DESIGN AND DEVELOPMENT OF FUZZY LOGIC BASED SENTIMENT ANALYSIS SYSTEM FOR ONLINE REVIEWS & SOCIAL MEDIA POSTS

Submitted in fulfilment of the requirement for the award of degree of

DOCTOR OF PHILOSOPHY

(Information Technology)

By

Srishti Vashishtha

2k17/Ph.D/IT/03

Under the supervision of

Prof. Seba Susan

Department of Information Technology

DEPARTMENT OF INFORMATION TECHNOLOGY

DELHI TECHNOLOGICAL UNIVERSITY

2022

CANDIDATE DECLARATION

I hereby declare that the work being presented in my Ph. D thesis entitled "**Design and development of fuzzy logic based Sentiment Analysis system for online reviews and social media posts**", in fulfilment of the requirement for the award of degree of **Doctor of Philosophy in Information Technology** and submitted to the Department of Information Technology of Delhi Technological University, Delhi. It is an authentic record of my own work under the supervision of **Prof. Seba Susan, Professor, Department of Information Technology.**

The matter presented in this Ph. D thesis has not been submitted for the award of any other degree elsewhere.

Signature

Srishti Vashishtha
(2k17/Ph.D/IT/03)

This is to certify that the above statement made by the student is correct to the best of my knowledge.

Prof. Seba Susan
Professor,
Department of Information Technology
Delhi Technological University, Delhi

ACKNOWLEGMENTS

I am extremely grateful to my supervisor, Prof. Seba Susan for introducing me to the intriguing field of research in artificial intelligence and machine learning with fuzzy logic. Starting with my first work on Fuzzy logic, she inculcated in me great zeal and interest for doing thorough research in this field. I thank her for accepting me as her student and also for her confidence in me without which this thesis would not have been possible. I also thank her for always listening to my ideas with lots of patience and select the best way out of any problem. I can never forget her encouragement and constant support in my research and I sincerely thank her for everything. This thesis would not have been possible without my family's strong support. I thank my parents for teaching me hard work and perseverance, my dearest husband for his cooperation and keeping me positive in the darkest hours.

Srishti Vashishtha

CERTIFICATE

This is to certify that the thesis entitled "**Design and development of fuzzy logic based Sentiment Analysis system for online reviews and social media posts**", which is being submitted by **Ms. Srishti Vashishtha** to the department of Information Department, Delhi Technological University, Delhi, for the award of the degree of **Doctor of Philosophy**, is a record of bonafide research work, carried out by her under my guidance and supervision. The thesis has reached the standards fulfilling the requirements of the regulations regarding the degree. The results contained in the thesis have not been submitted to any other university or institute for the award of any degree or diploma.

Prof. Seba Susan

Professor,

Department of Information Technology

Delhi Technological University, Delhi

ABSTRACT

Sentiment is a person’s frame of mind elicited when he/she confronts a specific topic, person, or entity. Comprehending and understanding humans’ views, beliefs, attitudes, or opinions towards a particular entity is sentiment analysis. This process of sentiment analysis can be automated using computational techniques. Users freely express their opinions on websites and social media platforms, we can use this data for analyzing and extracting the sentiment behind the data by applying the concept of fuzzy logic. Since real-world data is imprecise, vague, and not crisp, therefore fuzzy logic is required to deal with such subjective data. Fuzzy approach is based on the premise that key elements in human thinking are not just numbers but can be approximated to tables of fuzzy sets, or, in other words, classes of objects in which the transition from membership to non-membership is gradual rather than abrupt. Sentiment is basically human emotions, understanding human emotions. We have built brain inspired Sentiment Analysis (SA) framework to help machines emulate human inference of sentiment from natural language. We have developed five methodologies for addressing SA using fuzzy logic techniques. The first objective focuses on natural language words; SA is carried out by using these sentiment bearing words only by applying fuzzy logic. The next objective deals with the creation of key phrases and the computation of fuzzy scores for these phrases to perform SA. In the third objective different neuro-fuzzy networks machine learning models for SA are built. The subsequent objective focuses on social media platforms, its importance, and how social media posts can be analyzed using fuzzy concepts for SA. The last objective is about speech emotion recognition systems, and how emotions and sentiment can be evaluated from speech using various multimodal speech and text cues with fuzzy inferencing.

