ESP IJACT
ESP International Journal of Advancements in Computational Technology
Volume 1 / Issue 1 Paper Id: IJACT-V1I1P101 / Page: 1-4

Original Article

Sentiment Analysis of Product Reviews Containing Hindi Text

Vandana Yadav[1], Dr.Parul Verma[2], Dr.Vinodini Katiyar[3], Dr.Namrata Dhanda[4]

[1]Research Scholar, Amity Institute of Information Technology Amity University Uttar Pradesh, Lucknow, India.
[2]Assistan Professor, Amity University Uttar Pradesh, Lucknow, India.
[3]Professor, Department of Information Technology, Dr. Shakuntala Mishra University, Lucknow, U.P, India.
[4]Professor, Department of Computer Science & Engg. Amity University Uttar Pradesh, Lucknow, India.

Received Date: 05 March 2023 Revised Date: 18 March 2023 Accepted Date:01 April 2023

Abstract: *Sentimental analysis has gained popularity in recent years. Gathering massive volumes of data from the sources, utilizing the appropriate techniques or algorithms, and categorizing them are the key challenges in an emotional analysis. In the ever-expanding internet world of today, social media provides a platform for people to express their emotions. Along with changes in how things are done in all facets of our everyday lives, so have the ways in which people communicate their views or beliefs. Usually, people will speak in their own language or in a way that feels natural to them. The decision-making process depends on these specific reviews. The large amount of information gathered on social media is useless if opinions are not sorted in accordance with their thoughts. This study's data demonstrates whether a customer's reviews are positive, negative, or neutral.For this model, the customer reviews are first scraped from Amazon, after which they are given various sentiment evaluations using polyglot, and finally they are categorized as positive, negative, or neutral using text classification model.*

Keywords: *Social Media, Sentiment Analysis (SA),Text-based Classification, Polyglot*

I. INTRODUCTION

Sentimental analysis (SA) is the process of identifying a communicator's propensity or attitude by analyzing the contextual polarity of their writing or speech. The introduction of web 3.0 allowed users to conduct a conversation about the product or service with the manufacturer or service provider, changing the way people express their opinions. The majority of it is done online through blog posts, debates, item survey sites, internet-based activities, and so forth. Through social networking platforms like Google Plus, Twitter, Facebook, and others, users express their opinions, feelings, and mood. A significant amount of concept-rich information is being produced through social networks in the form of tweets, reviews, comments, discussions, blog posts, and so forth. Social media networks provide businesses a chance by providing a platform for interaction with their target audience for advertising. The majority of the time, a user's judgment regarding a product that is available online is heavily influenced by user-generated material. Regular users find it challenging to assess the vast amount of content that is generated by users on a regular basis. Therefore, there is a strong need to automate user reviews. SA has a very important role to play in this situation. Sentiment analysis helps customers decide whether the information about a product is appealing or not before they buy it. Advertisers and businesses use this information to better understand their goods and services so that they may cater to the needs of the user. So one can employ textual information retrieval techniques in the context of analysis. It mostly focuses on finding, gathering, or evaluating the actual information that is already present. Other textual content, however, may reflect subjective characteristics. This data focuses mostly on sentiments, attitudes, views, feelings, and assessments, which may constitute the core of SA.

II.SENTIMENT ANALYSIS

Making human language machine understandable using a process called natural language processing, which then applies various operations to it in order to extract meaningful information NLP is a branch of AI that facilitates communication between computers and people using language.

Many different Python libraries are available that can assist us in carrying out NLP tasks. Every library has a few distinct characteristics that set them apart from one another. Common NLP library functions include tokenization, stemming, lamentation, spell checking, etc.

The open-source Python package Polyglot is used to carry out various NLP operations. It is quick because it is built on NumPy. It distinguishes itself from the competition thanks to its extensive selection of specialized commands. It can be used for languages that do not support spacy and is comparable to spacy.

There have been several types of research carried out in this area. Research in the area of NLP and Hindi Language has focused mainly on Text Classification using Social media.

III. LITERATURE REVIEW

By contrasting the linguistic meaning of a word in a sentence with the context developed by Hindi WorldNet using a similarity-based approach that only works for the noun, **Prof. Bhattacharya** proposed the Hindi Word Sense disambiguation (WSD) method in 2008, addressing the issue of ambiguity resolution in Hindi for the first time (Bhattacharya et al. 2008).

Joshi et al. (2010) made an early effort while working on Hindi Language Sentiment Analysis. A backup plan was proposed by the researchers, using three techniques, including sentiment analysis in the In-language. Sentiment Analysis developed his own HSWN: Hindi SentiWordNet lexical tool with an emphasis on resources and machine translation. Researchers were able to reach an accuracy of 78.14 percent thanks to their work.

In another research sense-based features of WorldNet were given by **Balamurali et al. (2012,** they alsocarried out SA experiments for quality checks through travel review database.

