

# Sentimental Analysis on Hindi Text Use of Social Media Network Data

<sup>1</sup>AYAZ AHMED FARIDI, <sup>2</sup>TRYAMBAK HIWARKAR

Research Scholar PhD (CS), SARDAR PATEL UNIVERSITY BALAGHAT (M.P.), INDIA  
Professors, SARDAR PATEL UNIVERSITY BALAGHAT (M.P.), INDIA

Date of Submission: 15-09-2022

Date of Acceptance: 24-09-2022

**ABSTRACT:** Sentiment analysis means analyzing the sentiment in formation in order to draw the conclusion and understand the overall situation. Understanding machine data is important to a person because inputs such as product reviews, news, movie reviews, reviews, and blogs or any social comments contain emotion. Website and give output as positive or negative or neutral view. There are various type of algorithms and classification are present for sentimental analysis. This research gives a survey of different sentiments and attitudes used specifically for the Hindi language.

**Keywords:** -Sentiment analysis, Classifier, Sentiments Emoji, Gesture Text, NLP.

## I. INTRODUCTION: -

Sentiment analysis is the process of determining the emotional tone behind a series of words which is used to gain and understand the attitudes, opinions and emotions expressed from language. It is highly useful in social media monitoring as it allows us to gain an overview of the wider public opinion behind in topics. The applications of sentimental analyst are broad and powerful. The abilities to extract insights from media data of social world (twitter) is a practice that is being widely adopted by organizations across the Senti ward net world view. So we need asentiment analysis system for Hindi language. An application may include knowing how curious one is about an upcoming movie, participating in a marathon in a particular type of city, reviewing any new products in the market, changing written restaurants, etc. review into 5-star scale a cross various categories like food quality, service and value of money. Majority of the existing work has been done for English language. As the Internet is reaching to more people with in the world, there is huge increase in web content as people feel comfortable with their native language. So, there is

a need to pay attention in direction of sentiment classification for Hindi Language. In this paper, wear giving a survey of various classifiers and the approaches used for sentiment analysis specifically for Hindi language data.

## II. SENTIMENTANALYSIS: -

Sentiment Analysis is a natural language processing task that deals with the extraction of opinion from apiece of text with respect to a topic (Pang et al., 2008). Hindi is most commonly spoken language in the world. So, these information seen on Hindi-englishis important to be analyzed for the use of industries and government(s) [7] Sentiment analysis is very difficult for Hindi language due to numerous reasons as follows: [7]

- (1) Unavailability of well a notated standard corpora, therefore supervised machine learning algorithm scan not is applied.
- (2) Hindi is a resource scarce language; there are no efficient parser and tagger for this language. Limited resources available for this language like Hindi-Senti-Word-Net (HSWN). It consists of limited numbers of adjectives and adverbs. All the words are available in inflected forms. HSWN is created using the Hindi-Word-Net and English Senti-Word-Net (SWN).During the creation of this resource for Hindi language, it is assumed that all synonyms have the same polarity while all antonyms have the reverse polarity of aword.This assumption neglected word sense intensity in terms of polarity, however polarity in density of their word.
- (3) Important in opinion mining.

## III. LIMITATION:

The proposed approach relies on the availability of the resources needed to analyze the sentiment content in Hindi.

### 1.1 Machine Translation (MT) based Sentiment Analysis

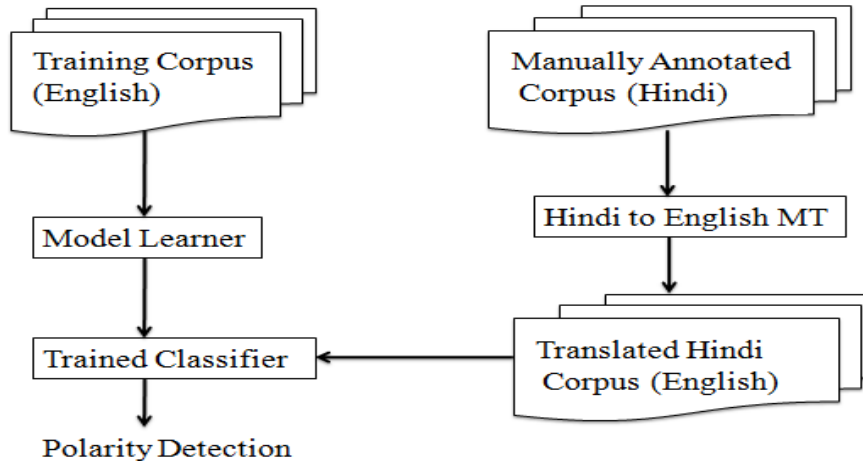


Figure 2: Procedure for MT-based Sentiment Analysis [2]

**1.2 Limitation:** Online Google translate is used to translate Hindi to English words.

#### 1.3 Resource based Sentiment Analysis

A classifier is implemented to uses cores in HSWN  
 Stemmer and stop word list are used.