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

This thesis contains an investigation of the fuzzy logic-based sentiment analysis systems by computing sentiment from natural language in the form of words, phrases, and speech. By doing so, we endeavour to bring about improvements in the existing sentiment analysis algorithms by amalgamating fuzzy logic concepts into sentiment analysis that helps in dealing with the fuzziness of natural language in a very efficient and automatic manner.

1.1 FUZZY LOGIC BASED SENTIMENT ANALYSIS- AN OVERVIEW

In today's life, one cannot imagine life without internet services due to the fast-growing development of web technology. The Internet has already changed the world, but the big changes that Internet will bring still lie ahead. There is an increasing demand for opinion-rich repositories like social media platforms, websites, and blogs; encompassing numerous online reviews. Using these, anyone can effectively apply information technology to understand and investigate the opinions of different entities. Thus, it has led to the formulation of different evaluation applications in this field. The digital universe had estimated to consist of 44 zettabytes of data at the beginning of 2020. In 2019, Google processed 3.7 million queries, Facebook saw one million logins, and YouTube recorded 4.5 million videos viewed every 60 seconds. The number of internet users has risen from 3.7 billion in 2018 to 4.5 billion in 2019. Data production is high, it is being produced every minute by internet applications like emails, google apps, WhatsApp, music apps, etc., and social websites like Facebook, Twitter, Instagram, etc. People can easily interact with each other at a global level, react to different events, and have freedom of expression on these web-based platforms. The process which makes use of the content posted by people for evaluating their expressions and opinions is popularly called Sentiment Analysis (SA) or Opinion Mining. This vast amount of data can be used as input to the SA process for prediction, marketing, research purpose, data mining, and many other purposes. For example, marketing is required to analyse the mindset of customers, what the customers need from a product or service can be analysed by conducting a survey or by collecting reviews about a product or service. Also, public sentiment regarding any social issue can be analysed easily. SA is a sequence of methods, techniques, and tools that aim to analyze subjective information like attitudes opinions, feelings, appraisals, and emotions about any individual, event, product, service, etc (Pang and Lee 2008). Followed by detection of polarity for the given analyzed information. This informative data can be in textual mode, audio, or video mode. The polarity determined by SA can be bipolar- positive or negative; tripolar- positive, negative, or neutral; and further it can have more granular levels like: strongly positive, positive, strongly negative, negative, or neutral (Liu 2012). The basic steps of this process include the collection of data (input), test preparation (pre-processing), feature extraction, feature selection, detection of sentiment, classification of sentiment, and presentation of output.

We are required to build a novel cognition-inspired sentiment analysis framework to help machines mimic human inference of sentiment from natural language. Sentiment Analysis along with Natural Language Processing (NLP) is used to study different relations existing between human emotions, feelings, and opinions. NLP is concerned with the development of computational models for various aspects of human language processing which includes both understanding & generation. The researchers collect information and knowledge on how humans understand and apply language to create models, tools, and techniques via NLP. NLP techniques can be applied at different levels- document, sentence, fine-grained, cross-domain, cross-lingual with supervised, unsupervised, and probabilistic generative model-based approaches (Sun, Luo, and Chen 2017).

