Long Short Term Memory is an upgraded version of RNN to overcome the problem vanish descend. A Support Vector Machine is a classifier which classifies the sentence using a separating hyperplane in between the data. Regression is a process of predictive analysis between the binary variables.

Thus by implementing sentiment analysis of the IMDB dataset using Deep Learning (MLP, CNN, LSTM), we received better accuracy in comparison with sentiment analysis using Machine Learning techniques (SVM, LR).

II. Proposed Methodology

A. SENTIMENT ANALYSIS

Sentiment analysis refers to the analysis of sentiments, opinions, and subjective text [1]. Sentiment analysis of public views using different tweets and reviews. It is a great tool for analyzing and predicting the many significant events such as box office performance of movies and general elections [2]. Public reviews are used to evaluate precisely a certain parameter, i.e., person, product or location and might be found on different websites like Amazon and Yelp. The opinions can then be categorized according to their polarity, i.e., into negative, positive or neutral. The purpose of sentiment analysis is to quickly and accurately determine the direction of user reviews [3]. The demand for sentiment analysis is raised due to the increased requirement of analyzing the hidden information which comes from social media in the form of unstructured data [4].

1) Features of Sentiment Analysis:

Sentiments contain a variety of featured values like tri-grams and bi-grams by means of polarities and combinations. So sentiments are being assessed both as negative and positive aspects through the numerous support vector machines, by using training algorithms. The neural networks employed for analyzing sentiment through the computation of its relationship with the help of labels. The data is then extracted at the context level. The conditional dependencies at several edges and nodes are calculated using Bayesian networks. The network itself learns and optimizes its learning and data accuracy to improve efficiency which can be attained on the social media platform. Data tokenization refers to produce negative and positive aspects of data. In order to attain a better accuracy in sentiment analysis for data acquired through social media, there is a need for the techniques to decrease the errors obtained. [5]

2) Sentiment Analysis as a multidisciplinary field:

The sentiment analysis is a multidisciplinary field because it includes various fields such as computational linguistics, information retrieval, semantics, natural language processing [6]. The classification for the approaches of sentiment analysis can

a) be done in three extraction levels feature or aspect level;
b) document level;
c) sentence level [5].

3) Techniques for Sentiment Analysis:

Sentiment analysis can be achieved using the two types of techniques, i.e., lexicon-based approach and machine learning based techniques [5].

a) Machine learning based techniques: These techniques are implemented by extracting the sentences and converting them into vectors. The features consist of words used in the daily speech from a dataset called bag-of-words. Machine learning contains three flavors at the sentence, i.e., Naive Bayes, Support Vector Machine (SVM) and Maximum Entropy.

b) Lexicon based or corpus-based techniques:
These techniques are based on decision trees such as K-Nearest Neighbors (k-NN), Conditional Random Field (CRF), Hidden Markov Model (HMM), Single Dimensional Classification (SDC) and Sequential Minimal Optimization (SMO), related to methodologies of sentiment classification. Machine learning approach has three categories:
I. supervised;
II. semi-supervised;
III. unsupervised.

B. DEEP LEARNING

Deep Learning was first proposed by G.E. Hinton in 2006. This is basically a part of a machine learning process which is able to get good accuracy with an increase in the dataset refers to Deep Neural Network [7]. The neural network is influenced by the human brain and it contains several neurons that make an impressive network. Deep learning networks are used to train both supervised and unsupervised categories [8]. Deep learning techniques include many models such as CNN (Convolutional Neural Networks), RNN (Recurrent Neural Networks), DBN (Deep Belief Networks) and many more. Neural networks are very beneficial in text classification and estimation, vector representation, word representation, and sentence modeling [9].

Applications of Deep Learning: Deep architecture consists of numerous levels of non-linear operations. The process of modeling the tasks of completed artificial intelligence problems makes the expectations that deep architecture will act well in semi-supervised learning such as Deep belief network (DBN) which attains great accuracy with respect to Natural language processing [10]. The Deep Learning techniques are equipped with improved software engineering, enhanced learning procedures and accessibility of computing power and training data [11]. It is inspired by neuroscience and has a splendid impact on a range of applications like speech recognition, NLP (Natural Language Processing) and computer vision. One of the basic challenges of deep learning is the number of layers and the number of hidden variables needed for each layer [12].

