

A Review on Sentiment Analysis on Hindi Language using Neural Network

Nikita Kolambe¹, Yashashree Belkhede² and Nikhil Wagh³

¹Department of Computer Science and Engineering,
Government College of Engineering Aurangabad, Maharashtra, India
Email id: nikita.kolambe@gmail.com

²Department of Computer Science and Engineering,
Government College of Engineering Aurangabad, Maharashtra, India
Email id: yashashreebelkhede@gmail.com

³Department of Computer Science and Engineering,
Government College of Engineering Aurangabad, Maharashtra, India
Email id: nikwagh100@gmail.com

Abstract-On widely spread network of the social media platforms people find it easy to express their opinions, feelings and thoughts. With the change in technology and development of social platforms people started expressing their thoughts in English as well as in their mother tongue for e.g. Hindi, Marathi etc. users of these platforms share their thoughts, reviews on movies, video, songs and product. These thoughts can be positive, negative and neutral. These expressions acts as a pool of information from which knowledge can be gained. So to identify sentiment from such data becomes essential and motivates research in analyzing sentiment from regional language such as Hindi. So the idea is to use neural networks for analyzing sentiment from Hindi data. The model will be trained by using Deep belief network (DBN) for classification of Data in Hindi Language into positive and negative sentiments.

Keywords: Artificial intelligence, extraction, Sentiment Analysis, Social Platform, WordNet.

I. INTRODUCTION

Natural Language processing comes under Artificial intelligence, also known as NLP. It includes interactions between computer and human language. Read, understand and gain new insights into human languages etc. are objectives of NLP [14].

Sentiment Analysis is the sub field of NLP that deals with the extraction of sentiment from a source of data. It tries to understand the data in human language and categorize it into positive, negative and neutral sentiments [2]. This data in human language can be found across different platforms, on which people share their thoughts like blogs, forums, social sites. Also people gives comments on a movie, video, songs are source of data for Sentiment Analysis. So the task of it is to classify the data and gain knowledge.

Day by day the data over the social sites is increasing. We can say that large amount of sector in this data is acquired by Hindi data as Hindi language ranked as fourth most spoken language in the world[1].So we proposed a model to classify Hindi data.

For training of the model some documents are required, Hindi WordNet which consist of parts of speech for Hindi words. It is used for tagging parts of speech to each word in the data for training of model. Hindi Senti WordNet is a document consist of polarity with part of speech. The polarity denotes if the word is positive or negative with score [10]. Hindi Stop words contains the words that adds no meaning to the text context. It is always better to discard

these words from source of data, which enables us to save the time during processing [11], [12]. Thus the preprocessing of data will include removal of stop words, tagging part of speech and polarity.

Neural Networks got more importance in the fields like object classification and data fitting because of its effective self-learning and adoption [15]. Deep Learning is an area added to machine learning and its objective is to achieve artificial intelligence [13]. Deep Belief Network (DBN) is an algorithm in deep learning. It is an efficient method that solves problems like low velocity and over fitting from neural network with deep layers. RBMs (Restricted Boltzmann Machine) are stacked and trained to produce a Deep Belief Network (DBN). RBM is made of two layer stochastic network. First is Visible layer and second is Hidden layer are two layer of RBM [15].

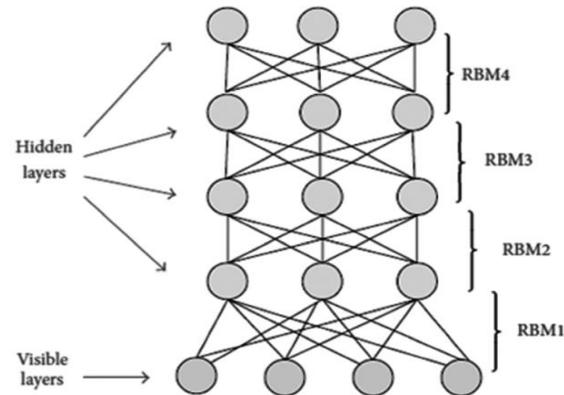


Fig.1 Deep Belief Network

II. RELATED WORK

Prof. Sumitra Pundlik [1] proposed a paper for Multiclass Classification for Hindi Language, in this paper they find out the different Hindi Speeches given by some leaders. This data is classified into multiple classes by finding out topics in it. They performed multiclass classification along with Sentiment Analysis with the help of ontology and they have used combination of Hindi Senti WordNet for finding polarity and LM Classifier to achieve accuracy.