We managed to develop a few solutions despite the ever-growing complexity of the problems. Both in analysis and design (synthesis) we follow key principles of abstraction and decomposition that help us to arrive at meaningful solutions. Computing systems around us are in abundance and operate on different principles of binary (Boolean or deterministic logic), numeric information, and fuzzy logic. A classical set is a set that has a fixed boundary, for example: “*All students having a height greater than 5 feet are tall*”. This distinction is intuitively unreasonable. The shortcoming is the sharp transition between inclusion and exclusion in a set. The fuzzy set is without any crisp boundary. Its transition from “belong to a set” to “doesn’t belong to a set” is gradual. Fuzzy logic is a methodology based on "degrees of truth" rather than the binary values "true or false" (1 or 0) (Zadeh, 2015). Modern computers are based on Boolean logic while fuzzy methodology is based on the assumption that vital elements in human thinking are not simply numbers but can be estimated to tables of fuzzy sets. Human reasoning logic is not the conventional two-valued logic it is the logic with fuzzy truths, fuzzy rules of inference, and fuzzy connectives. Fuzzy logic plays an elemental role in distinct aspects of the human thought process.

Opinion words are fuzzy and can be expressed apart from binary forms- true and false. For instance, the subjective words “*Fantastic*”, “*Bad*”, and “*Very Good*” can be effortlessly represented by fuzzy logic by assigning them to classes with some degree of membership. Further, fuzzy sets can be defined for these words based on inputs received from opinions. Since opinions are fuzzy and the definition of opinion words can be elucidated differently; fuzzy logic can be visualized as an effective method to accurately extract, analyze, categorize and summarize opinions.

The fuzzy logic-based sentiment analysis techniques developed in this thesis are used for evaluating the sentiment conveyed in online reviews and social media posts. The following are the research objectives which are described in five chapters 3-7:

- Development of fuzzy technique for highlighting of high sentiment cognition words (Chapter 3)
- Development of fuzzy technique for highlighting of high sentiment cognition phrases (Chapter 4)
- Design and implementation of neuro-fuzzy networks to detect sentiment (Chapter 5)
- Analysis of sentiment from social media posts based on fuzzy logic (Chapter 6)
- Design of multi-modal speech emotion recognition by fuzzy inferencing (Chapter 7)

All the problems except multimodal SA and speech emotion recognition, relies on the application of fuzzy logic-based algorithms, using text features, for identifying the sentiment contained in the text. The last problem is solved by applying supervised and unsupervised classification algorithms with fuzzy rules based on both text and speech features.

1.2 MOTIVATION BEHIND THE PROBLEMS ADDRESSED IN THE THESIS

Nowadays, numerous opinion-rich resources are easily available and popularly used by people to follow, explore and understand opinions about other entities. These resources- online review websites, social media websites, blogs, etc. have an abundance of opinions that can be used for different purposes. The real-world data obtained from these resources contain various forms of ambiguities and uncertainties. Fuzzy sets are well-suited to model such kinds of data. The amalgamation of fuzzy logic-fuzzy sets with other mathematical tools has delivered more robust and intelligent systems. Since fuzzy logic deals with logical reasoning, it will aid in the SA process by providing closer views of the exact sentiment values. Thereby, fuzzy logic will enable the users to take an effective decisions about the sentiment of the entities. The importance of fuzziness comes into play while dealing with natural language due to the presence of ambiguity in language. The human brain deciphers incorrect, uncertain, and incomplete sensory information. The linguistic variables present in natural language need to be expressed using the law of mathematics. Fuzzy set theory is equipped to deal with linguistic information as it includes mathematics that implements numerical computation by assigning membership functions to linguistic labels. (Ross 2004). The amalgamation of fuzzy logic concepts into sentiment analysis helps in dealing with the fuzziness of natural language in a very efficient and automatic manner.

In Sentiment Analysis, there is a need to create a bag of words from the textual data, which can be applied for further processing to ascertain the polarity of text. In most of the papers, authors have used in-built lexicon and dictionaries. Their SA methodologies work on these lexicons but not on other lexicons. We need to build algorithms based on standard dictionaries or lexicon, like WordNet, to bring uniformity so that anyone can use them. This will eradicate the biases as well as increase authenticity. Fuzzy logic can be employed on the bag of words created from these lexicons. Suppose we need to compute the sentiment of the following review:

"*The rooms were tidy, very comfortable, and the staff was amazing. I highly recommend this hotel for anyone visiting Shimla.*"

The bag of words can be created by selecting only those words which are significant for evaluating the sentiment behind the text. In this example, the bag of words is: '*room*', '*tidy*', '*very*', '*comfortable*', '*staff*', '*amazing*', '*highly*', '*recommend*', and '*hotel*'.