III. Block Diagram

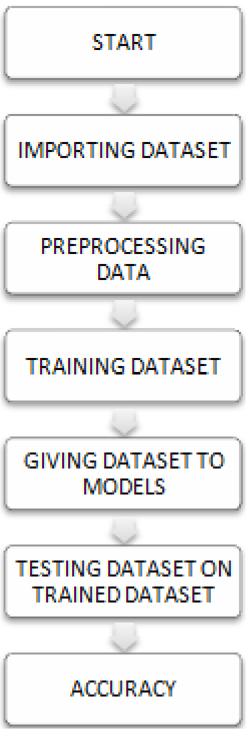

Fig. 1

IV. CLASSIFICATION OF ALGORITHM

A. Support Vector Machine(SVM)

Support-vector machines (SVMs) follows a supervised learning technique which is mostly associated with learning algorithms that analyze data used for classification and regression analysis mostly. Given a set of training examples with each method as belonging to one or the other of a category of an SVM training algorithm builds a model to assigns new examples to one prominent category or the other, making it a non-probabilistic binary linear classifier. An SVM model is an ideal representation of the examples as points in space, mapped so that the examples of the separate categories are firmly divided by a clear gap that is as wide as possible. New examples are then further mapped into that same space and predicted to belong to a category based on which side of the hyperplane they fall. SVMs can also perform non-linear classification using the kernel trick which is implicitly mapping their inputs into high-dimensional feature spaces.

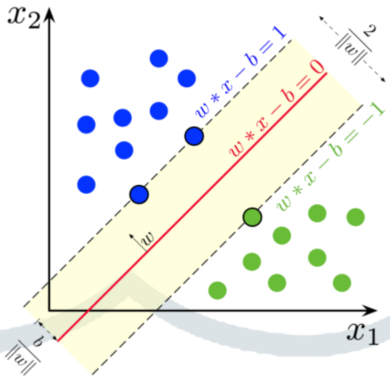

Fig 2

However, the issues faced by SVM are

- Requires full labeling of the given data.
- The SVM is only directly applicable for two-class tasks methods.
- Parameters of a model are difficult to interpret**.**
- It takes care of local minima but does not prevent overfitting.

B. Logistic Regression(LR)

The logistic model (or logit model) which is widely used a statistical model that in its basic form uses a logistic function to model which is a binary dependent variable. However, it can be modified to many more complex extensions. In regression analysis basically, the logistic regression (or logit regression) is the main estimating parameters of a logistic model (a form of binomial regression). A binary logistic model has a dependent variable with has two possible values which are basically a true or false, win/lose, these are represented by an indicator variable, where the two values are labeled either '0' or '1' depending on the data. In the logistic model, the log-odds for the value labeled "1" are a linear combination of one or more independent variables. Then the independent variables can each be a binary variable (two classes, coded by an indicator variable) or a continuous variable (or any real value). The labeling function converts log-odds to probability by updating the probability of the data labeled "1" to vary between 0 and 1. The unit of measurement for the log-odds scale is also known as a logit, from the logistic unit, hence the alternative names which are given to it. Besides, the analogous models with solemnly different sigmoid function instead of the logistic function are being used such as the probit model; the defining characteristic of the logistic model is that increasing one of the independent variables used multiplicatively scales the odds of the given outcome at a definite constant rate, with each dependent variable having its own parameter basically for a binary independent variable this generalizes the odds ratio.