A translation-based technique and bi-lingual dictionary were presented by **Bakliwal et al. (2012)** to establish Hindi Subjective lexicon. The authors used this lexicon to classify feelings and were able to get 79% of accuracy.
Mittal et al. (2013) published Sentiment Analysis on films and treated relations and negation of debate and attained 80.21% accuracy.

Arora(2013) used n-gram approaches as well as subjective lexicon to perform SA on various blogs and reviews written in Hindi for the films and related products.

Bansal et al. (2013) made use of deep learning methods for performing Sentiment Analysis on films and got 64% accuracy.

Bhattacharyya and Sharma (2014)suggested a method of bootstrap in all speech parts that is adjectives, nouns, to expand HSWN by adding the HindiWordNet existing. The researchers used a lexicon to test to Sentiment Analysis model over the field of movie and product reviews, achieving 87% accuracy.

Jha et al. (2015) came up with an opinion mining method for Hindi as well as unmonitored approach of POS tagging for performing Sentiment Analysis.

Sharma et al. 2015; Pandey and Govilkar2015 applied an unsupervised approach based on lexicon for performing Sentiment Analysis using HSWN and categorized the sentences among the neutral, negative, and positive category.They have treated the relationship of negations and debate.

Seshadri et al. (2016) published Sentiment Analysis for the tweets written in Hindi on dataset of SAIL-2015: “Sentiment Analysis Indian Language”.

Akhtar et al. (2016b) conducted Sentiment Analysis based on aspects for Hindi Language and created an annotated dataset for reviews on products. Researchers applied an SVM classifier for 54.05% accuracy. The researchers also experimented with 3-classifiers namely; SMO, DT, NB in MEKA: a “Multi-label/Multi-target Extension to WEKA” in mainly 4-different domains such as “Electronics”, “mobile apps”, “travels” and “movies” and concluded that performance of NB was better in mobile apps and electronics domain, whereas the decision tree technique reported being better for domains of movies and travel.

Akhtar et al. (2016a,c) first try to use a deep learning system such as CNN for working with Sentiment Analysis in Hindi and carried out the analysis on both sentence and aspects level.

IV. PROPOSED METHOD

The steps in the suggested procedure are as follows:

- Gathering feedback or reviews;
- Sentimental Classification; and
- Polarity Detection Using Polyglot API.

The aforementioned idea has been implemented in Python, and the source code for the aforementioned method is provided below;

- Using pip install polyglot, we will install Polyglot just like any other Python package
- **Importing Necessary Libraries**

We'll import polyglot and investigate its various features. As and when necessary, all functionalities will be imported.

- **Executing an Operation on the Data**

Let's initialize some text first, which we will use for doing various operations before performing the on our data.**detect = Detector(init) print(detect.language)**

- **Executing Sentiment Analysis**

It is used to determine the text's polarity.!polyglot download sentiment2.hi from polyglot.text import Text.

V. RESULTS AND DISCUSSION

For the findings, the researcher used some actual examples of product reviews that were found on an e-commerce website. The likelihood of a remark being both positive and negative is depicted here;hindi_text1 = "हालांकि इसके अधिकतर भाग सही स्टैंडर्ड रूप में दिखते हैं, फिर भी पूरे एप्प में साइड स्वाइप्स, एक्सेस और कार्यशैली का एक अन्य स्वरूप बढ़ा देते हैं।"

- # Sentiment Analysis using Polyglot

```
text1 = Text(hindi_text1)
sentiment = text1.polarity
print("Sentiment Polarity: ", sentiment)
if sentiment < -1.0:
print("Negative")
elif sentiment > 1.0:
print("Positive")
else:
print("Neutral")
```

- Sentiment Polarity=1.0 , Positive

The available result in this case is Polarity. It can range from -1 to +1 in value.

In this, 0 denotes a neutral statement, -1 is a severely negative statement, and +1 is a highly positive statement.

The sentiment polarity score for the aforementioned result is 1.0.

Since polarity = 1.0, it is obvious that the above-mentioned statement is positive in nature.

Consider one more example:

hindi_text2 = "इसपॉपअपबॉक्समेंपासवर्डचेंजकरने,
लॉगइनअलर्ट्सचालूकरनेऔरमौजूदाफेसबुकसेशनकोपूरीतरहसेबंदकरनेकेऑप्शनमिलेंगे।"

- Sentiment Polarity=0.0,Neutral

Consider one more example:

hindi_text3 = "बैंडविथकॉस्टकोकमकरनाशायदउसकाकारणहोसकताहै।"

- Sentiment Polarity=-1.0,Negative

VI. ADVANTAGES

The advantages of using the polyglot library for sentiment analysis of Hindi text are:

Multilingual support: Polyglot provides support for multiple languages, including Hindi, which makes it possible to perform sentiment analysis on texts written in Hindi.

