ISSN: 2096-3246
Volume 51, Issue 01, Jan, 2019

https://doi.org/10.5281/zenodo.12706884

Deep Belief Networks for Sentiment Analysis on Hindi Language

EDIGA KISHORE KUMAR GOWD, P VISWANATH, VIJAYA BHASKAR MADGULA

Assistant Professor [1,2,3]

KISHORE_EDIGA@GMAIL.COM, viswanath.p002@gmail.com, vijaya.bhaskar2010@gmail.com

Department of Computer Science and Engineering, Sri Venkateswara Institute of Technology, N.H 44, Hampapuram, Rapthadu, Anantapuramu, Andhra Pradesh 515722

Keywords:
Automata, Back-propagation, Gradient, Sentiment Analysis, Statistics, Unigram, WordNet.

Abbreviations: DBN, deep belief network; RBM, Restricted Boltzmann machine.

ABSTRACT

We now have a mountain of data due to the constant flow of information from various social media sites. It is crucial to digest data and extract emotions or important elements from it. An approach that may help with this is sentiment analysis. There has to be an English-like system that can decipher regional languages like Hindi for sentiment analysis. Machine translation is one of various emotion identification methods; nonetheless, it incurs the cost of translating across languages. This research presents a Deep Belief Network–based method for sentiment analysis of Hindi data. When it comes to Hindi data, this neural network model outperforms the machine translation method. An improvement in performance may be achieved by combining sentiment analysis with deep learning [4]. This sentiment categorization is best handled by a deep belief network, one of many deep learning neural network models [5]. To categorise Hindi reviews as either favourable, bad, or neutral, this study used a Deep Belief Network model trained using Supervised Learning.

ISSN: 2096-3246
Volume 50, Issue 02, April, 2018

https://doi.org/10.5281/zenodo.12706884

Introduction

Classifying provided data into positive, negative, and neutral categories is part of Sentiment Analysis, which falls within the purview of Natural Language Processing [1]. Neural network processing primarily aims to determine the purpose of a given text [10] [11]. Because it ranks fourth among the world's most spoken languages, Hindi was selected as the source language for sentiment analysis in this article [2]. One of the most promising new areas of machine learning is deep learning, and a particularly effective model within this field is the deep belief network [6]. When it comes to learning, Deep Belief Networks are strong enough to fix issues like poor velocity and overfitting [12]. Among the components of deep learning are multi-layered neural networks. The incoming data is abstracted and transformed into a more composite form by each layer as it learns [10]. The training of the neural network in this study is based on supervised learning, as the input data includes fixed labels like positive, negative, and neutral.DBN is a generative model that uses connections between hidden layers rather than values to generate new output. Several Restricted Boltzmann Machine (RBM) layers make up these layers [14]. A DBN using gradient descent and back propagation is the outcome of stacking RBM. The neural network model is trained for supervised learning using the back propagation technique. An efficient computation of the gradient according to weights is achieved by it. As a result, relative error is reduced. during training by adjusting input-related weights. The end result is a fine-tuned model that can accurately classify the incoming data [13].

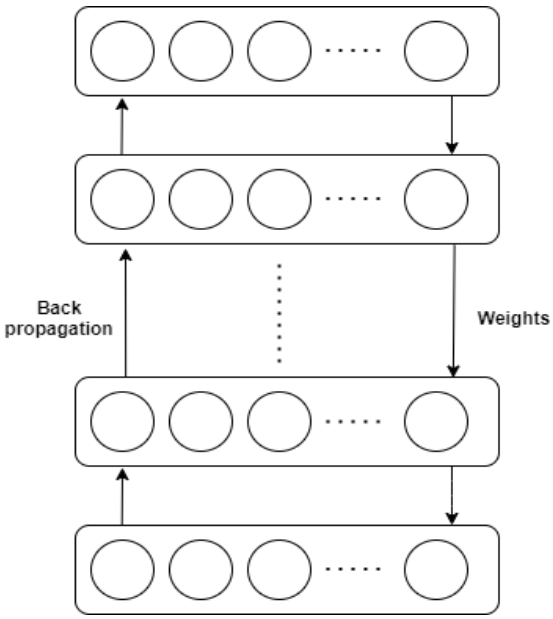

Figure 1. DBN

The above diagram visualizes the structure of a Deep Belief network. It learns with the weights assigned and feed forward to next layers. The back propagation deals with the minimization of deviation from desired output.