The sentiment of textual data can also be computed by considering a bag of phrases rather than a bag of words. Some of the bag of phrases for the above example can be: '*rooms were tidy'*, '*very comfortable*' and '*highly recommend*'. Most of the papers which compute sentiment using phrases have used non-fuzzy features like TF, TF-IDF, Pointwise Mutual Information (PMI), and Semantic Orientation (SO). Some of them are using their in-built dictionaries with their own assigned values. We can create phrases using fuzzy features to capture the impreciseness and vagueness of real-world text using standard lexicons.

In recent years, the significance of machine learning in NLP has increased a lot. It can automatically learn deep sentiment representation and capture meaningful features. But there has been no use of fuzzy logic with machine learning neural networks in SA. Linguistic hedges like modifiers concentrators and dilators can improve the efficiency of sentiment classification. We can construct neuro-fuzzy networks to enhance the performance of neural networks. Few linguistic hedges for the above hotel review are '*very*' and '*highly*', these are the concentrators because they are intensifying the sentiment. Words like '*somewhat*' and '*occasional*' are dilators as these words reduce the sentiment value. Some words like '*not*', '*nothing*', and '*although*' are negators that flip the sentiment polarity.

One issue of analysing a vast amount of social media data from websites such as Facebook and Twitter is that the English used there is very informal. Slangs, unconventional language, and shortened forms or abbreviations are commonly used. For instance, a sample tweet like this: "*SharmaJi ka beta fail ho Gaya!! #INDvPak #CT17Final.*" has Hindi language words written in English. While a human can read and understand a particular tweet, based on context and prior knowledge, likely, software would not be able to organise these kinds of tweets into a particular category. We need to build models using fuzzy concepts which can handle the above issues.

It has been observed that the sentiment expressed in online review posts is not only in textual form but also in vocal form. Multimodal SA models can be built which take into account both linguistic and acoustic features. In speech emotion recognition systems, in recent years authors have extracted speech cues and applied classifiers to categorize speech into different states of emotion. We can develop such models that can extract sentiment-based text cues along with speech cues, which can enhance the accuracy and performance of the system; also use fuzzy logic by either fuzzifying the features or using fuzzy classifiers. This fuzzy inferencing will help us to incorporate the impreciseness and vagueness of real-world speech signals.

1.3 ISSUES ADDRESSED IN THIS THESIS AND THEIR SOLUTIONS

It can be observed from the previous section the need for improved fuzzy logic-based algorithms, in the area of Sentiment Analysis, that can evaluate the opinion and sentiment information accurately by dealing with the ambiguity in natural language. After the introduction, the thesis is divided into 7 chapters. Chapter 2 includes a literature review of the state-of-the-art Sentiment Analysis field. Chapters 3-7 illustrate novel solutions to compute sentiment that handle the impreciseness and uncertainty contained in the natural language.

The fuzzy logic-based sentiment analysis solutions to the five problems in this thesis are very briefly described below:

- **Problem 1: A new fuzzy technique for highlighting of high sentiment cognition words is to be developed.**

  In Chapter 3 we propose a novel approach to shortlisting words that aid in sentiment cognition using a combination of fuzzy entropy, k-means clustering, and sentiment lexicon SentiWordNet. This challenging task of simulating the human cognition of words is addressed by developing a model that recognizes sentiment based on fuzzy scores automatically derived from SentiWordNet. Experiments on two benchmark movie review data sets—IMDB and polarity data sets by Pang and Lee—with training by long short-term memory neural networks, yields high accuracy for the proposed methodology as compared to other state-of-the-art methods of SA. In this Chapter, we also analyze the effect of fuzzification on word polarity sentiment scores. These word scores are obtained by deploying two lexicons: SentiWordNet and AFINN. Experiments are conducted on three benchmark datasets: polarity movie dataset by Pang-Lee, IMDB and hotel reviews dataset. The proposed approach is an unsupervised fuzzy logic-based approach for sentiment analysis of textual reviews; a fuzzy cardinality measure is formulated for the evaluation of word polarity scores. The fuzzy model has two versions based on the sentiment lexicon deployed in the model. The comparison of our fuzzy cardinality approach with other non-fuzzy state-of-the-art techniques reveals the superiority of our fuzzy approach.

