A Study on NLP Approaches for Emotion and Sentiment Interpretation from Text

Manvi Godbole; Shweta Bobate; Shruti Tijare
Department of MCA , G H Raisoni College of Engineering and Management
Nagpur, India

Abstract
As internet communication has gone through a boom with social media, humans have been more and more using text to express their emotions, so it has become crucial to come up with ways to accurately capture these emotions beyond mere positive or negative sentiment analysis; this research proposes a novel method employing deep learning and semantic text analysis (DLSTA) to detect precise emotions such as joy, sadness, anger, fear, disgust, and surprise in written language with high accuracy by taking advantage of word embeddings to capture the meaning of words and their context better than other methods available. The paper discusses the challenges of processing large-scale unstructured internet data and emphasizes the need for advanced NLP techniques to improve emotion detection accuracy.

Keywords: Sentiment Analysis, Emotion Detection, Natural Language Processing (NLP), Text Mining, Artificial Intelligence (AI) in text analysis

1. Introduction

The quick improvement of the web and social media has made an blast of content information, communicating client sentiments and opinions. Two unmistakable NLP methods sentiment analysis and emotion detection are utilized to get it human feelings from content. sentiment analysis names content as positive, negative, or impartial, though emotion discovery recognizes certain feelings such as delight, pity, outrage, fear, appal, and surprise.

But text-based feeling discovery is prevented by dialect dubiousness, contextuality, and the need of non-verbal signals. Lexicon-based strategies in the early days have created into profound learning models, utilizing semantic content investigation and word embeddings to accomplish higher precision. These procedures are connected broadly in healthcare, promoting, instruction, and social media analytics to screen opinions, degree mental wellbeing, and upgrade client engagement.

This article talks about distinctive opinion and feeling acknowledgment strategies, their disadvantages, and later improvements in NLP and profound learning. It moreover distinguishes real-world applications and conceivable improvements with advanced AI models [1],[2].

2. Background Study

2.1 Sentiments:

Nowadays, individuals from every part of the world extensively use social media platforms such as Instagram, Twitter and Facebook to exchange their opinions and views. Social media has now turned out to be one of the strongest methods of communication. Due to this reason, large quantities of data, referred to as big data, are generated on a daily basis. For this purpose, interpretation of such data effectively, sentiment analysis came into existence for helping to analyze it effectively. It is more significant than ever to understand user opinions, for businesses and organizations. Sentiment analysis, or opinion mining, is a technique utilized to ascertain used to positive, negative or neutral opinion

DOI: https://doi.org/10.5281/zenodo.15519708

of an individual. It operates through analyzing text using natural language processing (NLP) methods in order to determine the writer's attitude. Whereas sentiment analysis is primarily employed to label text as positive or negative, it doesn't always remain there. Opinions can be classified differently as well, for example, as agreeing or disagreeing, or graded on a scale of five steps: strongly disagree, disagree, neutral, agree.

2.2 Emotions:

Emotions are a crucial part of human life, impacting decision-making and enhancing communication. Emotion detection, also known as emotion recognition, is the process of identifying human emotions such as joy, sadness, and anger. Researchers have been actively working on automating emotion recognition, but detecting emotions from text remains challenging due to language ambiguities, evolving slang, and newly introduced terminologies. Although physical signals like heart rate, hand shaking, sweating, and voice tone also convey emotional states (Kratzwald et al., 2018), text-based emotion detection is much more complicated. In contrast to simple psychological states like happiness, sadness, or anger, emotion detection models tend to categorize emotions on a six-point or eight-point scale, depending on the used framework.

The term 'emotion' was derived in the seventeenth century from the French word emotion, which referred to a physical disturbance. Prior to the nineteenth century, mental states like passion, appetite, and affections were not regarded as emotions. In the nineteenth century, 'emotion' was officially established as a psychological term (Dixon, 2012). Emotions are currently classified into two categories of psychological models: dimensional models and categorical models [1],[3].

3. Methodology:

In this procedure, we are categorizing the input content under various feelings by finding the passionate substance from the provided English content. Passionate substance are verbs, adverbs, adjectives, expressions or any combination of these catchphrases.

For instance, "We are going on get-away. I'm exceptionally energized". The keyword "energized" symbolizes "bliss" or "delight", with such watch words feelings can be categorized. The source of input to the framework is content substance from social organizing destinations like product audits, comments, individual blogs, feedbacks etc. To begin with step is to distinguish the structure of content to be able to determine the calculation utilized for feeling classification. In this procedure, the structure is distinguished as follows [3]:

- Each content is a list of sentences
- Each sentence is a list of tokens
- Each token is a tuple of three components: a word form (the correct word that showed up in the content), a word lemma (a generalized adaptation of the word), and a list of related tags

For case, “We are going on a trip. I’m feeling exceptionally excited.” This input is structured as list of sentences as: [{We are going on a trip.}, {I’m feeling exceptionally energized.}]

Directly, each sentence in list is organized into list of tokens as: {‘We’, ‘are’, ‘going’, ‘on’, ‘a’, ‘trip’, ‘.’}, {‘I’,’’m’, 'feeling', 'exceptionally', 'excited', '.'