III. PROPOSED SYSTEM

For this research human sentiments plays a vital role. The system require data created by humans in Hindi language to classify. Thus the flow of the system starts from the collection of Hindi data and ends in classifying them into positive and negative categories. The figure number 2 shows the detailed structure of the system.Social media became a crucial part of all people's life. They feel easy to express opinion and their views about a topic[7]. This data present on the clouds

ISSN: 2096-3246
Volume 50, Issue 02, April, 2018
https://doi.org/10.5281/zenodo.12706884

of these platforms [17] and collected in a file will act as an input to the system. The data can be positive and negative Hindi sentences. This will help to train and test the system against its accuracyof classification.

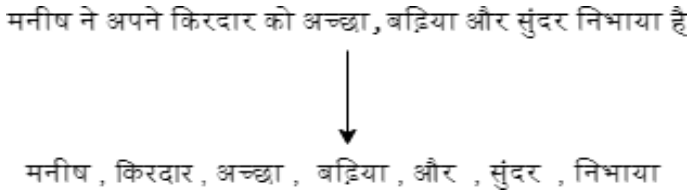

Figure 2. Stop words removal

Part of Speech tagging(POS)

We need to annotate data with language-specific grammatical terms if we want to get deeper insights from it. Every single word in the dataset has a tag linked to it. Nouns, pronouns, adjectives, adverbs, and so on may all serve as tags. [8]. Hello there! With the aid of WordNet, we can label every Hindi word. Stop words are only one example of a class of words that are useless as input. To exclude these kinds of terms from the data, the Hindi Stop words dataset is a great resource. These procedures are performed on the input data prior to it being fed into the neural network.

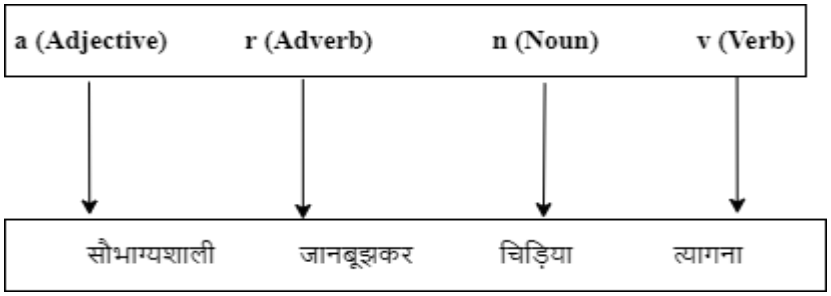

Figure 3. Parts of speech tagging

a. Tuple based score calculation

Hindi Senti WordNet contains scores of about all the Hindi words. These are for positivity and negativity of the word. Thus the positive and negative scores from dataset are taken and checked against positive score, accordinglylabels are attached to the words as a positive or negative word. The tuple based score calculation box in the architecture shows the algorithm for the same.

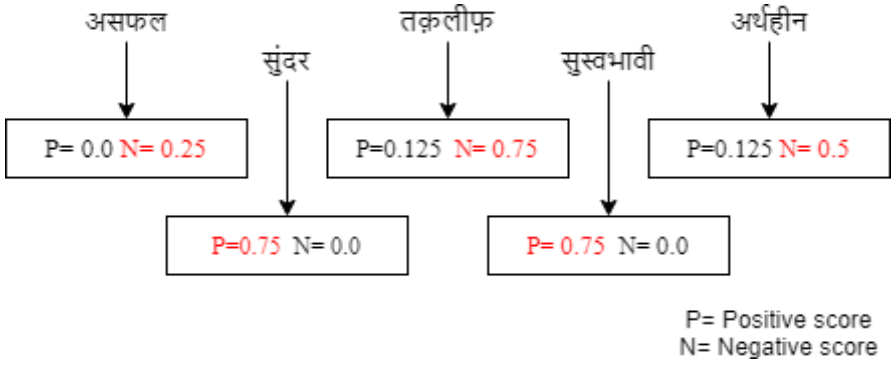

Figure 4. Polarities of words

ISSN: 2096-3246
Volume 50, Issue 02, April, 2018

https://doi.org/10.5281/zenodo.12706884

b. Training dataset

You can't feed the obtained data into the system to use it for data categorization. There has to be pre-processing. This leads to the creation of the training dataset for the Deep Belief Network. It is now time to add the processed data to the training dataset.