- **Problem 2: A new fuzzy technique for highlighting of high sentiment cognition phrases is to be developed.**

  In Chapter 4 we propose an unsupervised sentiment classification system that comprehensively formulates phrases, computes their senti-scores (sentiment scores) and polarity using the SentiWordNet lexicon and fuzzy linguistic hedges. Further, it extracts the keyphrases significant for SA using a fuzzy entropy filter and k-means clustering. We have deployed document-level SA on online reviews using n-gram techniques, specifically a combination of unigram, bigram, and trigram. Experiments on two benchmark movie review datasets- the polarity dataset by Pang and Lee and IMDB dataset, achieve high accuracy for our approach in comparison to the other state-of-the-art approaches for phrase-level SA. In this Chapter, a fuzzy logic-based methodology for dynamic plotting of mood swings from tweets has been introduced. The novelty of the paper is the use of linguistic hedges with fuzzy logic to compute the sentiment of the tweets. A comparison of our approach with existing methods, on real-time tweets extracted from online websites, confirms the superiority and efficiency of our method. The tweets used in our experiments are extracted from the timeline of the India Vs Pakistan final ICC world-cup match in June 2017. They reflect the moods of the Twitter users as the match progresses. Using this fuzzy logic-based approach, we successfully plot the dynamic mood vs time and compute the polarity of the sentiment at each time instant.

- **Problem 3: A new design and implementation of neuro-fuzzy networks to detect sentiment to be developed.**

  In Chapter 5 we propose MultiLexANFIS which is an Adaptive Neuro-Fuzzy Inference System (ANFIS) that incorporates inputs from multiple lexicons to perform sentiment analysis (SA) of social media posts. We classify tweets into two classes: neutral and not-neutral; the latter class includes both positive and negative polarity. This type of classification will be considered for applications that aim to explore neutral content posted by the users. Nowadays, there is a new trend, tasks like personality profiling and behavior profiling are executed by organizations for visa applications, military services, and intelligence agencies for tracking anti-nationalists or terrorist activities. Such tasks require sentiment analysis on online social network that classify neutral and non-neutral content about a person or group. Our ANFIS model combines natural language processing (NLP) with fuzzy logic to deal with the fuzziness of natural language in a very efficient and automatic manner. The optimization method used by the ANFIS system is a hybrid learning process, which associates the gradient descent with the least-squares estimate techniques to update the tunable hyperparameters. A novel neuro-fuzzy system: MultiLexANFIS that combines sentiment scores from multiple lexicons-VADER, AFINN, and SentiWord-Net, to classify tweets is proposed; a novel set of 64 rules for the Sugeno-type Fuzzy Inference System is formulated for the same. Single lexicon-based ANFIS variants to classify tweets are also developed and the comparison of fuzzy approaches with non-fuzzy approaches reveals the supremacy of the proposed neuro-fuzzy system for social sentiment analysis.

- **Problem 4: A new fuzzy logic-based technique is to be developed for the analysis of sentiment from social media posts.**

In Chapter 6 we evaluate the sentiment contained in social media data by applying a novel set of fuzzy rules associated with multiple lexicons and datasets. The Fuzzy Inference System (FIS) is developed by amalgamating the concepts of NLP and word sense disambiguation by applying a novel unsupervised nine fuzzy rule-based system to categorize the data into three sentiment classes: positive, negative, or neutral. The proposed system is deployed on nine different public Twitter datasets, three sentiment lexicons, four state-of-the-art approaches for unsupervised SA, and one state-of-the-art approach for supervised machine learning; for performing a comparative analysis. Usually, SA of twitter data is executed using a single lexicon but we have deployed an approach that integrates fuzzy logic with sentiment lexicons for improved sentiment classification. The results of this system can give a vision to researchers to select the best lexicon for handling social media. The proposed FIS is suitable for any lexicon and any two-class or three-class sentiment dataset. The experiments on benchmark datasets produce better performance for our system as compared to the state-of-the-art.

