# Sentiment Analysis using Fuzzy Logic: A Review

Saberi Goswami
*Dept. of Computer Science and Engineering*
*Techno India University, West Bengal,*
*Kolkata, India*
*saberigoswamidas@gmail.com*

Supratim Bhattacharya
*Dept. of Computer Science and Engineering*
*Techno India University, West Bengal,*
*Kolkata, India*
*bhattacharya.supratim@gmail.com*

Jayanta Poray
*Dept. of Computer Science and Engineering*
*Techno India University, West Bengal,*
*Kolkata, India*
*jayanta.p@technoindiaeducation.com*

*Abstract*—**Sentiment analysis is an automated system, which uses Natural Language Processing as a tool to analyze people's opinion, sentiment, emotional attitude from a data in textual or vocal form and is also well studied in data mining, text mining and web mining domains. The sentiments or opinions can be broadly classified into positive, negative or neutral.**

**This paper presents a comprehensive survey of the traditional models and the most recent state-of-the-art sentiment-based tools and techniques that are employed to classify and compare textual data on social media platforms. A new approach to analyze texts using a recent transformer model known as RoBERTa model has been discussed, which is aimed to overcome the problems, arise in the previous approaches.**

***Keywords—Sentiment Analysis, Fuzzy Logic, Natural Language Processing, Transformer model, RoBERTa.***

## I. INTRODUCTION

Sentiment analysis can be described as a contextual text mining method to identify and extract information from a source, usually used by business analysts, product managers, customer support directors, human resources, workforce analysts, and other stakeholders, for an understanding of the social sentiment towards a particular service, product, situation, person or business,

Various methods and models have been used in the research of sentiment analysis since the last few years. Methods like rule based and machine learning approaches, like Naïve Bayes, Logistic Regression, Support Vector Machines, or Neural Networks have been extensively used to create various Sentiment Analysis models over the years.

Fuzzy logic is an approach to find one of the multiple possible truth values for same variable. It is a process to solve problems with some open, vague data and heuristics to obtain accurate conclusions. The use of fuzzy logic for Sentiment Analysis has been of much help in the recent years. Human opinions may not always be either black or white. Fuzzy logic helps to find the opinions or sentiments which fall in the category of "grey" to provide an accurate sentiment to vague information.

This paper discusses the few ways Fuzzy Logic has been used for Sentiment Analysis and proposes a model using a Transformer [7] along with Fuzzy Logic.

## II. LITERATURE REVIEW

### A. Fuzzy Sentiment Analysis

The authors of the paper[1] propose a rule-based system for sentiment analysis using fuzzy membership degrees, to obtain more refined outputs. The performance of the proposed method is compared with commonly used sentiment classifiers like Decision Trees and Naïve Bayes and it is found that the result of the fuzzy based approach is better than the other algorithms. The fuzzy approach also defines different degrees of sentiment without using large number of classes.

Fuzzy Classifier resulted in better prediction accuracy across the Dual Sentiment Datasets. In case of the, the result of the fuzzy approach is similar to those of the machine learning approaches.

The paper[2] develops a model, built on the project proposed by Haque and Rahman (2014), to analyse the content of social media (such as tweet in Twitter) to understand customer feedback or opinion, which would be helpful to create a computer application that an organisation may use for the representation of customers' opinions on a product or service. Figure below shows the model [2]:

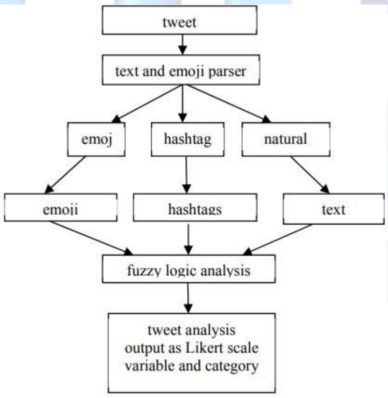


Fig. 1: Flowchart for Sentiment Analysis of Social Media data using Fuzzy Logic

The output passes through the fuzzy logic module, and the tweets are scored from 0 to 1, and assigned to 5 distinct classifications: strongly positive, positive, neutral, negative, strongly negative.

The authors of the paper[3], proposes an Aspect based Sentiment Summarization (ASFuL) with fuzzy logic by categorising opinions polarity as strong positive, positive, negative and strong negative. It also uses the Imputation of Missing Sentiment (IMS) mechanism to integrate the non-opinionated sentences. Fuzzy Logic has been used to find sentiment classes in the review.