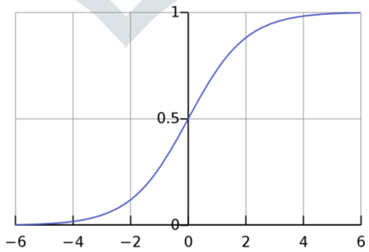

Fig. 3

C. Multilayer Perceptron(MLP)

A multilayer perceptron (MLP) is a class of feedforward artificial neural network used in deep learning. It foremost consists of at least 3 layers of nodes: an input layer, a hidden layer and an output layer in it. Except for the input nodes, each node is a neuron that basically uses a nonlinear activation function. MLP utilizes a supervised learning technique which is also called as backpropagation for training. Its multiple layers and non-linear activation distinguishes the MLP from a linear perceptron neural network. It can distinguish data that is not linearly separable using this method. The term "multilayer perceptron" doesn't refer to a single perceptron that has multiple layers. It instead contains many perceptrons that are organized into layers of networks. An alternative is "multilayer perceptron network". MLP perceptrons are not basically perceptrons in the strictest possible sense as we know that true perceptrons are formally a special case of artificial neurons that are used as threshold activation function mainly a Heaviside step function. MLP perceptrons can employ arbitrary activation functions to it. A true perceptron performs binary classification. The MLP neuron is free to either perform classification or regression which is depending upon its activation function to perform the task. The term "multilayer perceptron" defines the presence of multiple layers which can be composed of arbitrarily defined artificial neurons, and not specifically perceptrons. This interpretation avoids the loosening of the definition of "perceptron" to means of an artificial neuron in general.

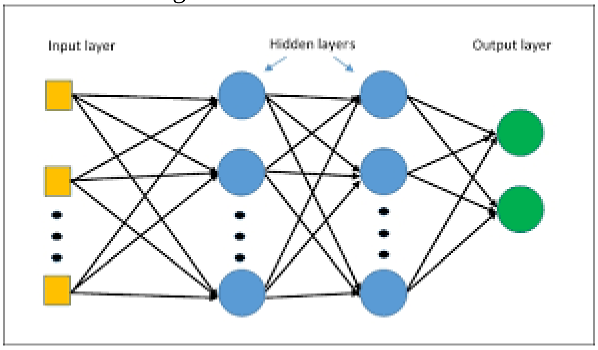

Fig.4

D. Conventional Neural Network(CNN)

A convolution neural network (CNN) is a deep neural network, employed for analyzing and classification of images mostly due to its good performance with respect to that domain. CNN's are regularized versions of multilayer perceptrons. They usually refer to fully connected networks, that is, each neuron in one layer is connected to all neurons present in the next layer. The "fully-connectedness" of these networks makes them suitable for over-fitted data. Typical ways of regularization include adding some form of magnitude measurement of weights to the loss function present in it. CNN takes a different approach toward regularization. They take advantage of the hierarchical pattern in data and later on assemble more complex patterns using smaller and simpler patterns in it. On the scale of connectedness and complexity, CNN is on the lower extremes on the scale. They are also known as a shift-invariant or space invariant artificial neural networks following their shared-weights architecture and translation invariance characteristics of its nature. These neural networks were inspired by biological processes in which the connectivity pattern between neurons resembles the organization of the animal variant visual cortexes present in them. Individual cortical neurons respond to stimuli only in a restricted region of the visual field which is also commonly known as the receptive field. These fields of different neurons partially overlap only to the extent that they cover the entire visual field which is being utilized. CNN's normally perform relatively lesser pre-processing as in comparison with other image classification algorithms used. This means that the network learns the filters which are used in traditional algorithms that were basically hand-engineered. This independence from prior knowledge and human effort in feature design is a major advantage which is mostly observed. They have various applications over the years in image and video recognition, recommender systems, image classification, medical image analysis, and natural language processing.

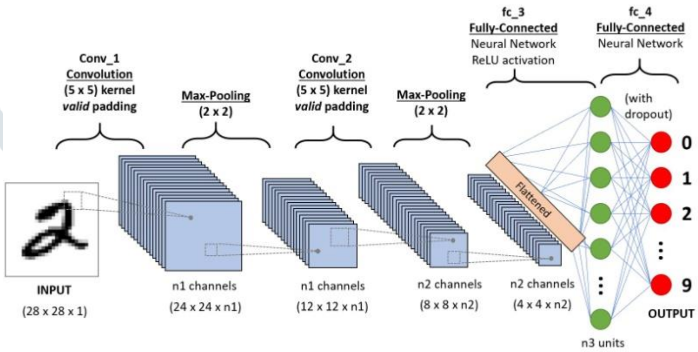

Fig.5

E. Long Short Term Memory(LSTM)

Long short-term memory (LSTM) is a modification made to the recurrent neural network (RNN) architecture which unlike standard feed-forward neural networks, it contains a feedback connection similar to recursive neural networks without the problem of vanish descent. It cannot only process single data points (such as images) but at the same time also entire sequences of data (such as speech or video) which is mostly the major requirement. For example, it is applicable to tasks such as unsegmented, connected handwriting recognition or speech recognition in everyday use. A common LSTM unit is composed of basically a cell, an input gate, an output gate and a forget gate used in it. The cell remembers values over arbitrary time intervals and then the three gates regulate the flow of information into the cell and out of the cell. LSTM networks are mostly used in classifying, processing and predicting time series data as there can be lags of unknown duration between important events in a time series of any frame. LSTMs were developed in areas to deal with the exploding and vanishing gradient problems that can be encountered when training traditional RNNs mostly. Relative insensitivity to gap length is a major advantage of LSTM over RNNs.