Text embedding: Polyglot provides text embeddings, which are vector representations of words, which can be useful in sentiment analysis as they capture semantic relationships between words.

Natural Language Processing (NLP) support: NLP activities like sentiment analysis, part-of-speech tagging, and named entity recognition—all of which are essential for sentiment analysis of Hindi text—are supported by the Polyglot library.

Integration with NumPy: Polyglot can be integrated with NumPy, a numerical computing library for Python, which makes it easier to perform mathematical operations on the results generated by Polyglot.

Open-source: Polyglot is an open-source library, which means that it is freely available to use, modify, and distribute, making it accessible to a wide range of users.

VII. CHALLENGES

The challenges using polyglot library are:

Resource intensive: polyglot library requires large amounts of memory and computational resources, which can make it difficult to run on low-end hardware.

Difficulty in installation: polyglot requires a lot of dependencies, and it can be challenging to set up and install on some systems.

Compatibility issues: polyglot may not work well with certain software, libraries, or programming languages, and compatibility issues may arise.

Limited support for some languages: polyglot may not support all the languages a user may require.

Lack of updates: polyglot library may not receive regular updates, which can result in bugs or compatibility issues.

High error rate: polyglot's performance may not be as accurate as desired, especially for low-resource languages.

VIII.CONCLUSION AND FUTURE WORK

Sentiment analysis of Hindi text is a challenging task due to the complex nature of the language, which includes different scripts, dialects, and sentiment polarities. However, with the help of Python libraries like polyglot and others, sentiment analysis of Hindi text has become easier and more accurate. These libraries provide support for NLP tasks, text embeddings, and integration with numerical computing libraries, which makes sentiment analysis of Hindi text more efficient and accurate.

In the Future work, with the development of machine learning and deep learning methodologies, researchers can seek to increase the accuracy of sentiment analysis of Hindi text by utilizing more sophisticated algorithms and models.Currently, sentiment analysis of Hindi text is well supported, but sentiment analysis of other Indian languages such as Bengali, Marathi, and Tamil still requires improvement.

IX. REFERENCES

[1] "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," A. Pak and P. Paroubek (2010). Pages 1320–1326 of the Proceedings of the Seventh Conference on International Language Resources and Evaluation.

[2] Social sciences and business intelligence in South Africa, Badenhorst, B. K. and Fitzgerald, W. (2012), "The application of text and sentiment analysis in market research." Available at: http://www.samra.co.za/wp-content/uploads/2013/05/Badenhorst Fitzgerald Text-andsentiment-analysis-Expert-QAMaterials.pdf. Southern African Marketing Research Association.

[3] Sentiment analysis and opinion mining are topics covered by B. Liu in Synthesis lectures on human language technologies, volume 5, Issue 1, pages 1-167.

[4] International Conferences Web Based Communities and Social Media (IADIS WBCSM 2012), "Towards Integrating Online Social Networks And Business Intelligence," Costa, P. R. S. and Souza, F. F. (2012). Available at: http://homepages.dcc.ufmg.br/fabricio/download/iadis-wbc2012.pdf.

[5] The first international workshop on ontology-supported business intelligence, Funk, A., and Bontcheva, K. (2008), "Opinion Analysis for Business Intelligence Applications," Proceedings, Article No. 3, available at: https://gate.ac.uk/sale/iswc08/obimusing/musing-om-iswc08.pdf.

[6] Po-Wei Liang and Bi-Ru Dai, "Opinion Mining on Social Media Data", IEEE 14th International Conference on Mobile Data Management, Milan, Italy, June 3–6, 2013, pp. 91–96, ISBN: 978-1-494673–6068-5, http://doi.ieeecomputersociety.org/10.1109/MDM.2013

[7] John K.Victor, Ilostanely Uzochukwu, Dr.N.Egu, 2021. "Twitter Sentimental Analysis" ESP Journal of Engineering & Technology Advancements 1(1): 1-5.

[8] "Profitability Analysis of Indian Information Technology Companies Using DuPont Model", Asian Journal of Management, 2019, Volume 9, Issue 3, pp. 1105–1108. Praveen Gujjar and T. Manjunatha.

[9] Praveen Gujjar and Prasanna Kumar H R, "Sentimental analysis for flowing text in Email Conversation," International Journal of Computer Science and Engineering (IJCSE), Volume 9, Issue 4, pp. 67–68 (2020).

[10] T. Manjunatha and Praveen Gujjar, "Performance Analysis of Indian Information Technology Companies Using DuPont Model", IUP Journal of Management Research, Volume 17, Issue 4, pp. 7–14, 2018.

[11] "Analytical mapping of opinion mining and sentiment analysis research between 2000-2015," Information Processing &Management, vol. 53, no. 1, pp. 122–150. R. Piryani, D. Madhavi, and V. K. Singh.