

# SENTIMENT ANALYSIS TECHNIQUES- A REVIEW

<sup>1</sup>Manjinder Kaur, <sup>2</sup>Nirvair Neeru, <sup>3</sup>Navjot Kaur

<sup>1</sup>M.Tech Research Scholar, <sup>2</sup>Assistant Professor, <sup>3</sup>Assistant Professor

Department of Computer Engineering,  
Punjabi University, Patiala.

**Abstract :** Sentiment Analysis is one of the fastest growing service and application in the mining industry. It is used in integration with machine learning algorithms. More and more industries and organizations have started to use this application in order to achieve better productivity and consumer experience. Different Deep Learning algorithms used for sentiment analysis like Convolutional Neural Network, Recursive Neural Networks and Deep Neural Networks are reviewed for sentiment analysis. There are different levels in sentiment analysis in which it is divided, which can be sentence based, document based or entity based. Sentiment Analysis has different approaches like Lexical Affinity, Statistical Method and Keyword spotting. Sentiment Analysis is divided into different techniques and they may have various sub-techniques. This paper explains the sentiment analysis and various algorithms, approaches related with this rapidly growing application.

**Index Terms -** Lexicon, Machine Learning, Deep Learning, RNN, CNN, DNN, Semantic Analysis, Keyword Spotting.

## I. INTRODUCTION

Technology is driving new advancements in different industries constantly these days. Technologies like Cloud Computing, Machine Learning, Deep Learning [5], Software Defined Networks etc have changed the industries and their work processes in recent years along with the enhanced business productivity that they bring. Almost all the industries are going for digitization of their work and to understand the customer requirements or what their customers think about the products or services they deliver. Computers can be programmed in order to learn and recognize things that helps them to achieve some specific operations. One problem that always arises is machine's inability to think and work like humans. But, neural networks [2,9] and its various advancements along with deep learning algorithms [9, 11] have come out largely and making advances in various industries while making machines learn, understand and implement different actions on its own.

Text Semantics are used in order to understand the work and context behind text and then it is used to solve different problems. With text semantics, meaning of text is found. Words and sentence integration make a lexical and contextual relation between various words that has a different relationship and hierarchy. Semantics acts as a centralized part to find relationships by getting any meaningful data from them. Semantics relates with context or meaning. Text Structure does not hold any importance here. There can be instances when this word arrangement gets the data meaning decided by entity based sentiment analysis algorithms and techniques like "Tesla can be the name of scientist" from "Tesla is a automobile company"

## II. SENTIMENT ANALYSIS DEEP LEARNING ALGORITHMS

Sentiment classification can be done using various algorithms and researchers have started to combine different machine and deep learning algorithms to get the better and more accurate results. In this part, work by various researchers in sentiment analysis using algorithms like CNN, RNN and DBF are explained.

**2.1. Convolutional Neural Networks (CNN)** – This algorithm brings a standard architecture to link the different sized sentences to fixed size scattered vectors. J onatas Wehrmann, Willian Becker, Henry E. L. Cagnini, and Rodrigo C. Barros[10] explains a CNN framework for visual sentiment to get visual sentiment analysis that helps in getting the sentiments related with visual content. Authors performed their work with python running on linux. Transfer learning method was used along with hyper parameters, while Google LeNet was used to gather weights. When CNN algorithm was applied, performance was improved vastly, so CNN algorithm was proposed over Google LeNet.A. Severyn and A. Moschitti [13] proposed a CNN based system for sentiment analysis. Author's main focus was to get weight of different CNN parameters to train the model in accurate manner. No new feature was used and word embedding is performed using a neural language and it has used a wide group of tweets. Activations were used along with sentence matrix pooling and various convolutional layers. Regularization was improved using a dropout method. Six tests were used and author's model comes first in accuracy.

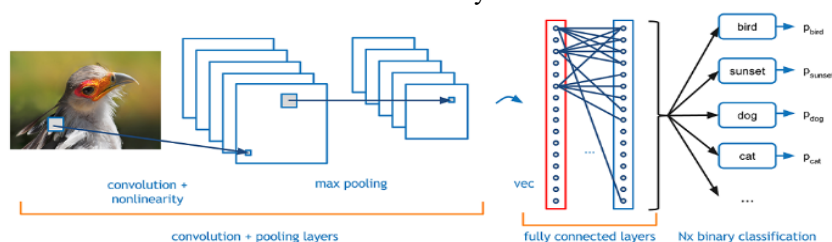


Figure 1.1 – CNN Layers

Convolutional Layers in CNN applies learned filters to input images which helps in creating feature maps summarizing the presence of those features in input. Convolutional Layers are effective and by stacking them in deep models helps in layers which are close to the input in learning low-level features like lines and high level features like some particular object. They have a high accuracy in image recognition problems. Limitation of using CNN is that even small movements in the position feature in input image results in a different feature map. This limitation happens while one is rotating, shifting, or making some other minor changes in the input