c. Initialization of DBN

To begin training the neural network, this is the first step. Words in Hindi that have been tagged as positive or negative make up the training dataset that the model is trained against. Input is sent to its sublayers using DBN feeds. There is a chance that anything may go wrong during training. This inaccuracy represents a deviation from the intended result. The inaccuracy will grow exponentially if the model is trained with this mistake and not addressed. Reverse propagation is carried out after a certain number of iterations to guarantee that the weights are adjusted in accordance with the inaccuracy. We keep doing this training until we have an error-free output that is very near to the target output.

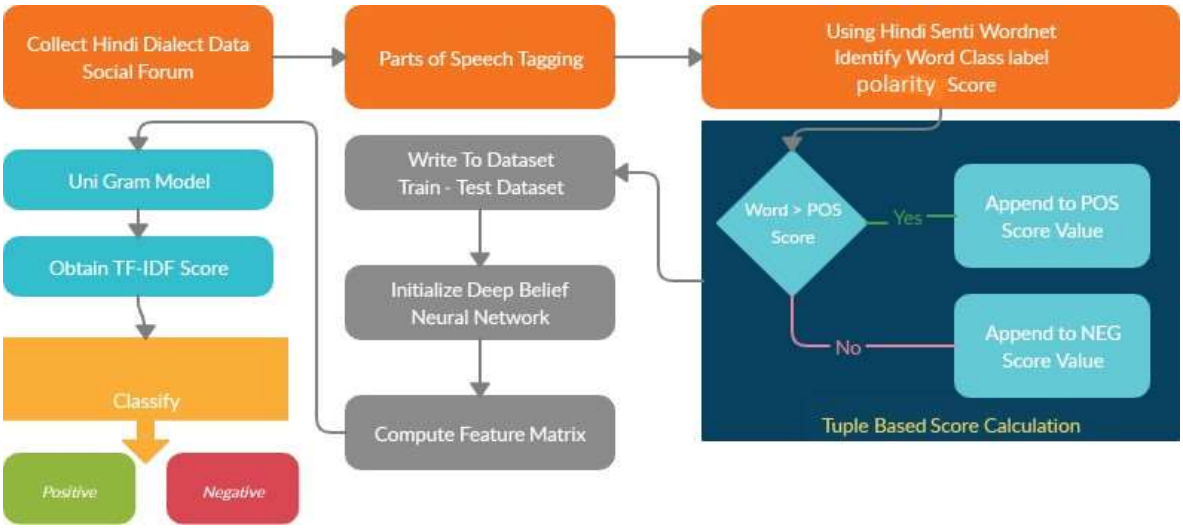

Figure 2. Proposed Architecture

d. Feature matrix

Every machine learning technique uses a set called as feature set or feature matrix. This matrix shows the columnsand its variable to be processed under the training. These are nothing but the lines of observation. This extractionor selection of features used for initialization of the DBN model.

e. Uni-Gram model

Unigram model is one of the type of Language model. The statistical language model is responsible for finding probability distribution of a given sequence of words. It gives probability value to the complete sequence. In unigram model is based on one state finite automata. This model gives probability to each word but this probability depends upon that word's individual probability in the document [15]. Thus the sentiment will be identified depends on the words occurrence in that sentence.

A ES
https://doi.org/10.5281/zenodo.12706884
ISSN: 2096-3246
Volume 50, Issue 02, April, 2018

f. TF-IDF Score

It is term frequency inverse document in statistics. It is useful in identifying the importance of the word in a document. It also works on the occurrence of a word in a document [16].

g. Classification

At this stage the decision is taken about sentiment of the given data by the model. In the initial stage tuple based classification was done, here we are concerning with the polarity of given testing data. After the polarity given by the model, the decision is made by summing up the votes assigned to polarities. Hence results in classification ofsentiment into positive and negative.