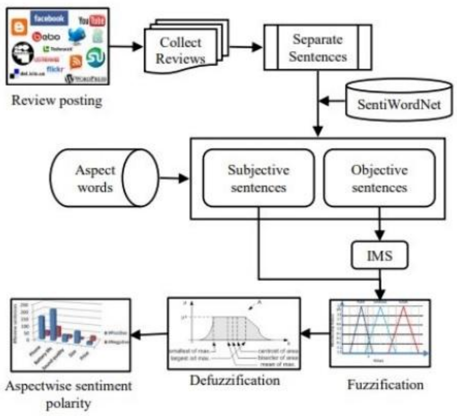


Fig. 2: AsFulModel for Sentiment Analysis

Authors of [4] introduce Fuzzy sentiment analysis using convolutional neural network, which is a popular machine learning approach for sentiment analysis. The concept of fuzzy sets is used to express the sentiment degree of a sentence. Comparison of the Euclidean distance analysis shows that this method is more efficient than the other methods.

This research performs sentiment analysis in the form of sentiment degree in a text. The raw data needs to be cleaned and the data preparation process, including tokenization, encoding and padding is carried out. The data is then divided into training and test data, and convolutional neural network model is applied. The results of the method are probability values of a text having a negative, neutral or positive sentiment class and the results are modified into fuzzy set. The flowchart is presented in Figure 3.

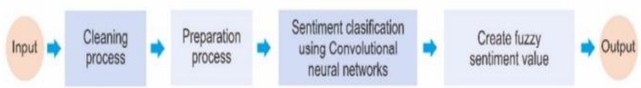


Fig. 3: Flowchart for Fuzzy Sentiment Analysis using Convolutional Neural Network

The paper[5] proposes an innovative approach based on fuzzy natural logic. The main aim is to show that Fuzzy Natural Logic helps in accurate extraction of both semantic meaning and sentiment analysis.

Paper [6] identifies tweet sentiment during Covid-19 by using a combination of fuzzy logic methods with artificial intelligence. The experiment was performed on around 200 simpler twitter data, with several pre-processing techniques; using twitter API- would become inefficient with larger and more difficult datasets.

The Fuzzy rule-based system (FRBS) in [7] extracts the sentiment and classifies the polarity into negative, neutral, and positive.The outputs of the FRBS need to be enhanced, because the rules between the variables overlap at times. To solve this, CSA is used to optimize the outputs and obtain the best result. Like most optimization algorithms, the crow search algorithm has limitations of "slow convergence and easy fall into local optimum".

### B. *Research Gaps*

- The accuracy of the discussed methods mainly depends upon the data or corpus, so the systems become unstable with changes in the size and contents of the data.
- It was also found that ambiguous, vague and short conversational texts are the most difficult to analyze.
- In the paper which implements the Fuzzy Natural Logic method, there was no approach to determine the degree of positivity, negativity or polarity of the expressions.
- The LSTM model is quite slow and without data augmentation, the accuracy of the result is quite low.
- The use of optimization algorithms, like the crow search algorithm has slow convergence and the results fall in local optimum

### C. *Transformer Model*

- Transformer is a deep learning neural network model [8] that learns context and meaning by tracing relationships in sequential data.
- Transformer model uses encoder-decoder architecture.
- The encoder processes the input iteratively one layer after another, while the decoder consists of decoding layers that perform same operation on the encoder's output.

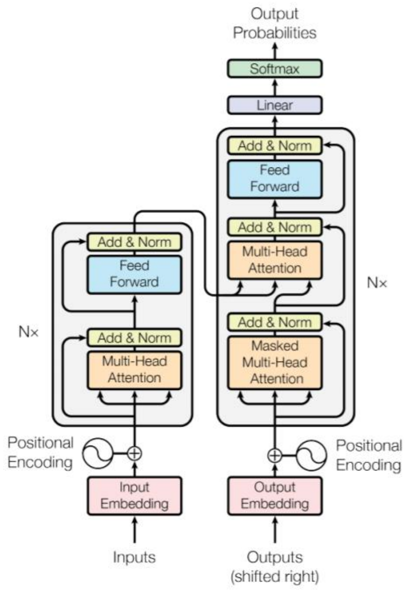


Figure 1: The Transformer - model architecture.

*a) The Encoder Block has the following parts:*

- **Embedding Space-**Where the word sequence is converted into Vectors.
  **Word sequence→ Embedding → Positional Embedding → Final Vector (Context).**
- **Multi-Head Attention Part-**Which works on self attention mechanism [8], i.e. the part of input it should focus on
- **Feed Forward Network:** Here Parallelization is applied as the attention vectors are independent of each other unlike other neural networks, which make it faster.

*b) The Decoder Block has the following parts:*

- **Embedding layer and Positional encoder:** Like the encoder block
- **Masked Multi-Head Attention Part:** So that it predicts the next word itself using previous results.
- **Encoder-Decoder Attention Block:** The mapping takes place here, resulting in attention vectors.
- **Feed-forward Unit:** Output vectors become easily acceptable by another decoder block or a linear layer.
- **Softmax Layer:** Transforms the input into the highest probability distribution.

