A.V.COLLEGE OF ENGINEERING
DEPARTMENT OF COMPUTER APPLICATIONS

"CAS NEWS LETTER"

Volume: 17 **Month: Oct'25** **Issue: 1**

A Natural User interface (NUI) is a system for human-computer interaction that the user operates through intuitive actions related to natural, everyday human behavior. ANU I may be operated in a number of different ways, depending on the purpose and user requirements.

This is the emerging field in computer science. I encourage all the students to know about this kind of new technologies. This article is very much useful to about the methodologies involved in NUI

WISH YOU ALL SUCCESS...!!

DR.S.SELVAMUTHUKUMARAN

LEARER-WRITER

J.JAWAHIRA SANOFAR-I MCA

S.SHIFFAYA FATHIMA-I-MCA

SENTIMENTAL ANALYSIS ON CUSTOMER REVIEW

INTRODUCTION

Sentiment analysis is the process of identifying and extracting the emotional tone of a block of text to determine if the underlying sentiment is positive ,negative , or neutral.it is a natural language processing(NLP) technique that uses computational linguistics and machine learning to analyze text from sources like social media , customer review, and emails to understand public opinion and improve brand reputation or customer services

It moves beyond simple keyword counting to understand the underlying attitude—positive, negative, or neutral—expressed in customer feedback.

HOW SENTIMENTAL ANALYSIS WORKS

1. Data Collection & Pre-Processing

Gathering text data from various sources (reviews, social media) and cleaning it (removing noise, tokenization).

2. Feature Extraction

Identifying relevant features like individual words, phrases, and punctuation that carry emotional weight.

3. Sentiment Scoring/Classification

Using models (lexicons or algorithms) to assign a polarity score (positive, negative, neutral) to the text.

4. Visualization & Reporting

Presenting the results through charts and dashboards for easy interpretation and decision-making.

KEY TECHNIQUES DRIVING SENTIMENT ANALYSIS

Sentiment models rely on different foundational approaches to accurately classify text data:

Rule-Based Approach

In rule-based sentiment analysis algorithms, the system automatically tags input data based on a set of predefined rules to identify the polarity of user sentiments. To execute these rules, NLP techniques are leveraged. These techniques include stemming, part-of-speech tagging, parsing, lexicons, and tokenization.

An example of a rule-based sentiment analysis model can be seen in the following scenario. Lists of words are applied to different sentiments. Negative words can be "horrible", "worst", "bad" and positive words can be "great", "love" "amazing", for example. Next, an input sentence is analyzed where the count of polarized words are stored. This count is then processed into buckets of different classes, such as "positive", "negative", or "neutral", based on the highest count of polarized words associated with the given sentiments.

Rule-based algorithms are simple and easy to implement, however, they often overlook the complexities of text and word combinations. Rule-based or automatic algorithms can be used, depending on how advanced you want your sentiment analysis model to be.

Pros: Highly transparent, easy to interpret.
Cons: Time-consuming to build, struggles with context and irony

Lexicon-Based Approach

Lexicon-based sentiment analysis is a technique that utilizes a sentiment lexicon, which is a collection of words and phrases tagged with sentiment values. These values indicate the polarity (positive, negative, or neutral) and intensity (strength) of the sentiment associated with each word. The primary goal of this approach is to aggregate the sentiment scores of individual words to determine the overall sentiment of the text.

Components of a Sentiment Lexicon

1. Sentiment Words: Words that carry emotional weight, such as "happy," "sad," "love," and "hate."
2. Polarity Scores: Numerical values indicating whether a word is positive (eg, +1), negative (eg, -1), or neutral (eg, 0).
3. Intensity Scores: Values representing the strength of the sentiment, often on a scale (eg, from -5 to +5).

How It Works

The process of lexicon-based sentiment analysis typically involves the following steps:

1. Tokenization: Breaking down the text into individual words or phrases.
2. Normalization: Converting words to their base forms (eg, "running" to "run") and handling negations (eg, "not happy" to "unhappy").
3. Lexicon Lookup: Matching each word in the text to the sentiment lexicon to retrieve its sentiment score.
4. Aggregation: Summing or averaging the sentiment scores to determine the overall sentiment of the text.

Pros: Fast and language-independent (with appropriate lexicon).
Cons: Less accurate with slang or domain-specific terminology

Machine Learning (ML)

Sentiment analysis using machine learning is a technique in which data is automatically analyzed to determine if customer interactions are positive, negative, or neutral. Sentiment analysis machine learning uses natural language processing to design the core building blocks of the data analysis process. Understanding how to do sentiment analysis is crucial for companies today. Companies set up a unique strategy for sentiment analysis, starting with

data collection. Following data collection, a text cleaning occurs. For example, among all the text data captured, punctuation and stop words are not helpful in the sentiment analysis process, so they are removed during the cleaning process. The goal is to reduce words to their step before moving on to the analysis process. Following text cleaning, the pre-set algorithms can analyze the data to determine its sentiment or emotional meaning. At the end of the sentiment analysis process, companies can view the data and better understand customer interactions.