image. Pooling Another limitation of using CNN is that it involves high computation costs and good GPUs. They also need to have large amount of training data. Layer is added to the convolutional layer after applying non-linearity to the feature maps which are outputted by convolutional layer. Pooling Layer works on every feature map in separate manner in order to generate same amount of pooled feature maps.

**2.2 Recursive Neural Networks (RNN)**

RNN uses supervised learning method with a tree type structure which is settled before training and with every node having different matrix. Reassembling of input is not needed in Recursive Neural Networks. W. Li and H. Chen[14] explains a generalized framework which is scalable also and does identifies the malware. RNN comes under deep learning and is used for sentiment analysis. Quality of Service (QoS) or Quality of Product (QoP) checks are made using thread classification and sampling of snowballs and analyzes on the basis of customer feedback. Data is fetched from a Russian forum website. A tree based sentiment bank was used along with recursive tensor based network with online reviews in the Russian forum website.

To get the validation and to know the amount of accuracy that author’s model has achieved, author’s work has been compared with Naïve Bayes, KNN and SVM Models. Results shows that deeplearning algorithms provide much better outcomes than using some shallow classifiers. Figure 1.2 below shows RNN as a tree like network structure with zero time entity in input sequence and that input have to be processed in a tree type structure.

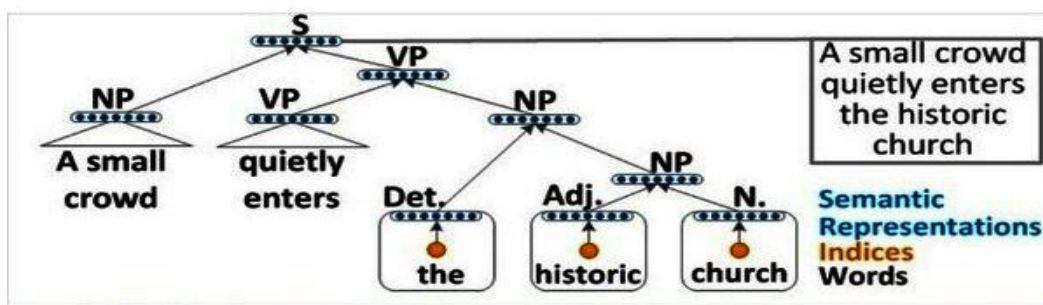


Figure 1.2 – Recursive Neural Network

Recursive Neural Networks is one of the type of Recurrent Neural Networks which extends to the tree structure. Recursive Neural Network has fixed number of branches. In binary tree, hidden vector of current nodes are computed from hidden state vectors of child nodes. This computation is done in sequential manner from leaf towards root node. Advantages include better hierarchical representations. RNNs has a disadvantage that it does not keep track of any long term dependency. Another drawback is that it cannot be stacked into high number of deep models.

**2.3 Deep Neural Networks**

H.Yanagimoto, M. Shimada and A. Yoshimura[15] explained DNN architecture to gather the information on similarity of the documents. This architecture was trained using various market related news which helps in production of vectors for articles. Data is taken using T&C News and a cosine based similarity gets calculated from labeled articles and by using that document polarity was considered. Better results were achieved in predicting article based similarity using the author’s technique. Figure 1.3 displays deep neural networks using intermediate layers which were build using various abstraction layers like in Boolean circuits. Hidden layers are present between input and output layer.

