International Journal of Scientific Research & Engineering Trends
Volume 7, Issue 2, March-April-2021, ISSN (Online): 2395-566X

"A Model for Sentimental Analysis of Twitter Data in Hindi Polarity: A Study"

Suraj Prasad Keshri[1], Neelam Sahu[2]
[1]Research Scholar Ph.D (CS), [2]Assistant Professor
[1]Information Technology & computer science
[2]Dr.C.V.Raman University, Kota Bilaspur (C.G.), India
suraj.softtech11@gmail.com[1], neelamsahu16@gmail.com[2]

***Abstract*- This research explores real-time feelings of analysis. Twitter / WhatsApp posts are based in Hindi, with a sense of classification on a three-way scale, negative, positive and natural. The efficiency of various methods such as speech (POS) tagging and part of the stop word extract are compared and WordNet is proposed to improve the Hindi Sentimental Word. The current work of analysis on sentiment has been done in mining. There are resource-rich languages such as English; while data in Indian languages is still relatively low web content in languages like Hindi has grown rapidly over the past few years. In addition, as the length of social media increases, WhatsApp has become a hotbed of user content on platforms such as Twitter and Facebook SA of tweets made in English. There is not an analysis of similar data in Hindi. In this paper, we perform real-time SA on a live stream of Hindi tweets. Twitter social media application is a popular site Forum to share opinions of millions of people there are users on a variety of topics. The Analysis and opinion mining websites are rich sources of data for sentiment. A Twitter user can often use hash tags, then the same tag is used to group similar posts. A group of people can become a hash tag trend Attract a special discussion to as many people as possible to participate in it.**

***Keywords*- Sentiment Analysis (SA), Natural language Processing, HSWN multilingual Senti Word Net, sentiment polarity detection.**

I. INTRODUCTION

The primary goal of this research is to analyze to capture. The capture sentiment of personal tweets in Hindi. A sentiment is associated with a given hash tag. Since Twitter uses the Latin script, Devanagari does not use the hash tag. We collect tweets for a specific hash tag. Average counts on all tweets with the tag of sentiment focused on analysis of the context containing the hash tag. They determine a sense of belonging.

The Millions of people are using social networking sites like Face book, Twitter, WhatsApp etc. to express their feelings, give thoughts and share our opinion about their daily lives. The Internet age has changed the way people open up, express ideas. This is mainly done through blog posts, social media, online forums, websites, etc. The field of emotion classification is an exciting new Research direction due to large number of real Hindi world Application where it is important to know public opinion in social media Sentence analysis is studied to make better decisions.

The Analyzes public opinion and feelings Services such as negative, positive or natural words [2]. It is important to know what other people think with him the rapid growth of user-generated data on the web is people need to understand online reviews, blogs, web sites to use social media and online analytics Generated data / review has been generated. Through of the user experience, you can get to know the merits and demerits of the result Shared by people on social media that can be useful to them, application developers can make decisions and improve automated analysis of your products or services online based on public opinion

A deep understanding of emotional analysis is required to draw an opinion. The capability of most of the existing models is considered unsatisfactory by the machine. A dataset is a dictionary in which opinion stores have positive or negative words with their polarity.

For example, "खुश" "उत्कृष्ट" "बुरा" "उबाऊ" and so on, these words are used as Hindi datasets [3]. Most user opinions of the current sentiment analysis model are available in the literature because many opinions are available as SentiWorldNet is very difficult, which can be the polarity of all words, which can be used.

II. POLARITY FRAME WORK

We have presented our framework in which we have interpreted the Hindi language from the collection, and analyzed the sentiment and categorized the opinion of Twitter WhatsApp. We considered the social things posted

© 2021 IJSRET

by the user currently as a hash tag to express our opinion about the political thought. We archived the retrieved social media tweets in the database, translated the Hindi tweets and created the dataset.