Algorithm

- For each word in the document,
- Applies top word removal and stemming

(depending on the variant of the experiment)

- Look up the sentiment scores for each word in the H-SWN.
- Assign a polarity to a word based on the maximum of the scores
- Assign to a document the polarity which majority of its words possess.

Sense Consideration	Stemming	Stop Word Removal	Accuracy
Most Commonsense	No	No	56.35
	Yes	Yes	53.96
All Senses	No	No	60.31
	No	Yes	57.53
	Yes	Yes	55.95

Figure 3: Results for resource-based sentiment analysis [2]

Algorithm for creation of H-SWN

- 1.3.1 For each synset in the SWN, repeat 2 to 3
- 1.3.2 Find the corresponding synset in Hindi Word Net
- 1.3.3 Project the cores of a syn set in SWN to the corresponding synset in Hindi Word Net.

Experiment	Accuracy%
In-language sentiment analysis	78.14
MT-based sentiment analysis	65.96
Resource-based sentiment analysis	60.31

Figure 4: Comparison of approaches [2]

**2.4 Limitation:** Even though the word is present in the data base the correct sentence with meaning is not accurate i.e. the meaning of the original

language changes. This can be overcome by words ended is ambulation. Paper [2] Hindi Subjective Lexicon: A Lexical Resource for Hindi Polarity.

#### IV. CLASSIFICATION:-

Bakliwal et al. (2012) created lexicon using a graph based method. They explored how the synonym and antonym relation can be exploited using sample graph traversal to generate the subjectivity lexicon. Their proposed algorithm achieved approximately 79% accuracy. [3] Observation

- Developed lexicon of adjectives and adverbs with polarity scores using Hindi Word Net
- Developed notated corpora of Hindi Product Reviews.
- Simple scoring method classification is used.
- Shallow parser is used to identify adjectives & adverbs in a sentence.
- Negation handling was performed.
- Adjective scoring-Baseline stem Nh gave maximum accuracy of 78.41 and Adjective & adverb scoring for base line stem NH gave 79% accuracy.

**5.1 Limitation:** Extending your research work to include Word Sense Disambiguation (WSD) and symbolic form can result in better values for words in which the nature is dual. I use it with adjectives and adverbs in my research done, this function can be extended to other parts of the text (verbs and nouns). Paper [3] Sentiment Analysis in Hindi Naman Bansal and UmairZ.Ahmed. Advisor: Amitabha Mukherjee used Semi-Supervised approach to train Deep Belief Network on small percentage of labelled data (150) and assign polarity to unlabeled data. 300 sentences dataset stored in xml. Sentence level sentiment is adopted. They report 71% accuracy using DB Non-English language, 76% using active deep learning and 64% on Hindi language using Deep Belief Network. [4]

#### V. OBSERVATIONS:

Movie reviews were obtained from IITB for research purpose and content from jagran.com web site was collected and we restored in XML format so that it is easily to parse. These supervised learning method based on DBN architecture is divided into two steps: first is pre-training model is constructed by greedy layer-wise unsupervised learning using Restricted Boltzman Machine (RBM) in which all the labeled data together with L unlabeled data are utilized to find the parameter space W with N layers. Second: fine log like LaHood using gradient descent method is used. The parameter space W is retained negative log like LaHood cost function using Labeled data to find the parameters space only according to labeled

data. Experiments carried out for multiple configurations of number of neurons in the hidden layers. The best configuration for the deep belief network founded was Five-layer network, One Input, Three Hidden and One Output Layer.

**4.1 Limitation:** The sentences are marked incorrectly as negative or positive even though they are not. Negation handling is missing. Paper [4] A Hybrid Approach for Twitter Sentiment Analysis Namita Mittal et al. (2013) introduced an approach for automatically classifying the sentiment of Twitter messages. A 3 stage hierarchical model is proposed by the author; first: labeling with emoticons; second: using predefined list of words with strong positive or negative sentiments; third: tokens are weighted based on subjectivity lexicon. [5]

**4.2 Observations:** The author used 60,000 tweets. Rules for assigning weight to the two classes which are not present in the positive as well as negative list of words. Accuracy obtained without discourse by Sent iWord Net was 64.694, proposed probability-based method 71.116, Senti Word Net then probability based method 69.511, probability based method then Senti-Word -Net 71.83, Hybrid approach (SWN+Probability) 72.563. With discourse results obtained for the accuracy were; for Senti Word Net 66.052, proposed probability based method 71.625, Senti Word Net then probability based method 70.193, probability based method then Senti Word Net 73.35, Hybrid approach (SWN+Probability) 73.72. There shows that in corporation of discourse makers improves the sentiment classification accuracy.