All these tokens are depicted help as triple of qualities, i.e., tuple. For event, token "energized" has its qualities as ('excited', 'excite', 'VB'.)

3.1 Content Processing:

Content Handling Preprocessing is done on the substance a few time as of late applying the calculations on the input. It changes over unrefined input into a more organized

DOI: https://doi.org/10.5281/zenodo.15519708

shape, making it less complex and more beneficial to get ready.

There are a few preprocessing methods utilized: Cleaning: Deals with complement, end words, duplicate letters, capitalization, etc. Comment: Names tokens with Part-of-Speech (POS) labels. Normalization: Standardizes the input for basic openness. Highlight Extraction: Extricates and finds noteworthy highlights relating to a particular application or task.

3.1.1 Removing Punctuation:

Removing Punctuation in order to derive useful keywords from the provided input for processing, punctuation characters have to be discarded since they are irrelevant tokens in the input form. A simple method is to divide the input into words by whitespace and subsequently use string translation to remove all.

3.1.2 Rehashed Character:

Now-a-days individuals on social media do not entirely take after language structure. Their feelings are communicated through words like "ohhhh," "wowww," "cooool," etc. They will type in things such as “I likeee it” in arrange to emphasize the word “like”. In any case, computers don’t get it that “likeee” is one of the variation of “like” so they must be told. This strategy kills pointless rehashed characters to deliver appropriately designed, significant words found in the English dictionary.

3.1.3 ReplacementForNegativeExpressions

Among other contractions, text frequently uses won't for will not, can't for cannot, I'm for I am, I'll for I will, and that's for that is. Usually, preprocessing enlarges these contractions to increase uniformity and clarity.

3.1.4 Stop Phrases:

Preprocessing is mostly used to eliminate terms that are superfluous and do not contribute to applications, such as search engine searches. Stop words are unnecessary words that increase processing time and take up more database space. Depending on the system, the list of stop words changes. English stop words that are frequently used include: a, an, the etc.

3.1.5 Stemming:

In the English dialect, numerous words exist in distinctive shapes, such as things, verbs, and descriptive words. For illustration, the

word "pull in" has different shapes: "pulls in" (thing), "pulling in" (verb), and "alluring" (descriptive word). All these words share the common root "draw in" by expelling additions like "-s", "-ing", and "-ive". Putting away all these varieties in a database is wasteful and expends pointless memory.

Stemming makes a difference address this issue by expelling prefixes and postfixes to change over words into their root frame, known as the stem. In any case, the coming about stemmed word may not continuously be a appropriate word in the English word reference. The primary advantage of stemming is that it decreases database estimate whereas progressing recovery accuracy.

For example:

"exciting," "energize," "energized," and "energizes" are all stemmed to "excit", which is not a important word in standard English.

3.1.6 Lemmatization:

Lemmatization is comparable to stemming, but instep of creating a stem, it changes over a word into its legitimate root shape, known as the lemma. The key distinction between stemming and lemmatization is that lemmatization considers the morphological structure of a word, guaranteeing the result is a real word found in the dictionary.

Lemmatization is more exact but moreover slower than stemming since it requires a phonetic investigation of the word's root shape. To decide the redress lemma, part-of-speech (POS) labeling must be performed in advance, as words

DOI: https://doi.org/10.5281/zenodo.15519708

can exist in diverse shapes such as things, descriptive words, verbs, and adverbs
For example:
"exciting," "energize," "energized," and "energizes" are lemmatized to "energize" (in verb shape), protecting the rectify lexicon word.

3.2.Creatingemotion-Based,Word Dictionaries In YAML:
The development of a data dictionary, which is a YAML file with a list of words grouped and labeled with the respective emotion tags, is the second step. Six different dictionaries were developed, each for one of Ekman's six primary emotions: surprise.yml, disgust.yml, anger.yml, fear.yml, happy.yml, and sad.yml (remark: anger was copied in your text). Example: joyful.yml excite: [content] enjoyable: [happy] pleasant: [content] Sad.yml Sadness: [sad] gloom: [depressing] Disappointed: [depressed] Enhanced classification in emotion tests is enabled by the aid of each dictionary in word-to-emotion category mapping.

3.3 Tokenization and POS Tagging:
During this step, the input content is partitioned into littler units known as tokens. Based on the structure characterized over, each word is considered a token when a sentence is tokenized, and each sentence is considered a token when a passage is tokenized.
Tagging is the errand of relegating parts of discourse to words. NLTK offers the word_tokenize() work for tokenization, and the pos_tag() work for tagging.
Example of POS-tagged and tokenized text: tagged_token: [[('We', 'We', ['PRP']), ('go', 'going', ['VBG']), ('vacation', 'vacation', ['NN']), ('excite', 'excited', ['VBN'])]]
Common POS Tags
The taking after are a few of the commonly utilized Parts of Discourse (POS) tags:
1. NN (Thing): Apple, Orange, Bat, Taj Mahal.
2. NNP (Legitimate Thing): India, John, Amazon.
3. PRP (Pronoun): He, She, It, They, I.
4. JJ (Descriptive_word): Extraordinary, Best, Beautiful.
5. RB (Verb modifier): Gradually, Continuously, Very.
6. WP (Wh-Pronoun): What, Which, When, Who.
7. CC (Conjunction): And, Or, But, Either, Since.