Machine learning in sentiment analysis is necessary due to the vast amount of data collected through texts, phone calls, reviews, and other methods. A comprehensive analysis of the data collected is almost impossible if attempted through manual processes. Machine learning adapts to the needs of individual users and is a scalable solution for data analysis. Sentiment analysis is vital for companies to elevate their sales and marketing efforts, adjust social media strategies and strengthen a consistent brand message. In addition, sentiment analysis can help companies conduct market research to gain an understanding of competitors

Pros: Highest accuracy, handles nuance and sarcasm well.
Cons: Requires large training data, results can be difficult to explain (black box).

CONFUSION MATRIX IN SENTIMANTAL ANALYSIS

A confusion matrix is a fundamental tool for evaluating the performance of classification models, particularly in the context of sentiment analysis. It provides a comprehensive overview of how well a model's predictions align with the actual outcomes, helping practitioners understand the strengths and weaknesses of their models.

The structure of confusion matrix:

	Prediction Positive	Predictive Negative
Actual Postive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

True Positive (TP): The model correctly predicts a positive sentiment.
True Negative (**TN**): The model correctly predicts a negative statement
False Positive (FP): The model incorrectly predicts a positive sentiment
False Negative (FN*):* The model incorrectly predicts a negative sentiment
Example : 70 reviews were actually positive, and the model predicted 65 as positive (TP) and 5 as negative (FN).

STEPS INVOLVED IN SENTIMENTAL ANALYSIS

1.Tokenization

The process of breaking down text into individual words or tokens. This is a crucial step as it allows the model to analyze each word separately.
For example, the sentence "I love this product!" would be tokenized into ["I", "love", "this", "product", "!"].

2.Cleaning the data

This step involves removing noise and irrelevant information from the text data. This can include:

- Removing punctuation and special characters
- Removing URLs and hyperlinks
- Removing HTML tags and entities
- Handling out-of-vocabulary (OOV) words

3.Removing the stop words

Stop words are common words like "the", "and", "a", etc. that don't add much value to the sentiment of the text. Removing these words helps reduce the dimensionality of the data and improves model performance.
Example: Removing stop words from "I love this product" → ["love", "product"]

4.classification

This is the core step in Sentiment Analysis, where the text is classified as positive, negative, or neutral. The classification can be done using various techniques, including:

- Rule-based approaches
- Machine learning algorithms (supervised or unsupervised)
- Deep learning models (e.g., CNN, LSTM, BERT)

Example: Classifying "I love this product" as positive sentiment

5.Apply supervised algorithm for classification

In this step, a supervised machine learning algorithm is trained on labeled data to learn the patterns and relationships between the text features and sentiment labels. Some popular algorithms for Sentiment Analysis include:

- Naive Bayes
- Support Vector Machines (SVM)
- Random Forest
- Gradient Boosting

6.Calculation

This step involves computing sentiment scores or probabilities to determine the sentiment intensity.

Example: Calculating a sentiment score of 0.8 for "I love this product" (0 = negative, 1 = positive)

The Challenges: Limitations and Biases

Sarcasm

people tend to use sarcasm as a way of expressing their negative sentiment, but the words used can be positive(eg:”I am so glad that the product arrived in one piece!”). in such cases, sentiment analysis tool can classify the feedback as positive which is reality is negative

Ambiguity & Negation

Ambiguity and negation are major challenges in sentiment analysis because they complicate the interpretation of a text true sentiment Ambiguity arises from issues like sarcasm, irony, idioms and context-dependent word meaning, while negation challenges occur when words like “not” reserve the polarity of a sentence, making it difficult for algorithm to accurately classify sentiment without understanding the full context and scope of negation

Data Quality

It include inaccuracies, incompleteness and inconsistencies in the data which can be caused by unstructured or irrelevant text, lack of content, sarcasm and noise. These issues can lead to unreliable model and por decisions, making it crucial to address them through data cleaning and other quality assurance processes

Handling Imbalanced Datasets:

- Sentiment analysis datasets can be imbalanced, with more positive or negative samples than neutral ones.

Example: A dataset with 80% positive reviews and 20% negative reviews may not provide enough examples of negative sentiment

Intensity and Degree:

- Understanding the intensity or degree of sentiment, such as "very good" vs. "good," can be challenging.

Example: "This product is amazing" (vs. "This product is good").

Rush your ideas to

P. JAYAPAL MCA.,M.Phil,

AssistantProfessor/CA Editor – CAS

Mail-Id:editor.cas@gmail.com,

pjavcce@gmail.com