**Deep neural network**

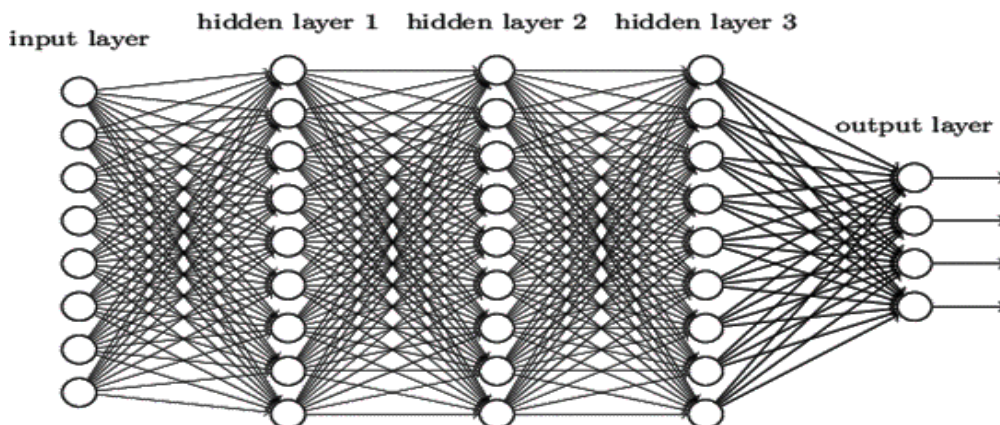


Figure 1.3 – Deep Neural Networks

DNN is an efficient tool used for Deep Learning. DNN processes input data with every processing gathers more abstract information. In the DNN, there are inputs using which the hidden layer processes simple characteristics, for example edges. In the second level, complex parts are formed and at the output layer analysis is completed. Advantage of using DNN is high performance in various domains. Another benefit is that it adapts well to new problems. Disadvantage is that it needs extreme level computation and it is not easy to grasp what is learned.

### III. SENTIMENT ANALYSIS

Sentiment Analysis is a popular application comes under text analytics where it is integrated with web applications, mobile apps, etc. Sentiments are analyzed using various resources which can be from tweets, Movie Reviews [16] from International Movie Database(IMDB) etc. Text body is analyzed in order to find and learn various opinions with different factors like mood or modality. Sentiment Analysis mainly works with subjective context. This is mainly as objective context does not express any emotions or mood. Sentiment Analysis is mainly used in Social Network [1] sites like Facebook and Twitter [1], ecommerce platforms like Flipkart or Amazon, movie review websites like IMDB in order to find the user opinions. Text based data mainly is of two types, one is objective type, also known as the one which is based on facts, other one is subjective, also known as opinions based. Somewebsites like Google Rewards and IMDB etc. have opinions based data that shows emotions, judgements of people. Opinion Mining can also be expressed using NLP techniques, machine learning and deep learning algorithms to get the data like mood or modality. This data is then used to find the polarity of the text based document. Polarity means to get the sentiments in textual data.

**Levels** –Sentiment Analysis has various levels explained below:

- **Document** –This level gets the sentiments from the document. Document used in mainly based on single topic. Various types of texts which are based on comparative learning is not used in this level.
- **Sentence** – This level finds the sentiment in the sentences. This level expresses the sentiment in a sentence like if the sentence is positive or negative etc. If the sentence does not finds any opinion, then it is of neutral view. This level related with subjectivity classification. In this level, a sentence shows the good or bad polarity.
- **Entity Based** – This level involves detailed analysis. The main goal at this level is to fetch the aspect of text. For example, in a car review by customer, “Mileage is good, but pickup is poor.” Aspects here are Car Mileage and pickup speed. Two different tasks are integrated at this level, one is finding the aspect of the text and other one is to classify the aspect.

**Sentiment Analysis Approaches** – There are various methods which can be used for sentiment analysis.

Sentiment Analysis is used to compute sentiments from the textual data on different levels like sentence, document, and entity based levels. In large number of cases, it is computed at the document level. Polarity Analysis[8,10] gives scores to access the different levels of positivity and negativity with the emotions expressed under the document. Sentiment Analysis is divided in to 3 major types:

**Keyword Spotting** – Using this method, text can be classified in to many unambiguous words. Like documents and sentence, there are some specific words holding importance when used in sentiment analysis. This is one of the most intuitive method which is used in fetching emotions from text. Figure below shows the keyword spotting method:



Figure 3.1 – Keyword Spotting Method

- **Lexical Affinity** – This type brings an probabilistic similarity to somerandom words having particular emotion. This is the advanced version of keyword spotting. Emotion is analyzed on the related words and a probabilistic affinity is integrated to the particular emotion on a set of words than using predefined words. This method has some drawbacks like as it depends on corpus based text.
- **Statistical Method** – This method determines the target of intuitive keywords. Calculation of the word coexisted numbers is also done on the basis of large training corpus. Some of the challenges of this method is that it does not work well with new or unseen words.