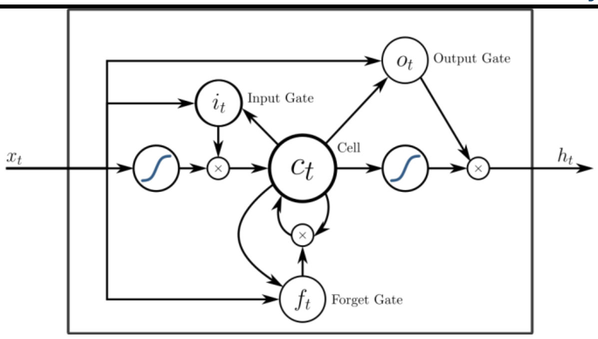

Fig.6

V. Result

Using IMDB dataset on the methods mentioned above the accuracy achieved is given below:-

1. SVM

```
Logistic Regression
Accuracy Score:  50.848000000000006  %
```

Achieved accuracy: 50.85%

2. LR

```
Support Vector Machine
Accuracy Score:  49.612  %
```

Achieved accuracy: 49.61%

3. MLP

```
- 25s - loss: 1.8745e-04 - acc: 1.0000 - val_loss: 0.6296 - val_acc: 0.8689
Accuracy: 86.89%
```

Achieved accuracy: 86.89%

4. CNN

```
- 51s - loss: 0.0015 - acc: 0.9999 - val_loss: 0.7149 - val_acc: 0.8682
Accuracy: 86.82%
```

Achieved accuracy: 86.82%

5. LSTM

```
Training Step: 6250  | total loss: 0.32710 | time: 113.487s
| Adam | epoch: 010 | loss: 0.32710 - acc: 0.8597 | val_loss: 0.38767 -
val_acc: 0.8280 -- iter: 20000/20000
```

Achieved accuracy: 82%

Achieved Accuracy In Tabular Format:

Neural Network	SVM	LR	MLP	CNN	LSTM
Accuracy	**50.85%**	**49.61%**	**86.89%**	**86.82%**	**82%**

VI. Conclusion

The aim of this paper is to provide results that deep neural networks are able to provide better accuracy and adaptability with the increase in the dataset. Support Vector Machine (SVM), Logistic Regression (LR) was able to deliver with a large dataset. Thus the accuracy obtained was only 50% which is not adequate. However, with the help of deep neural nets like MLP, CNN and LSTM we were able to get an accuracy of 80% and above. This is because deep neural networks are able to learn better with large training dataset and the accuracy also subsequently increases.

VII. Acknowledgment

I will take this opportunity on behalf of our project group to thank all the help provided by teachers and friends who were always there and without whom this project completion would not have been possible. Our respected guide, Prof. Pratibha Sudhakaran under whose guidance we were able to make smart choices towards our project. Ma'am was instrumental in not only solving our pettiest doubts patiently but also gave us the necessary push needed to focus on attaining the goals of the project. Also, Dr. Suprava Patnaik whose advice and recommendation turned out to be the concrete basis of our research. Furthermore, I would thank all the other teachers and my colleagues who're motivation and aid helped us get this far.