A system is proposed by Reddy Naidu [2] for Sentiment Analysis of Telugu Language. In this paper they have used Telugu Senti WordNet to perform Sentiment Analysis of Telugu e-Newspaper sentences. The proposed system consist of two phase Sentiment Analysis. At initial stage it identifies the subjectivity classification of newspaper sentences as subjective or objective. It treated Objective sentences as Neutral. Then Subjective sentences are further classified into Positive and non-positive sentences.

Charu Nanda, Garima Nanda and Mohit Dua [3] proposed a paper for Sentiment Analysis of Movie reviews in Hindi Language. This paper describes a system that classifies Hindi reviews into Positive and Negative for user to easily identify the polarity of reviews on a movie. Random Forest method and SVM that is Support Vector Machine is used for classification.

Akshi Kumar and Ritu Rani [4] proposed a system for Sentiment analysis using neural network. The proposed paper for this system describes sentiment analysis on the Twitter data. They implemented two types of probabilistic neural network (PNN) to analyze sentiment from the tweets. PNNs is the first model having one value of smoothing parameter and the second model is PNNC designed with different values of smoothing parameter for each class. Also they concluded from the results that PNNC has been found better in performance than PNNs.

Parul Sharma and Teng-Sheng Moh [5] proposed a paper on Indian Election prediction using Sentiment Analysis. They used Twitter Archiver tool for collection of Hindi Language tweets from Twitter. The prediction of election is

performed by using three approaches. And they are Dictionary Based, SVM (Support Vector Machine) and naive bayes classifiers. They got higher accuracy by using SVM algorithm than the other two.

Impana P and Jagadish S Kallimani [6] proposed a paper on sentiment analysis (cross lingual) for Indian Regional Languages. They described the resource poor language which have less data for extraction and resource rich language which have sufficient dataset required for sentiment extraction. They performed sentiment analysis of two languages using Bilingually Constrained Recursive Auto-encoder (BRAE) and WordNet datasets.

Swagat Ranjit [7] proposed a system for Foreign Rate Exchange Prediction. The prediction is performed using Artificial Neural Network (ANN) and Sentiment Analysis. Sentiment Analysis is applied using Naive Bayes and Lexicon Based algorithm and thus analyzed the opinion from traders trading in financial market and predicted the combined sentiment. They concluded that ANN based model is able to forecast the Foreign Exchange rates precisely.

Yaser Maher Wazery [8], in their paper they described the Sentiment Analysis with the help of Deep Neural Network. They used two major approaches of Sentiment Analysis. The first one is by using machine learning approaches such as Decision tree, Support Vector Machine (SVM), K-nearest neighbor and Naive Bayes. A Recurrent Neural Network using Long Short-Term Memory (LSTM) is another approach using Deep Neural Network. They have used three Twitter datasets such as Airline datasets, IMDB and Amazon to perform Sentiment Analysis. And they concluded RNN-LSTM showed higher accuracy from their overall approaches.

III. PROPOSED SYSTEM

In the recent trends of machine learning, Deep Neural Network Algorithm covered the problems of most of the machine learning approaches in different domains [8]. Neural Networks consist of multiple learning layers which makes them more effective learning approach. Adaptive learning, parallelism, fault tolerance and generalization are some features of neural network which provides high performance [4].

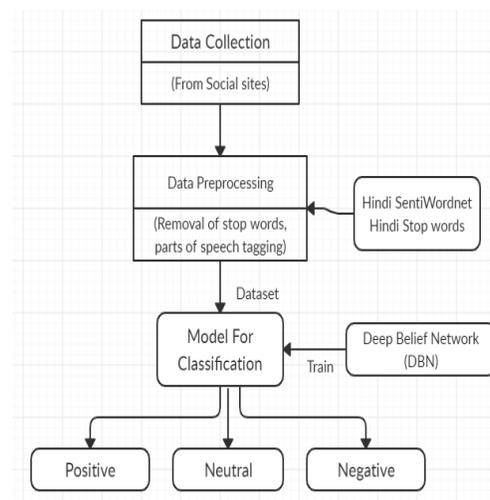


Fig. 2. Proposed System

Thus Deep belief neural network approach in deep neural network can be used in analyzing sentiment from Hindi data. Figure 2 shows the proposed system for classification of Hindi data, the data collected from social sites will be preprocessed using Hindi Senti WordNet and Hindi Stop Words to make data processing faster. Then prepared dataset is given as input to train the model using deep Belief Network (DBN). After training process sentiment can be analyzed from given data into positive, negative and neutral.