**4.2.1 Limitation:** Rules for handling sarcasm is missing. Paper [5] Sentiment Analysis of Hindi Review based on Negation and Discourse Relation Namita Mittal et al. (2013) proposed method for improving HSWN. 770 reviews were taken as input-gram & POS- tagged N-gram approaches were also used. Fleiss kappa coefficient value 0.8092 & data set has 104 words.

**4.3 Observations:** An algorithm for negation handling and discourse relation is given. Accuracy of 82.89% for positive views, 76.59% for negative and overall 80.21%. for improved HSWN + negation Discourse. [6]

**4.3.1 Limitations:** The size of the data set is small. Paper [6] A Survey on Sentiment Analysis and Opinion Mining Techniques Aman deep Kaur and Vishal Gupta (2013) gave the process of Sentiment

Analysis is divided in to 5 steps: process of Sentiment Analysis for Text (Lexicon generation), Subjectivity detection, sentiment polarity detection using Network Overlap Technique, sentiment structuration, sentiment summarization visualization tracking. [7]

**4.4 Observations:** There are four approaches to predict the polarity of a word. In this First stage an interactive game is provided which identify the polarity of the words. In the Second strategy, a bilingual dictionary is developed for English and Indian Languages. In the 3rd stage, word net expansion is done using antonym and synonym relations. In the fourth approach, are annotated corpus is used for learning.

**4.4.1 Limitation:** subjective lexicon can be developed for the unexplored languages which does not have a word net. The basic resources like parsers, name density recognizers, morphological analyzers, and part of speech tagger need to be improved to reach the state of accuracy. Paper [7] Sentiment Classification in Hindi Sneha Mulatkar (May 2014) used WSD algorithm (Sense disambiguation) is given and is used to find the correct sense of the word on a context. SVM is used. Cxv Term presence, vs term frequency, term position is described. [8]

**4.5 Observations:** SVM separate the 2 categories and build a wide gap which are root words and

some are some of the affixes. Steps of algorithm contains: stop words removal, sorting single column word line and stemming performed on sorted list in which each word is compared with next 10 words assuming that minimum that 10 morphological variants is present in list. If word is present as substring in next word then the word is broken as substring + remaining characters of word. Sub string is treated as root/ base for m and remaining characters are treated as affix.

**4.5.1 Limitation:**

Term position is important. Words appearing in the 1st few sentences and last few sentences in a text are given more weight tagged than those appearing elsewhere.

The following table presents the summary of literature. Paper [8], A Framework for Sentiment Analysis in Hindi using HSWN Pooja Pandey & Sharvari Govilkar (2015) used Hindi Senti Word Net (HSWN) for Hindi movie review to find the overall sentiment associated with the document. Polarity of the words in the review are extracted from HSWN and then final aggregated polarity is calculated which can sum either positive, negative or neutral. Synset replacement algorithm is used to find polarity of those words which don't have polarity associated with it in HSWN. Negation and discourse relations which are mostly present in Hindi movie review are also handled to improve the performance of the system. [9]

Table 1: Summary of Literature Survey for Sentiment Analysis for Hindi Language

Title of the paper	Author & Year of publication	Observations/Remarks
A Fall-back Strategy for Sentiment Analysis in Hindi: a Case study	Aditya Joshi, Balamurali AR, Pushpak Bhattacharya, 2010	They used 3 approaches: training classifier, translate given document to English & develop lexical resource called Hindi-Senti Word Net (H-SWN). Naive approach, Senti Word Net 1.1 is used. WordNet linking is used to map synsets of Word Nets of different languages. POS tagging is done. 250 Hindi movie reviews. Rapid Miner 5.0 for document classification. Lib SVM type-Clear for classification. TF-IDF gave highest accuracy of 78.19% & 60.31% for resource-based SA.