### CONCLUSION

Sentiment Analysis is used by different organizations for their business promotion and growth. This is an application of machine learning and deep learning algorithms which are used in integration with the text mining to fetch meaningful information from unstructured data. The data gathered using sentiment analysis is very beneficial for customers and product based companies which uses this information to know better about the products and emotions of the people by analyzing some feedbacks of the product or analyzing comments of people on election etc. This is used in large extent in social network sites like facebook or twitter etc. to know about the people sentiments on some specific politician [17], celebrity etc. This application is also very popular in ecommerce industry as ecommerce giants like Amazon, Flipkart, Myntra etc. are using sentiment analysis to get people mood and emotions on various products. With the time, sentiment analysis is growing rapidly in different businesses. CNN, RNN and DNN are some of



the algorithms which are used to get the sentiment analysis with good accuracy. This paper provides a review on what sentiment analysis is and its different types, algorithms and approaches.

## REFERENCES

- [1] **Kiruthika M. ,SanjanaWoonaa and PriyankaGiri(2016)**,"Sentiment Analysis of Twitter Data",International Journal of Innovations in Engineering and Technology(IJIET).
- [2] **Ming Jiang, Liqiang Jin, Feiwei Qin, Min Zhang, Ziyang Li(2016)**," Network Public Comments Sentiment Analysis Based on Multilayer Convolutional Neural Network", IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData).
- [3] **MetinBilgin and IzzetFatihSenturk(2017)**," Sentiment Analysis on Twitter data with Semi-Supervised Doc2Vec", UBMK International conference on computer science and engineering.
- [4] **Shiyang Liao, Junbo Wang, Ruiyun Yu, Koichi Sato, Zixue Cheng(2016)**," CNN for situations understanding based on sentiment analysis of twitter data", International Conference on Advances in Information Technology, IAIT.
- [5] **Oscar Araque ,IgnacioCorcuera-Platas, J.Fernando Sánchez-Rada, CarlosA. Iglesias(2017)**," Enhancing deep learning sentiment analysis with ensemble techniques in social applications", . Retrieved from-<http://dx.doi.org/10.1016/j.eswa.2017.02.002>
- [6] **Vishal A. Kharde and S.S. Sonawane(2016)**,"Sentiment Analysis of Twitter Data: A Survey of Techniques", International Journal of Computer Applications (0975 – 8887) Volume 139 – No.11.
- [7] **ShufengXiong, HailianLv , Weiting Zhao, DonghongJi(2017)**," Towards Twitter sentiment classification by multi-level sentiment-enriched word embeddings", Elsevier.
- [8] **NehalMamgain, Ekta Mehta, Ankush Mittal and Gaurav Bhatt(2016)**," Sentiment Analysis of Top Colleges in India Using Twitter Data", International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT).
- [9] **AliakseiSeverynabdAliakseiSeveryn(2015)**," Twitter Sentiment Analysis with Deep Convolutional Neural Networks", ACM. ISBN 978-1-4503-3621-5/15/08.
- [10] **JonatasWehrmann, Willian Becker, Henry E. L. Cagnini, and Rodrigo C. Barros(2017)**," A Character-based Convolutional Neural Network for Language-Agnostic Twitter Sentiment Analysis",IEEE.
- [11] **AdyanMarendraRamadhani and Hong Soon Goo(2017)**," Twitter Sentiment Analysis using Deep Learning Methods", International Annual Engineering Seminar (InAES).
- [12] **Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. (2011)**. Learning Word Vectors for Sentiment Analysis. The 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011).
- [13] **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
- [14] **W. Li and H. Chen**, Identifying top sellers in underground economy using deep learning-based sentiment analysis, Proc. - 2014 IEEE Jt. Intell. Secur. Informatics Conf. JISIC 2014, pp. 6467, 2014.
- [15] **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.
- [16] **T. P. Sahu and S. Ahuja**, "Sentiment analysis of movie reviews: A study on feature selection & classification algorithms," *2016 International Conference on Microelectronics, Computing and Communications (MicroCom)*, Durgapur, 2016, pp. 1-6. doi: 10.1109/MicroCom.2016.7522583
- [17] **A. Lopardo and M. Brambilla**, "Analyzing and Predicting the US Midterm Elections on Twitter with Recurrent Neural Networks," *2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA, 2018, pp. 5389-5391. doi: 10.1109/BigData.2018.8622441
- [18] **Yang, X., Macdonald, C. & Ounis, I.** Inf Retrieval J (2018) 21: 183. <https://doi.org/10.1007/s10791-017-9319-5>