IV. COMPARISON OF DIFFERENT METHODOLOGIES

Table1: Comparison of classification techniques

Parameters and Techniques	Parent	Working	Use	Remarks
Naive Bayes	It comes from the family of Probabilistic Classifiers in Machine Learning	This classifier considers that the value of a specific feature is independent of the value of any other feature	Used for classification of text and solves problems having multiple classes	There is chance of inaccuracy if any non-labeled data is found in dataset
Support Vector Machine (SVM)	Type of Supervised learning in Machine Learning	-It uses a hyper plane in N-dimensional space (N - Number of feature) to classify data points distinctly -It selects the hyper plane (decision boundary) that has maximum margin	Supports both classification and regression tasks but mostly suitable for classification tasks	-Does not perform well when dataset have more noise - not suitable for large datasets
Probabilistic Neural Network (PNN)	It is a feed forward type of Neural Network in neural networks	It is a multilayer feed forward neural network consist of a smoothing parameter that plays major role in classification of data	Widely used in solving problems like classification and pattern recognition	Complexity increases with large database classification problems
Convolution Neural Network (CNN)	It is a division of Deep Neural Networks	It uses convolution (an operation in mathematics) for matrix multiplication in at least one layer of neural network	Suitable for analyzing visual imagery such as image and medical image classification, video recognition	It has ability to solve large classification problems but it is expensive when we increase computations
Recurrent Neural Network (RNN)	It is a type of Artificial Neural Networks	-Connections among the nodes of RNN make a directed graph in a temporal sequence -it make use of internal memory to process variable length input sequences	Used for speech recognition and handwriting recognition	Training of RNN is a very complex and difficult task
Deep belief Networks	It is a class of Deep Neural Network in Machine Learning	-It is a generative graphical model consist of multiple hidden layers -It is composition of RBMs (Restricted Boltzmann Machine)	DBN is used for Sentiment classification, image and audio classification	-It works better with human language complexity -DBN with feature selection overcomes its drawbacks (computationally expensive, time consuming)

There are different techniques in machine learning for classification of data. And thus the selection of a method among them is a tedious task. The table shows the comparison between these machine learning techniques using parameters such as its parent, structure and use. SVM and PNN are used for classification but lacks in performance as size of dataset increases. Naive Bayes is less efficient when dataset have some unlabeled data. CNN works better for analyzing image data but gradually becomes expensive with increased computations. DBN is also computationally expensive but DBN with feature selection overcomes this problem. And when each layer of DBN is trained greedily one at a time, results in one of the effective deep learning algorithm [16], [17], [18], [19], [20], [21], [22], [23], [24], [25].

V. CONCLUSION

The different social media platforms contains wide variety of data in different languages on their cloud storage [25]. Thus this motivates sentiment identification from such data which includes views of people. Sentiment classification of data in Hindi can be implemented by different machine learning techniques. But Deep Belief Networks (DBN) results in better classification than several machine learning techniques in use [17]. Also it performs well with human language complexity [16]. So for regional language such as Hindi it will give better results and thus DBN will be used for sentiment analysis of Hindi language.

VI. FUTURE SCOPE

India has wide diversity of languages. Thus we can get data over social media in different regional languages. In future the sentiment analysis can be done using Deep Belief Networks (DBN) in different regional languages such as Marathi.