Hindi Subjective Lexicon: A Lexical Resource for Hindi Polarity Classification	Bakliwal etal.(2012)	Created lexicon using a graph based method. They explored how the synonym and antonym relations can be exploited using sample graph traversal to generate the subjectivity lexicon. Their proposed algorithm achieved approximately 79% accuracy.
Sentiment Analysis in Hindi	Naman Bansal and Umair Z. Ahmed. Advisor: Amitabh Mukherjee, IIT Kanpur	They used Semi-Supervised approach to train Deep Belief Network on small percentage of labelled data (150) and assign polarity to unlabeled. 300 sentences dataset stored in xml. Sentence level sentiment is adopted. Data preprocessing stage, Deep belief architecture divided in 2 stages: pre-training model & fine-tuning step. They report 71% accuracy using DBN on English language, 76% using active deep learning and 64% on Hindi language using DBN.
A Hybrid Approach for Twitter	Namita Mitta etal. (2013)	A 3 stage hierarchical model for automatically classifying twitter messages is proposed by the author; 1) labeling the emoticons; 2) using predefined list of words with strong positive or negative
Sentiment Analysis		Sentiments; 3) tokens are weighted based on subjectivity lexicon. Maximum Accuracy obtained by hybrid approach with discourse was 73.72%.
Sentiment Analysis of Hindi Review based on Negation and Discourse Relation	Namita Mittal, Basant Agarwal, Garvit Chauhan, Nitin Bania, Prateek Pareek	Method for improving HSWN is proposed. 770 reviews. N-gram & POS-tagged N-gram approaches used. Fleiss kappa coefficient value 0.8092 & dataset has 104 words. Accuracy of 82.89% for positive reviews, 76.59% for negative and overall 80.21%. for improved HSWN + negation + Discourse.
A Survey on Sentiment Analysis and Opinion Mining Techniques	Amandeep Kaur, Vishal Gupta, 2013	The process of Sentiment Analysis is divided into 5 steps: process of Sentiment Analysis for Text (Lexicon generation), Subjectivity detection, sentiment polarity detection using Network Overlap Technique, sentiment structuration, sentiment summarization - visualization-tracking.
Sentiment Classification In Hindi	Sneha Mulatkar, May 2014	WSD algorithm (Sense disambiguation) is given and is used to find the correct sense of the word on a context. SVM is used. Term presence, vstem frequency, term position are described.



<p>A Framework for Sentiment Analysis in Hindi using HSWN</p>	<p>PoojaPandey,SharvariGovilkar,June2015</p>	<p>Existing HSWN is improved with the help of English Senti WordNet, where sentimental words which are not present in the HSWN are translated to English and then searched in English Senti WordNet to retrieve their polarity. Sentiment is extracted by finding the overall polarity of the document; which can be positive, negative or neutral. During pre-processing tokens are extracted from sentence and spell check is performed. Rules are devised for handling negation and discourse relation which highly influence the sentiments expressed in the document .Finally, overall sentiment orientation of the document isDetermined by aggregating the polarity values of all the sentimental words in the document.</p>
---	--	---

## VI. CONCLUSION

Sentimental analysis is analyzing the sentiment information from the input like product review, movie review, news review, comments from blogs or any social web site in order to draw the conclusion and understand the overall situation. In order to understand data machine needs to understand what sentiments from the input are and give output as positive or negative or neutral reviewing English languageA great deal of work has been done in Emotion and Emotion, as English is a global language, but sentiment analysis needs to be done in other languages as well.

## REFERENCES

- [1] Daniel Jurafsky and James H. Martin, Speech and Language Processing, "An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition", University of Colorado, Boulder, Pearson Education series, 2011.
- [2] Aditya Joshi, Balamurali A R, Pushpak Bhattacharya, Dept of Computer and Science Engineering, IIT Bombay, IITB-Monash Research Academy, IIT Bombay, Mumbai, India-400076, "A Fall-back Strategy for Sentiment Analysis in Hindi: a Case study", Proceedings of ICON2010:8<sup>th</sup> International Conference on Natural Language Processing.
- [3] AkshatBakliwal, Piyush Arora, Vasudeva Varma, "Hindi Subjective Lexicon: A Lexical Resource for Hindi Polarity Classification", Proceedings of the Eight International Conference on Language Resources & Evaluation (LREC), 2012
- [4] Naman Bansal and Umair Z Ahmed, Advisor: Amitabha Mukherjee, Department of Computer Science and Engineering, Indian Institute of Technology, Kanpur, India, "Sentiment Analysis in Hindi"
- [5] NamitaMittaletal., "A Hybrid approach for Twitter Sentimental is", 2013
- [6] Namita Mittal, Bansat Agarwal, GarvitChauhan,NitinBania, PrateekPareek, "Sentiment Analysis of Hindi review based on Negation and Discourse Relation",InternationalJointConferenceonNatural LanguageProcessing,pg.45-50,Nagoya,Japan,14-18 Oct 2013.
- [7] Amandeep Kaur and Vishal Gupta, "A Survey on Sentiment Analysis and Opinion Mining Techniques", Journal of Emerging Technologies in Web intelligence. VOL 5, no.4, Nov2013
- [8] SnehaMulatkar, "Sentiment Classification in Hindi", International Journal of Scientific & Technology research (IJSTR2014),Vol.3,Issue5, May2014
- [9] Pooja Pandey, SharvariGovilkar, "A Framework for Sentiment Analysis in Hindi using HSWN", International Journal of Computer Applications (IJCA) (0975-8887), Vol.119-No.19, June2015
- [10] MukeshYadav,VarunakshiBhojane,"DataAn alysis&SentimentAnalysisforUnstructuredD ata",InternationalJournalofEngineeringTech nology,managementandAppliedSciences(IJE TMAS), Vol.2,Issue 7,Dec 2014 ISSN 2349-4476