REFERENCES

[1] QuratTulAin, Mubashir Ali, AmnaRiaz, AmnaNoureen, Muhammad Kamran, Babar Hayat and A. Rehman, Sentiment analysis using deep neural network: A Review (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 6, 2017
[2] A. V. Yeole, P. V. Chavan, and M. C. Nikose, Opinion mining for emotions determination, ICIIECS 2015 - 2015 IEEE Int. Conf. Innov. Information,
Embed. Commun.Syst., 2015.
[3] B. Heredia, T. M. Khoshgoftaar, J. Prusa, and M. Crawford, CrossDomain Sentiment Analysis: An Empirical Investigation, 2016 IEEE 17th Int. Conf.
Inf. Reuse Integr., pp. 160165, 2016.
[4] F. Luo, C. Li, and Z. Cao, Affective-feature-based sentiment analysis using SVM classifier, 2016 IEEE 20th Int. Conf. Comput. Support. Coop. Work
Des., pp. 276281, 2016.
[5] M. Haenlein and A. M. Kaplan, An empirical analysis of attitudinal and behavioral reactions toward the abandonment of unprofitable customer
relationships, J. Relatsh.Mark., vol. 9, no. 4, pp. 200228, 2010.
[6] J. Singh, G. Singh, and R. Singh, A review of sentiment analysis techniques for opinionated web text, CSI Trans. ICT, 2016.
[7] E. Aydogan and M. A. Akcayol, A comprehensive survey for sentiment analysis tasks using machine learning techniques, 2016 Int. Symp. Innov.Intell.
Syst. Appl., pp. 17, 2016.
[8] M. Day and C. Lee, Deep Learning for Financial Sentiment Analysis on Finance News Providers, no. 1, pp. 11271134, 2016.
[9] P. Vateekul and T. Koomsubha, A Study of Sentiment Analysis Using Deep Learning Techniques on Thai Twitter Data, 2016.
[10] Y. Zhang, M. J. Er, N. Wang, M. Pratama, and R. Venkatesan, Sentiment Classification Using Comprehensive Attention Recurrent Models, pp.
15621569, 2016.
[11] S. Zhou, Q. Chen, and X. Wang, Active deep learning method for semi-supervised sentiment classification, Neurocomputing, vol. 120, pp. 536546,
2013.
[12] L. Deng, G. Hinton, and B. Kingsbury, New types of deep neural network learning for speech recognition and related applications: an overview, 2013
IEEE Int. Conf. Acoust. Speech Signal Process., pp. 85998603, 2013.
[13] S. Bengio, L. Deng, H. Larochelle, H. Lee, and R. Salakhutdinov, Guest Editors Introduction: Special Section on Learning Deep Architectures, IEEE
Trans. Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 17951797, 2013.
[14] L. Arnold, S. Rebecchi, S. Chevallier, and H. Paugam-Moisy, An Introduction to Deep Learning, Esann, no. April, p. 12, 2011.
[15] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, Deep learning for visual understanding: A review, Neurocomputing, vol. 187, pp. 2748,
2016.
[16] X. Ouyang, P. Zhou, C. H. Li, and L. Liu, Sentiment Analysis Using Convolutional Neural Network, Comput. Inf. Technol. Ubiquitous
Comput.Commun.Dependable, Auton.Secur.Comput.Pervasive Intell.Comput.(CIT/IUCC/DASC/PICOM), 2015 IEEE Int. Conf., pp. 23592364, 2015.
[17] R. Socher and C. Lin, Parsing natural scenes and natural language with recursive neural networks, Proc. , pp. 129136, 2011.
[18] T. Mikolov, K. Chen, G. Corrado, and J. Dean, Efficient Estimation of Word Representations in Vector Space, Arxiv, no. 9, pp. 112, 2013.
[19] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, A Convolutional Neural Network for Modelling Sentences, Acl, pp. 655665, 2014.
[20] Y. Kim, Convolutional Neural Networks for Sentence Classification, Proc. 2014 Conf. Empir. Methods Nat. Lang. Process.(EMNLP 2014), pp.
17461751, 2014.
[21] T. Mikolov, K. Chen, G. Corrado, and J. Dean, Distributed Representations of Words and Phrases and their Compositionality, Nips, pp. 19, 2013.
[22] Z. Wu, T. Virtanen, T. Kinnunen, E. S. Chng, and H. Li, Exemplar-based unit selection for voice conversion utilizing temporal Information,
Proc.Annu.Conf.Int.SpeechCommun.Assoc.INTERSPEECH,pp.30573061, 2013.
[23] K. S. Tai, R. Socher, and C. D. Manning, Improved semantic representations from tree-structured long short-term memory networks, Proc. ACL, pp.
15561566, 2015.
[24] A. A. Al Sallab, R. Baly, and H. Hajj, Deep Learning Models for Sentiment Analysis in Arabic, no. November. 2015.
[25] C. Z. Jiajun Zhang, Neural Networks in Machine Translation: An Overview, IEEE Intell. Syst., pp. 17241734, 2015.
[26] Aditya Timmaraju and Vikesh Khanna. Sentiment Analysis on Movie Reviews using Recursive and Recurrent Neural Network Architecture. Stanford
University, Stanford, CA – 94305, USA (2016).