The model was tested and split on a training dataset to obtain accuracy results for each Polarity and personal things were computed using three different libraries with 600 social Hindi sentiment word use of tool SentiWorldNet. The training model HSWN Web and SVM classifiers built into the training dataset Visual Basic Dot Net Classifier. The following elaborates the tools and techniques that help us in Hindi emotional analysis.

Some Examples:

- **Subjective:** - अक्षय और करीना की यह फ़िल्म बहुत अच्छी है (There is an opinion in this sentence, it talks "अच्छी "About the feelings of the film and the writer and therefore it is subjective) .The analysis of emotions can be classified into three parts on the basis of expressed feelings.
- **Positive Sentence**: मोदी जी महान पुरुष है
- **Negative Sentence:** यहाँ के लोग बहुत खराब है
- **Natural Sentence:** मैं दोपहर में 2 बजे भोजन कर लेता हूं

1. Components of thought Mine:

1.1 Thought Holder: It is the holder of a special opinion. It can be a person. Or a group has a thought in terms of blogs and reviews Holders are the people who write these reviews or blogs.

1.2 Thought Objective: the object is an object to which the thought holder is expressing.

1.3 Thought Orientation: The orientation of an opinion on an object determines whether .The opinion of a thought holder about an object is positive, negative or neutral.

2. Layers of Thought Mine:

2.1 Sentence Layer: The sentence layer analyzes the documents at the sentence layer. The sentences are analyzed individually and are classified as positive, negative or neutral.

2.2 Document Layer: The document layer classifies the entire document as a single polarity Positive, negative or neutral.

2.3 Aspects Layer: Analyzes the aspect layer and identifies the aspect to go deeper Analyze and categorize features and features for a given document in one sentence accordingly, positive, negative or neutral.

III. HINDI LANGUAGE CHALLENGES

1. Order Words:

Plays an important role in identifying the word arrangement in a sentence. The subjective nature of the text Hindi is a free sequence language i.e. subject, object and Verbs can come in any order while English is the language of a certain order i.e. subject Followed by an action and followed by an object. The Word order plays an important role in decision making. The polarity of a text [5], with slight changes and changes in the text in the same set of words the word order affects the polarity aspect.

2 Morphological language words:

Handling the word modal is also a big deal Challenge for Hindi language. Meaning in Hindi language form where a lot of information is written in words compared to the Hindi language Add another word for additional information.

3. Handling spelling language words:

Similar words in Hindi language the meaning can be with different spellings, so it is quite complicated for everyone. It is enough to have such words in a word and even while training a model All spelling words can be complicated to handle

4. Lack of resources language words:

Lack of sufficient resources, equipment and annotated corpora also adds especially when we face challenges by addressing the problem of sentiment analysis working with Hindi languages.

5. Reference Resolution language words:

This is the problem of determining multiple expressions that refine the same thing for example "सीता ने खाना खाया ।वह सोने चले गई।" in the "वह" second sentence refers to सीता that is an entity. It is important to identify these Context Relationships for Aspect-Based Sentiment Analysis.

IV. RESEARCH OBJECTIVES

The basic objective of this research work is a model for meaning classification of Hindi word expression using various classification techniques.

The main objectives of the proposal model are given below-

- The proposed research on social media will classify public expression of Hindi on Twitter, Whatsapp, using a hybrid approach to the political and social domains.
- The Accuracy tool will be developed online for better accuracy and analysis.
- Research will be based on Hindi sentence level.
- The proposed research will provide better accuracy and efficient results than previous research.
- Statistical methodology and machine learning algorithms will be used to optimize previous research algorithms.
- It will detect the number of entries in the lexicon, database and corpus size.

© 2021 IJSRET

V. METHODOLOGY

1. Lexicon Method:
The Lexicon method uses sentimental dictionaries with words and matches them with data to determine polarity. It adds sentimentality to the sentiment of by saying how to make positive, negative and natural words contained in the dictionary [8].