- **Problem 5: The design of multimodal sentiment analysis and speech emotion recognition systems with fuzzy inferencing are to be developed.**

  In Chapter 7 computation of the sentiment using an ingenious set of fuzzy rules has been employed to categorize the review as positive or negative sentiment. The confidence score from the supervised Support Vector Machine (SVM) classification of text and speech cues is considered the input variable for the fuzzy rules. The fusion of fuzzy logic with acoustic and linguistic features for classifying sentiment contributes to a new exemplar in multimodal sentiment analysis. In Chapter 7 an unsupervised Fuzzy Inference System (FIS) is developed for Speech Emotion Recognition (SER) that incorporates audio and text features. The extracted features are pitch, energy, and textual sentiment score. The proposed system is based on the Mamdani Fuzzy Inference model and is capable of determining four emotions: happy, sad, angry, and neutral. The SER-FIS has three variants based on the sentiment lexicon- AFINN, SentiWordNet, and VADER selected for computing textual sentiment score as the text feature. A set of new eleven novel fuzzy rules based on audio and text cues for SER is formulated. The comparative analysis of all variants of the proposed SER-FIS with six state-of-the-art supervised machine learning methodologies for SER indicates the superiority of our model. Investigations reveal that few speakers in speaker-dependent SER have received higher accuracies than the others and the proposed unsupervised FIS can handle multiple datasets without any training while the supervised machine learning algorithms fail for cross-dataset evaluation. The experiments conducted on speech datasets: SAVEE and RAVDESS, indicate that SER-FIS has achieved higher accuracy and f1-scores in comparison to the other state-of-the-art methods.

Chapter 8 provides conclusions and discussions of the work presented and future work.

CHAPTER 2

LITERATURE REVIEW

Sentiment analysis is considered to be a subdivision of data mining, artificial intelligence, natural language processing, and computational linguistics; that also acquires aspects from psychology and sociology. The concept of sentiment analysis was started in the early 2000's, it was founded by Prof. Lillian Lee (Cornell) and Prof. Bo Pang & colleagues. Pang *et al*. 2002 worked on movie review documents by determining the sentiment of each review using machine learning techniques- Naive Bayes (NB), maximum entropy classification, and Support Vector Machines (SVM). Each review is labelled as either positive or negative. The features applied to machine learning techniques were: unigrams, bigrams, frequency vs presence of word feature, and position of the word. In the same year, an unsupervised sentiment classification approach was proposed by Turney 2002. This approach employs the average semantic orientation of the phrases in the review that consists of adjectives or adverbs, for categorizing reviews into two classes: recommended or not recommended. The average semantic orientation is computed by PMI-IR algorithms that use Pointwise Mutual Information (PMI) and Information Retrieval (IR) to calibrate how much similar are the given pairs of words or phrases in the review. In 2008, Pang and Lee 2018 presented a survey that covers the fundamentals and basic applications of sentiment analysis, and its challenges and additionally contain a list of free resources such as lexicons and data sets. The work in this field started gaining popularity from 2008 onwards and is increasing exponentially since then (Mäntylä *et al*. 2018).

2.1 Impact of social media on Sentiment Analysis

Social Media are affecting consumers' preferences by shaping their mental outlook and behaviors. The impact of the internet, principally via social networking, on people's purchasing behaviour, has expanded over the years. The most popular social media websites used for analysing human perspectives and opinions are- Facebook (Neri *et al*. 2012, Siganos *et al*. 2014), YouTube (Han and Kim 2017), and Twitter (Khan *et al*. 2014, Bravo-Marquez *et al*. 2014, Siddiqua *et al*. 2016, Trupthi *et al*. 2017, Arslan *et al*. 2017). To check and measure customers' loyalty and sentiment towards a product, service, or any event on social media, SA systems can be developed. Neri *et al*. (2012) executed SA on around 1000 Facebook posts about news articles and Wang *et al*. (2012) built a system for real-time analysis of public sentiment published in the year 2012 towards presidential candidates in the U.S by analysing Twitter posts.