3.4 Tagging Words from the Dictionary:
In this step, emotional words are marked in the text and tagged based on a pre-determined dictionary using tags like happy, sad, fear, anger, disgust, or surprise.
Sentiment tagging example:
sentiment_tag: [[('We', 'We', ['PRP']), ('go', 'going', ['VBG']), ('vacation', 'vacation', ['NN']), ('excite', 'excited', ['happy', 'VBN'])]]
Upon tagging, non-emotional content is removed by separation rules:
Rule 1 (Before "But") – Ignore text before "but," as the text after it is a reflection of the dominant emotion.

3.5 Sentiment Score Calculation:
The sentiment measure is calculated by analyzing the frequency of emotion tags such as happiness, sadness, fear, anger, disgust, and surprise. This method follows a simple, naive approach to sentiment analysis, where each occurrence of these emotion tags contributes to the overall sentiment score. However, this approach does not consider the context or intensity of emotions, making it less accurate for complex textual data. More advanced techniques, such as machine learning and natural language processing (NLP), provide a deeper understanding of sentiment by analyzing word associations, tone, and contextual meaning.

3.6 Sentimental Flow:
The emotional score, already generated with a simple method, can be rendered even more precise by incorporating two additional data dictionary files. The dictionaries refine the strength of emotions, making the sentiment more descriptive and closer to Ekman's six universal emotions happiness, sadness, anger, fear, disgust, and surprise instead of being

DOI: https://doi.org/10.5281/zenodo.15519708

merely general sentiment tags. By increasing the dataset with additional emotion-specific terms and weighted scores, this method embraces subtle emotional changes, resulting in a more natural and accurate sense of sentiment.

3.7 Polarity Shifters and Inverters:

The second is fixing polarity flips in a sentence. If left uncured, they would cause improper sentiment analysis. For instance, in the sentence "The food at that place is not bad," "not" would under normal circumstances carry a negative sense, but preceded by "bad," it flips the sense upside down, thus the sentiment becomes positive. For such situations to be dealt with, an inversion data dictionary and a new polarity flips dictionary are defined. The final step is dealing with polarity flips in a sentence. They will provide incorrect sentiment analysis if they are not treated. For example, in the statement "The food at that place is not bad," the word "not" typically implies a negative, but when placed in front of "bad," reverses the sense, so that the attitude becomes positive. In order to cope with such an eventuality, a new dictionary of data for inversion and polarity reversals is suggested [5].

4. Future Work and Conclusion

This study identifies DLSTA's strengths and areas for improvement in detecting emotion from text. Future studies can focus on fusing multimodal approaches, raising contextual understanding, and optimizing accuracy in processing sarcasm and evolving language styles. Real-time applications in healthcare, education, and customer care can further boost emotional comprehension. With the evolution of AI and NLP, detection of emotion will become even more precise, promoting human-computer interaction and decision-making [4],[5].

5. References

[1] Kumar, S., Roy, P. P., Dogra, D. P., & Kim, B.-G. (2023). "A Comprehensive Review on Sentiment Analysis: Tasks, Approaches and Applications." arXiv preprint arXiv:2311.11250. This report provides an extensive overview of sentiment analysis, discussing various tasks, methodologies, and applications across different domains.

[2] Sharma, R. (2020). "Sentiment Analysis and Emotion Detection Using Deep Learning Techniques." M.Tech Thesis, Department of Computer Science, IIT Delhi.

[3] Liu, B. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers.

This foundational text provides an in-depth exploration of sentiment analysis techniques and their applications.

[4] Kratzwald, B., Ilić, S., Kraus, M., Feuerriegel, S., & Meinel, C. (2018). Deep Learning for Emotion Recognition on Small Datasets using Transfer Learning. Proceedings of the 2018 ACM Conference on Web Search and Data Mining (WSDM '18), 464–472.

[6] Dhruvi D. Gosau, Himangini J. Gohil, Prof. Hardik S. Jayswal, "A Review on Emotion Detection and Recognization from Text using Natural Language Processing"

[7] Grover, S., et al. (2022). Emotion Detection from Text and Analysis of Future Work: A Survey. RIF Analitica.

[8] Shivhare, Shiv Naresh, and Saritha Khethawat. "Emotion detection from text." arXiv preprint arXiv:1205.4944 (2012).

[9] Amandeep Kaur, Vishal Gupta "A Survey on Sentiment Analysis and Opinion Mining Techniques"

[11] Lin Yue[1,2,3,5] · Weitong Chen[3] · Xue Li[3,4] · Wanli Zuo[5] · Minghao Yin[1,2,5]. "A survey of sentiment analysis in social media."

[12] Armin Seyeditabari, Narges Tabari, Wlodek Zadrozny "Emotion Detection in Text: a Review"

[13] Bharat Gaind, Varun Syal, Sneha Padgalwar "Emotion Detection and Analysis on Social Medi

DOI: https://doi.org/10.5281/zenodo.15519708