2. Dictionary-Based:
It is based on the usage of commonly occurring words manually collected and annotated. The set grows by searching for synonyms and antonyms of a dictionary. An example of that dictionary is SentiWordNet, which is use to develop a thesaurus called HindiSentiwordnet.

3. Corpus-Based
The purpose of the corpus-based approach is to provide dictionaries related to a specific domain. These dictionaries originate from a set of opinion words that grow through the use of statistical techniques through the search for related words [9].

VI. PROPOSED ALGORITHM

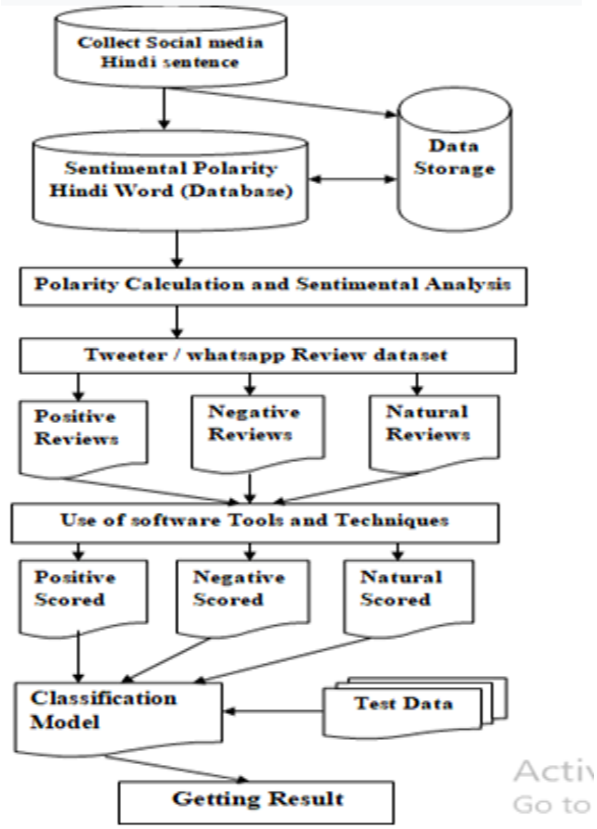

Fig 1. Framework of social media (twitter/whatsapp) classification.

VII. WORKING PROCEDURE

1. STEP-1:
(Input Phase): Input any text into the system. It breaks down into whole sentences because the proposed sentence classification model works at the sentence level.

2. STEP-2: (Text Preprocessing Phase):
To get high accurate results, some pre-processing action is applied to the given text or sentence (special characters, symbols, logos, question marks, URLs, smiles etc.) sentence. Which has no meaning and we cannot get the correct sentiment classification results by the presence of those noises.

3. STEP-3: (Parts of Speech Tagging Phase):
Not all words in a given sentence are useful for sentence classification, as some words capture emotion. Hence the Hindi sentimental words of the word and word group after parts of speech tagging are characteristic of their respective tags.