There are some unsupervised approaches to evaluate sentiment from tweets (Ortega *et al*. 2013) and review documents (Agarwal *et al*. 2013 (a)). In the former paper, the authors created a system that relies on unsupervised word sense disambiguation to attain contextual word polarity. A rule-based classifier and a lexicon- SentiWordNet (Baccianella *et al*. 2010) is used for computing the polarity of a tweet. The second unsupervised approach investigates the review documents by computing the semantic orientation of the phrases using the PMI method and the overall sentiment orientation of the document is ascertained by combining the polarity values of all the phrases in the document. Ghiassi *et al*. 2013 introduced a supervised technique for reducing features using n-grams and also applied statistical analysis to create a sentiment lexicon specifically for Twitter. The use of emoticons on social media has increased swiftly in modern times. People prefer to use only emoticons or a combination of text and emoticons to express their feelings. Yadav and Pandya 2017 handled several problems, like sarcasm detection, multilingualism, handling acronyms and slang language, encountered during SA of textual information with emoticons.

2.2 Applications of Sentiment Analysis

Sentiment Analysis is being applied in a variety of fields- agriculture (Valsamidis *et al*. 2013), news (Li *et al*. 2014a, Li *et al*. 2014 b, Rao *et al*. 2014, Jiang *et al*. 2016), movies (Bhoir *et al*. 2015), hotel industry (Songpan 2017, Park *et al*. 2020, Mostafa 2020), contextual online-advertisement (Adamov and Adali 2016), Indian union budget (Shakeel and Karwal 2016) and banks (Bruno *et al*. 2016, Bidulya *et al*. 2016). A supervised term weighting scheme was proposed to enhance the performance of SA by Deng *et al*. 2014. This scheme is based on how important is a term in a document and how can the importance of a term express sentiment by considering the correlation between terms and sentiment polarity. In (Singh *et al*. 2018, Kumar and Singh 2019), the authors have explored the effect of the demonetization on the public and Indian financial market using SA. Public opinions about demonetization are collected from tweets across the whole country. Consumers express their opinions about products and services via reviews, these reviews are analyzed by different companies for marketing, advertisement, and enhancing their product quality using SA. The computation of sentiment from consumer reviews can be for products (Shivaprasad and Shetty 2017), mobile handsets (Vu *et al*. 2016, Singh 2017), online games (Sirbu *et al*. 2016), and different products collected from Amazon.com (Salehan and Kim 2016).

2.3 Word-level Sentiment Analysis

Sentiment analysis is the mechanism of determining and computing the opinions, attitudes, and feelings expressed using natural language by people. Words are the fundamental building block of language. Every human language spoken or written is composed of words. Word-level approaches to Natural Language Processing (NLP) are the first step toward understanding the natural language. In this category under tasks, we primarily centralize on those fuzzy logic-based techniques which evaluate sentiment from individual words in the text. Andreevskaia and Bergler 2006 developed a dictionary-based Sentiment Tag Extraction Program (STEP) to create a fuzzy set of English sentiment-bearing words for use in sentiment tagging systems. This method extracts only adjectives from the WordNet dictionary by applying the STEP. For each word a score was calculated that can be interpreted as a measure of the degree of membership of each word in the fuzzy sentiment classes. Jusoh and Alfawareh 2013 recommended the application of a fuzzy lexicon and fuzzy sets to determine the degree of positive and negative sentiment. The fuzzy max operator is used to deciding the level of positive or negative degree in a review. But for a robust system a larger fuzzy lexicon is required to cover more words and building a manual lexicon is a tedious and time-consuming task. A Sentiment Fuzzy Classification algorithm with parts of speech (POS) tags was developed by Mouthami *et al*. 2013 to improve the accuracy of sentiment classification on Movies reviews dataset. The text is transformed into numbers by extracting adjectives using POS and then computing the product of The ermn Frequency- Inverse Document Frequency (TF-IDF) of these adjectives. Three sentiment fuzzy sets are created- positive, negative and neutral using the semi-trapezoid distribution as the membership function. The fuzzy sets are created and explained well but the classification process is not described. A three-layer sentiment propagation model (TLSPM) to determine the fuzzy sentiment membership of texts was developed by Zhao *et al*. (2014). In TLSPM first the relationships between texts, topics and words are constructed. Then the sentiment score of texts, topics, and words by their sentiment neighbours are computed, and the final sentiment score is achieved through continuous iterations. A supervised fuzzy SVM is employed to classify the product reviews of amazon shopping website.