### D. RoBERTa Model

- **RoBERTa (Robustly Optimized Bidirectional Encoder Representation from Transformer Pre-training Approach) [10]** was proposed by the Meta Research team.
- It is the most recent model (2019) built on the previous BERT transformer model [9], which is used by Google for common NLP applications.
- The BERT model was modified by changing the key hyper parameters and removing the next-sentence pre-training objective to obtain the RoBERTa model.
- The RoBERTa model has been trained with much larger mini-batches and learning rates are much higher in comparison to the previous BERT model.

### E. RoBERTa Model Corpus

The model is trained on five English-language corpora of varying sizes and domains, of over 160GB of uncompressed text.

- **BOOKCORPUS** plus **English WIKIPEDIA** - original data used to train BERT. (16GB).
- **CC-NEWS-** English portion of the CommonCrawl News dataset. - 63 million English news articles. (76GB after filtering).
- **OPENWEBTEXT**, an open-source recreation of the WebText corpus- web content extracted from URLs shared on Reddit with at least three upvotes. (38GB).
- **STORIES**- subset of CommonCrawl data filtered to match the story-like style of Winograd schemas. (31GB).

## III. Results And Discussions

From the papers mentioned above we can see that hybrid models usually yield more accurate results as compared to the traditional or learning models. Models using fuzzy logic are shown to be more accurate for opinionated texts or data.

Paper [1] table illustrates different values for the membership degrees to the negative and positive class, a single output, a dual output, and sentiment intensities output. For sentiment intensities, we use the categories from the multi-sentiment dataset. The first four rows show the dominant of one sentiment. Row five shows only one sentiment. Rows six and seven show the presence of both sentiment in the text. Row eight shows what happens when the text has no sentiment.

| No | Degree value for Class= Negative | Degree Value for Class= Positive | Single output | Dual output | Intensities output |
|---|---|---|---|---|---|
| 1 | 0.985 | 0.015 | Negative | 98% Negative and 2% Positive | Negative |
| 2 | 0.000 | 0.951 | Positive | 0% Negative and 95% Positive | Positive |
| 3 | 0.000 | 1.000 | Positive | 0% Negative and 100% Positive | Positive |
| 4 | 0.942 | 0.058 | Negative | 94% Negative and 6% Positive | Negative |
| 5 | 0.000 | 0.328 | Positive | 0% Negative and 33% Positive | Somewhat positive |
| 6 | 0.229 | 0.771 | Positive | 23% Negative and 77% Positive | Positive and somewhat negative |
| 7 | 0.529 | 0.471 | Negative | 53% Negative and 47% Positive | Somewhat positive and somewhat negative |
| 8 | 0.000 | 0.000 | Unclassified | Unclassified | Neutral |

Table I: Comparison of Fuzzy Sentiment Analysis from Multi-sentiment Dataset

[4] Proximity test of two vectors is performed using Euclid distance to show that fuzzy sentiments represent sentiments of a sentence more smoothly. The results are shown in Fig.

The table shows a comparison between Semantic Orientation Calculator and Fuzzy Natural Logic, where [5] allows gradient description of the semantics of each word. Same scale [0,1] is used by all the words instead of evaluating in a 10-point scale. The conversion is shown in the table below.

## IV. Problem Statement

In paper [1], the authors propose to use fuzzy approach for polarity classification and categories of emotions.
Authors in [2] have created a twitter sentiment analysis model using social bots.

| System/Word | *good* | *excellent* | *masterpiece* | *bad* | *terrible* | *disaster* |
|---|---|---|---|---|---|---|
| **SO-CAL** | 3 | 5 | 5 | -3 | -5 | -4 |
| **FNL** (*positive*) | 0.8 | 1 | 1 | 0.2 | 0 | 0.1 |
| **FNL** (*negative*) | 0.2 | 0 | 0 | 0.8 | 1 | 0.9 |

Table II: Sentiment Analysis using Different Approaches

Another model with a 3-Phase mechanism called ASFuL is proposed [3] for aspect-based sentiment summarization by using Fuzzy Logic to classify sentiments from product reviews.
In [4], the authors use convolutional neural networks along with fuzzy logic, to analyse sentiment in fuzzy form. The paper sets a recommendation for further research work to apply the concept of fuzzy logic in the classification process, until the output results are in the form of fuzzy set.
The Fuzzy Natural Logic model proposed in [5] presents only an approximation and has to further work on solving the problem of compositionality in semantic orientation.

Roberta Model [6] can be defined as a replication study of BERT pre-training (Devlin et al., 2019). BERT was found to be undertrained and hence an improved method was proposed for training BERT models, called RoBERTa, which can exceed the performance of the post-BERT approaches. The modifications include: (1) training the model longer, with more data; (2) removing the next sentence prediction objective; (3) training on longer sequences; and (4) dynamically changing the masking pattern of training data.