4. STEP-4: (Hybrid Approach Phase):
There are two models used under the Hybrid Approached they are:

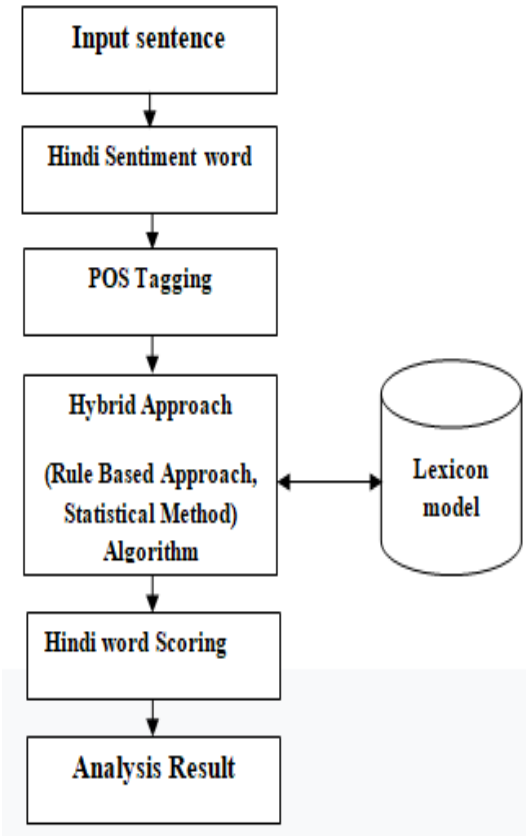

Fig 2. Working procedure Using Lexicon model.

4.1 Rule Base Model: This model includes a set of negative handling rules because polarity can be changed to

© 2021 IJSRET

positive in some situations and vice versa Negative handling is the biggest problem to perform Sentimental Hindi word classification.

4.2 Statistical-Based Model: To achieve accuracy, a database of all sentimental word with their statistical score is developed under this model. The positive Sentimental words assign a positive value (between 0 to 1) and negative to (-1 to 0).

5. STEP-5: (Result Representation Phase):
The system gets the value of every sentimental word present in the sentence and the sum will be calculated. If the count sum is positive, the sentence is positive otherwise negative. The Sentence in neutral if sum means zero.

VIII. EXPECTED OUTCOME AND SOCIAL IMPORTANCE OF THE RESEARCH WORK

All words on Twitter / WhatsApp have equal possibilities to be negative and positive. The senses are the overall emotion of each word Score of zero. Twitter/WhatsApp net worth becomes zero due to equal number of positives and negatives Words. No words are given polarity due to the fact that scores do not appear in HSWN.

We will use a set of machine learning techniques to classify Hindi sentences as positive, negative or neutral based on polarity. The technique also gives good enough results and better performance with maximum entropy and index.

The research vector can be increases to improve the feature vector related sentence recognition process. We are using HindiSentiWordNet in research. This can give a better view of the content in a better way which will be helpful for the users.

As the Hindi user twitter is growing in whatsapp (social media) areas, our project aims to efficiently divide Hindi texts to provide better performance to the user.

REFERENCE

[1] Farhadloo M. et al. (2013),” Multi Class Sentiment Analysis Clustering and Score Representation”, IEEE 13th International Conference on Data Mining Workshops, pp. 904 to 912.
[2] Nongmeikapam, et al., (2014), "Verb Based Manipuri Aanalysis”, International Journal on Natural Language Computing (IJNLC) Vol. 3, No.3, pp. 113-119.
[3] Balamurali A., et al. (2012), “Cross-Lingual Sentiment Analysis for Indian Languages using Linked WordNets”, COLING 2012, pp. 73-8.
[4] K. M. Anil Kumar et al., (2015), “Analysis of Users’ Sentiments from Kannada Web Documents”, Eleventh International Conference Communication Networks, vol. 54, pp. 247-256.
[5] Arora P., et al., (2013), Sentiment Analysis For Hindi Language, MS Thesis IIITH.
[6] Mittal N., et al., (2013), Discourse based sentiment analysis for Hindi reviews. In PReMI, pages 720–725.
[7] Arora P., et al., (2012), Hindi subjective lexicon generation using wordnet graph traversal. In Journal Proceedings of CICLing, New Delhi, India, March.
[8] Kastrati, Z., et al., (2019), Integrating word embeddings and document topics with deep learning in a video classification framework. Pattern Recognition Letters, 128, pp.85-92.
[9] Devlin, J.,et al.,(2018), Bert: Pre-training of deep bidirectional transformers for language understanding, preprint:1810.04805.
[10] Loshchilov, I. and Hutter. (2017), Decoupled weight decay regularization. Preprint arXiv: 1711.05101.
[11] Parikh, R. and Movassate, M., (2009), Sentiment analysis of user-generated twitter updates using various classification techniques. CS224N Final Report, 118.

© 2021 IJSRET