REFERENCES

- [1] Prof. Sumitra Pundlik, Prachi Kasbekar, Gajanan Gaikwad, Prasad Dasare, Akshay Gawade, Purushottam Pundlik, “ Multiclass classification and class based sentiment analysis for Hindi language”, 2016 Intl. Conference on Advances in Computing, Communication and informatics (ICACCI), Sept. 21-24, 2016, Jaipur, India, pp. 512-518.
- [2] Reddy Naidu, Santosh Kumar Bharti, Korra Sathya Babu, Ramesh Kumar Mohapatra, “Sentiment Analysis Using Telugu SentiWordNet”, IEEE WiSPNET 2017 conference, pp. 666-670.
- [3] Charu Nanda, Mohit Dua, Garima Nanda, “Sentiment Analysis of Movie Reviews in Hindi Language using Machine Learning”, International Conference on Communication and Signal Processing, April 3-5, 2018, India, pp. 1069-1072.
- [4] Akshi Kumar and Ritu Rani, “Sentiment Analysis Using Neural Network”, 2016 2nd International Conference on Next Generation Computing Technologies (NGCT-201) Dehradun, India 14-16 October 2016, pp. 262-267.
- [5] Parul Sharma, Teng-Sheng Moh, “Prediction of Indian Election Using Sentiment Analysis on Hindi Twitter”, 2016 IEEE International Conference on Big Data, pp. 1966-1971.
- [6] Impana P, Jagadish S Kallimani, “Cross-Lingual Sentiment Analysis for Indian Regional Languages”, 2017 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), pp. 1-6.
- [7] Swagat Ranjit, Shruti Shrestha, Sital Subedi, Subarna Shakya, “Foreign Rate Exchange Prediction Using Neural Network and Sentiment Analysis”, International Conference on Advances in Computing, Communication Control and Networking (ICACCCN 2018), pp. 1173-1177.
- [8] Yaser Maher Wazery, Hager Saleh Mohammad, Essam Halim Houssein, “Twitter Sentiment Analysis using Deep Neural Network”, 2018 14th International Computer Engineering Conference (ICENCO) , Cairo, Egypt, pp. 177-182.
- [9] Siyuan Chen, Chao Peng, Linsen Cai, Lanying Guo, “A Deep Neural Network Model for Target-based Sentiment Analysis”, 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brail, pp. 1-7.
- [10] Internet Link: <http://cdn.iiit.ac.in/cdn/ltrc.iiit.ac.in/icon/2014/proceedings/File70-p154.pdf>

- [11] Internet Link: <https://tdil-dc.in/>
- [12] Internet Link: <https://data.mendeley.com/datasets/bsr3frvvc/1>
- [13] Internet Link: [http://deeplearning.net/tutorial/DBN.html\(DBN\)\(DB\)](http://deeplearning.net/tutorial/DBN.html(DBN)(DB))
- [14] Internet Link: https://en.wikipedia.org/wiki/Natural_language_processing
- [15] Yuming Hua, Junhai Guo, Hua Zhao, “Deep Belief Networks and Deep Learning”, 2015 International Conference on Intelligent Computing and Internet of Things (ICIT), pp. 1-4.
- [16] Patrawut Ruangkanokmas, Tiranee Achalakul, and Khajonpong Akkarajitsakul, “Deep Belief Networks with Feature Selection for Sentiment Classification”, 2016 7th International Conference on Intelligent Systems, Modelling and Simulation, pp. 9-14.
- [17] Ruhi Sarikaya, Geoffrey E. Hinton, and Anoop Deoras, “Application of Deep Belief Networks for Natural Language Understanding”, IEEE/ACM Transactions on Audio, Speech and Language Processing, Vol.22, No.4, pp. 778-784, April 2014
- [18] Internet Link: https://en.wikipedia.org/wiki/Support-vector_machine
- [19] Internet Link: https://en.wikipedia.org/wiki/Probabilistic_neural_network
- [20] Internet Link: https://en.wikipedia.org/wiki/Naive_Bayes_classifier
- [21] Internet Link: https://en.wikipedia.org/wiki/Convolutional_neural_network
- [22] Internet Link: https://en.wikipedia.org/wiki/Recurrent_neural_network
- [23] Internet Link: https://en.wikipedia.org/wiki/Deep_belief_network
- [24] Internet Link: <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>
- [25] Nikhil Wagh, Vikul Pawar and Kailash Karat, “Implementation of stable Private Cloud using Openstack with Virtual Machine Results”, International Journal of Computer Science and Engineering & Technology (IJCET-19) Vol 10, Issue 2, pp. 258-269, March-April 2019