[27] A. Severyn and A. Moschitti, Twitter Sentiment Analysis with Deep Convolutional Neural Networks, Proc. 38th Int. ACM SIGIR Conf. Res. Dev. Inf.
Retr. - SIGIR 15, pp. 959962, 2015.
[28] L. Yanmei and C. Yuda, Research on Chinese Micro-Blog Sentiment Analysis Based on Deep Learning, 2015 8th Int. Symp. Comput.Intell.Des., pp.
358361, 2015.
[29] Q. You, J. Luo, H. Jin, and J. Yang, Joint Visual-Textual Sentiment Analysis with Deep Neural Networks, Acm Mm, pp. 10711074, 2015.
[30] C. Li, B. Xu, G. Wu, S. He, G. Tian, and H. Hao, Recursive deep learning for sentiment analysis over social data, Proc. - 2014 IEEE/WIC/ACM Int.
Jt. Conf. Web Intell.Intell.Agent Technol. - Work.WI IAT 2014, vol. 2, pp. 13881429, 2014.
[31] R. Socher, A. Perelygin, and J. Wu, Recursive deep models for semantic compositionality over a sentiment treebank, Proc. , pp. 16311642, 2013.
[32] W. Li and H. Chen, Identifying top sellers in underground economy using deep learning-based sentiment analysis, Proc. - 2014 IEEE Jt. Intell.
Secure.Informatics Conf. JISIC 2014, pp. 6467, 2014.
[33] C. Baecchi, T. Uricchio, M. Bertini, and A. Del Bimbo, A multimodal feature learning approach for sentiment analysis of social network multimedia,
Multimed. Tools Appl., vol. 75, no. 5, pp. 25072525, 2016.
[34] H. Yanagimoto, M. Shimada, and A. Yoshimura, Document similarity estimation for sentiment analysis using neural network, 2013 IEEE/ACIS 12th
Int. Conf. Comput. Inf. Sci., pp. 105110, 2013.
[35] R. Silhavy, R. Senkerik, Z. K. Oplatkova, P. Silhavy, and Z. Prokopova, Artificial intelligence perspectives in intelligent systems: Proceedings of the
5th computer science on-line conference 2016 (CSOC2016), vol 1, Adv. Intell. Syst. Comput., vol. 464, pp. 249261, 2016.
[36] A. Hassan, M. R. Amin, A. Kalam, A. Azad, and N. Mohammed, Bangla Text (BRBT) using Deep Recurrent models.
[37] T. Chen, R. Xu, Y. He, Y. Xia, and X. Wang, Using a Sequence Model for Sentiment Analysis, no.August, pp. 3444, 2016.
[38] P. Ruangkanokmas, T. Achalakul, and K. Akkarajitsakul, Deep Belief Networks with Feature Selection for Sentiment Classification, Uksim.Info, pp.
16, 2016.
[39] G. Zhou, Z. Zeng, J. X. Huang, and T. He, Transfer Learning for Cross-Lingual Sentiment Classification with Weakly Shared Deep Neural Networks,
Proc. 39th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. - SIGIR 16, pp. 245254, 2016.
[40] R. Ghosh, K. Ravi, and V. Ravi, A novel deep learning architecture for sentiment classification, 3rd IEEE Int. Conf. Recent Adv. Inf. Technol., pp.
511516, 2016.
[41] R. Goebel and W. Wahlster, Integrated Uncertainty in Knowledge Modelling and Decision Making, Proc. Int.
Symp.Integr.Uncertain.Knowl.Model.Decis.Mak.(IUKM 2011), vol. 1, pp. 362373, 2011.
[42] A. Graves, N. Jaitly, and A. R. Mohamed, Hybrid speech recognition with Deep Bidirectional LSTM, 2013 IEEE Work.Autom Speech
Recognit.Understanding, ASRU 2013 - Proc., pp. 273278, 2013.
[43] C. N. dos Santos and M. Gatti, Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts, Coling-2014, pp. 6978, 2014.
[44] K. Ravi and V. Ravi, Sentiment classification of Hinglish text, 2016 3rd Int. Conf. Recent Adv. Inf. Technol. RAIT 2016, pp. 641645, 2016.