IV. RESULTS

While there are a number of machine learning methods for doing sentiment analysis on the data, we settled on a neural network approach. The model utilised to categorise the Hindi user evaluations was a deep belief network. The model was trained using user-reviewed Hindi Sentences and Hindi WordNet. Machine learning consists of two stages: training and testing. During training, we achieved an accuracy of around 98%. Additionally, we determine important index factors that demonstrate our system's performance in relation to the provided Hindi data.

Approach	**Key Index parameters**	**values**
DBN	Specificity	95.174
	Sensitivity	92.918
	Accuracy	95.92
	F measure	93.74

Table 1

In the table we can check for the different key Index parameters and respective values. The values are calculated over different test cases. The table shows the mean values for each key index parameters. In the testing phase we got accuracy around 95.92 percent, which is mean of accuracies in testing phase.

ISSN: 2096-3246
Volume 50, Issue 02, April, 2018
https://doi.org/10.5281/zenodo.12706884

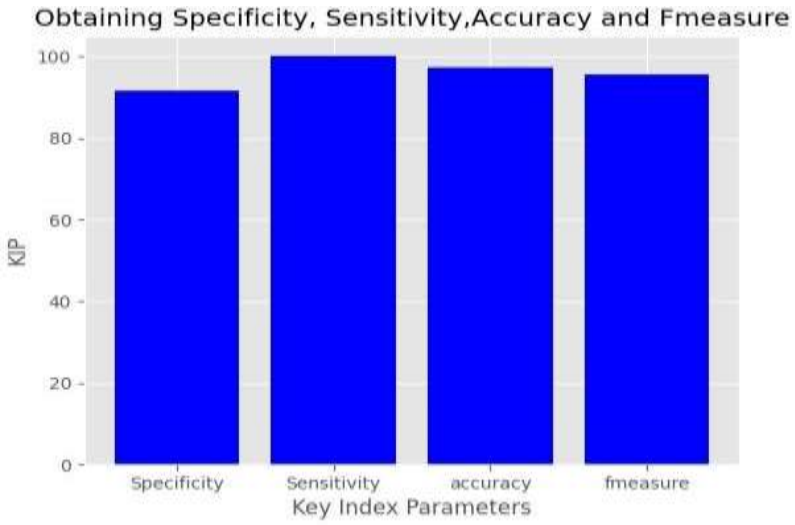

Figure 3. Key Index Parameters

The DBN system of sentiment classification when applied on a test case the system gives output as positive sentimentor negative with the key index parameters like specificity, sensitivity, accuracy and f measure using a statistical graph.

The above graph shows the output for one of the test case in testing phase. The table 1 shows the values calculated bythe taking the mean of each test case output given by the system.

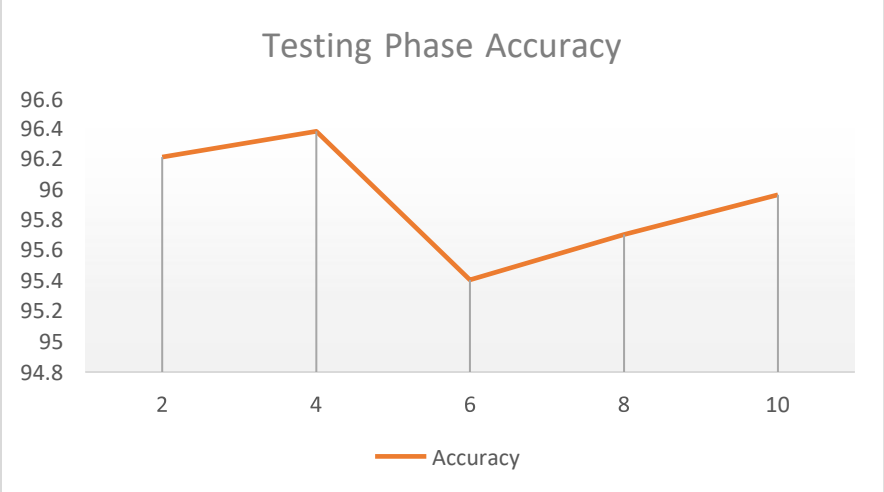

Figure 4. Testing Phase Accuracy

During testing, the system is put through its paces using various test cases and its output is double-checked. Next, we expand the number of test instances and verify the outcome. The graph makes it easy to see how the system performs as the number of test cases increases. Up ahead, the number of test instances rose by 2, 4, 6, etc. Graphing the system's accuracy in analysing sentiment from Hindi data reveals an average of 95.22%.