The paper [7] presents a hybrid model of Transformer and Recurrent Neural Network, where the Roberta Model has been used for Sentiment Analysis using Long Short-Term Memory, and referred to as the RoBERTa-LSTM model. Data augmentation and text pre-processing is performed and the cleaned corpus is passed into the RoBERTa-LSTM model for training and sentiment analysis. RoBERTa efficiently encodes the words into word embedding and LSTM captures the long-distance dependencies. The experimental results demonstrate that the accuracy has improved from 85.89% to 91.37% after data augmentation.

### A. *Proposed Model*

An efficient hybrid model for Sentiment Analysis can be created using the RoBERTa model and Fuzzy Logic. The RoBERTa model acts as an efficient transformer and encodes the words into word embedding and the Fuzzy Logic can be applied to perform Fuzzy Sentiment Analysis.

### B. *Algorithm and Flowchart*

*a) Flowchart:*

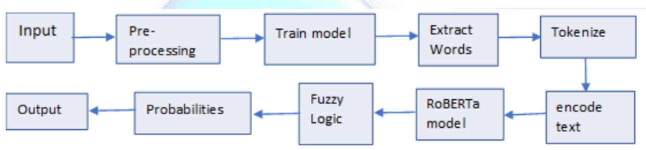


Figure 8: Fuzzy Sentiment Analysis using RoBERTa-Fuzzy model.

## V. CONCLUSIONS:

Sentiment Analysis is a vast area of research with continuous advancement of techniques and approaches required to overcome all the challenges associated with this field of Natural Language Processing. The models mentioned in this paper are from the recent research works in the last five years, where the authors have tried to find the best solution to the previous problems, using Fuzzy Logic. The model proposed in this paper is another such model using the most recent and updated transformer model, known as RoBERTa. This is combined with fuzzy logic for more accurate analysis of sentiments from texts with vague meanings or polarity. The proposed model is in experimental phase. This can be further modified with its successful application for overcoming the accuracy and other major challenges of Sentiment Analysis.

## REFERENCES

[1] Fuzzy Approach for Sentiment Analysis Chris Jefferson, Han Liu, IEEE and Mihaela Cocea, IEEE 2017

[2] Applying fuzzy logic for sentiment analysis of social media network data in marketing Karen Howellsa *, Ahmet Ertugana, 2017

[3] A Framework for Aspect Based Sentiment Analysis Using Fuzzy Logic A. Jenifer Jothi Mary1 and L. Arockiam2 2018

[4] Fuzzy Sentiment Analysis Using Convolutional Neural Network Sugiyarto, Joko Eliyanto, Nursyiva Irsalinda1 and Meita Fitrianawati

[5] Adrià Torrens Urrutia, M. Dolores Jiménez-López & Vilém Novák, "Fuzzy Natural Logic for Sentiment Analysis: A Proposal", 978-3-030-86887-1_6 , (2021).

[6] Devi Ajeng Efrilianda , Erika Noor Dianti , Oktaria Gina Khoirunnisa, "Analysis of twitter sentiment in COVID-19 era using fuzzy logic method", Journal of Soft Computing Exploration, 2746-0991, (2021).

[7] Mazen Sharaf AL-Deen , Yu lasheng, Gamil R. S. Qaid, "Study on Sentiment Classification Strategies Based on the Fuzzy Logic with Crow Search Algorithm", (2022).

[8] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, Illia Polosukhin, "Attention Is All You Need", 1706.03762v5, (2017).

[9] Jacob Devlin Ming-Wei Chang Kenton Lee Kristina Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", 1810.04805v2 , (2019).

[10] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, Veselin Stoyanov, "RoBERTa: A Robustly Optimized BERT Pretraining Approach", 1907.11692, (2019).

[11] RoBERTa-LSTM: A Hybrid Model for Sentiment Analysis with Transformer and Recurrent Neural Network, Kian Long Tan, Chin Poo Lee, Kalaiarasi Sonai Muthu Anbananthen, And Kian Ming Lim. Feb 2022

[12] Challenges of Evaluating Sentiment Analysis Tools on social media, Diana Maynard, Kalina Bontcheva 2016

[13] Who cares about sarcastic tweets? Investigating the impact of sarcasm on sentiment analysis Diana Maynard, Mark A. Greenwood 2014

[14] Boiy, Erik and Marie-Francine Moens. A machine learning approach to sentiment analysis in multilingual Web texts. Information retrieval, 2009.

[15] Wang, Dong and Yang Liu. A pilot study of opinion summarization in conversations. in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL-2011). 2011.

[16] A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools, D. P. Acharjya School of Computing Science and Engineering VIT University Vellore, India 632014, Kauser Ahmed P School of Computing Science and Engineering VIT University Vellore, India 632014