In a study that Charu Nanda, Mohit Dua, and Garima Nanda presented, they analysed Hindi sentiment using a system they developed using the Random Forest method and Support Vector Machine (SVM). Accuracy rates of 91.01 and 89.73 were achieved in this study [3]. For this reason, we advocated a DBN-based system, which is a

A ES
ISSN: 2096-3246
Volume 50, Issue 02, April, 2018
https://doi.org/10.5281/zenodo.12706884

neural network-based approach to classification; we achieved testing accuracy of 95.22% and training accuracy of 98%, outperforming the SVM method.

v. CONCLUSION

Deep Belief Network is one of several machine learning neural network models that excels at processing complicated human input. DBN enhances system performance with feature selection. Because of this, as compared to other machine learning methods, Deep Belief networks trained using Hindi's WordNet provide superior sentiment analysis results. When evaluated using Hindi data, the neural network model had an average accuracy of about 95.92 percent. The data was more accurately categorised as good or negative. Obtaining high-quality training datasets is also crucial for data classification; with better datasets in the future, DBN will be able to classify data more accurately.

REFERENCES

[1] Reddy Naidu, Santosh Kumar Bharti, Korra Sathya Babu, Ramesh Kumar Mohapatra, "Sentiment Analysis Using Telugu SentiWordNet", IEEE WiSPNET 2017 conference, 666-670.

[2] Prof. Sumitra Pundlik, Prachi Kasbekar, Gajanan Gaikwad, Prasad Dasare, Akshay Gawade, Purushottam Pundlik, " Multiclass classification and class based sentiment analysis for Hindi language", 2016 (ICACCI)Intl.Conference on Advances in Computing, Communication and informatics, Sept. 21-24, 2016, 512-518.

[3] Charu Nanda, Mohit Dua and Garima Nanda, "Sentiment Analysis of Movie Reviews in Hindi Language usingMachine Learning", International Conference on Communication and Signal Processing, April 3-5, 2018, 1069-1072.

[4] Qurat Tul Ain, Mubashir Ali, Amna Riaz, Amna Noureen, Muhammad Kamran, Babar Hayat and A. Rehman "Sentiment Analysis Using Deep Learning Techniques: A Review" International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 8, No. 6, 2017

[5] Nikita Kolambe ,Yashashree Belkhede and Nikhil Wagh,, "A Review on Sentiment Analysis on Hindi language Using Neural Network", International Journal of Analytical and Experimental Model Analysis (IJAEMA) Volume XII , Issue IX September/2020

[6] 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, 177-182.

[7] Yashashree Belkhede, Dr. Praveen Shetiye and Dr.Avinash Gulve, "Review on election prediction using machine learning technique", International Journal of Analytical and Experimental Model Analysis (IJAEMA) Volume XII , Issue IX September/2020

[8] Raksha Sharma, Pushpak Bhattacharyya "A Sentiment Analyzer for Hindi Using Hindi Senti Lexicon" Proceedings of the 11th International Conference on Natural Language Processing, December 2014, 150–155

[9] Internet Link: https://tdil-dc.in/

[10] Internet Link: http://deeplearning.net/tutorial/DBN.html(DBN)(DB)

[11] Internet Link: https://en.wikipedia.org/wiki/Natural_language_processing

[12] Internet Link: https://en.wikipedia.org/wiki/Support-vector_machine

[13] Internet Link: https://en.wikipedia.org/wiki/Restricted_Boltzmann_machine

[14] Internet Link: https://en.wikipedia.org/wiki/Deep_belief_network

[15] Internet Link: https://en.wikipedia.org/wiki/N-gram

ISSN: 2096-3246
Volume 50, Issue 02, April, 2018

https://doi.org/10.5281/zenodo.12706884

[16] Internet Link: https://en.wikipedia.org/wiki/Tf%E2%80%93idf

[17] Nikhil Wagh, Vikul Pawar and Kailash Karat, "Implementation of stable Private Cloud using Openstack withVirtual Machine Results", International Journal of Computer Science and Engineering & Technology (IJCET-19) 10(2): 258-269, March-April 2019