Bing and Chan (2014) proposed a fuzzy big data algorithm called FMM system for extracting public opinion from the ambiguous social media data- Twitter dataset. FMM uses a fuzzy logic technique to assign numeric fuzzy values to each word, computes TF-IDF to select top 40% words and uses gaussian fuzzy membership function (MF) for fuzzy sets. In the pro-posed method, the data is transformed into matrixes, fuzzy association rule mining is applied and the fuzzy partitions are defined as [Positive+, Positive, Neutral, Negative, Negative-]. Haque (2014) proposed a SA system based on fuzzy logic for classifying tweets with the help of SentiWordNet lexicon (Baccianella *et al*. 2010). The tweets are categorized into positive and negative classes on the basis of their computed scores and the weights. Manual weights have been assigned to frequently used terms; this task should be automatic. Also, the assigned manual weights are dependent on only 100 extracted tweets, it should be independent so that it can be applied to any piece of text.

De Sousa *et al*. (2015) introduced a SA method for estimating the degree of importance of comments on products created by web users. In this paper the FIS is applied to online product reviews. It is composed of three input variables: author reputation, number of tuples, and percentage of correctly spelled words and one output variable: importance degree of the comment. To compute the sentiment, the sentiment lexicon SentiLex-PT is used to ascertain the semantic orientation of dataset. The benefit of this method is that it reduces the task of analysing the un-countable user reviews of product or service, and avoids noisy comments. In Wang *et al*. (2015), authors proposed an unsupervised fuzzy computing model to identify the polarity of Chinese sentiment words. This paper is demonstrated in three aspects: i) consists of computing the sentiment intensity of sentiment morphemes and sentiment words using three existing Chinese sentiment lexicons, ii) a fuzzy sentiment classifier is constructed and a corresponding classification function of the fuzzy classifier by virtue of fuzzy sets theory and the principle of maximum membership degree, and iii) sentiment words datasets have been constructed to exhibit the performance of their model. This methodology proves that there is fuzziness in natural language and by applying fuzzy concept these problems in SA can be dealt efficiently.

A fuzzy linguistic summarization approach is built for opinion mining of hotel reviews by Dundar et.al (2016). This work targets to create a decision tool that can be used for generating short summary sentences and classify text into following sentiment classes: positive, negative and neutral. Fuzzy quantified sentences are developed by normalizing each adjective to [0, 1] range using the most frequent adjective. The polarity score of sentiment bearing words is computed using SentiWordNet lexicon but how is the total score of review computed is not discussed. Appel *et al*. (2016) developed a hybrid system using NLP techniques, SentiWordNet lexicons, and fuzzy sets to evaluate the semantic orientation of sentences. The system is developed in hybrid mode at distinct levels: (a) techniques deployed by the sentiment classifiers, and (b) the approaches employed to create lexicon. The concept of graduality expressed through fuzzy sets has been demonstrated. Fuzzy granulation i.e., the linguistic discrimination is implemented using linguistic variables and trapezoidal membership functions to represent the subjective classification of sentences into positive or negative. This system achieves significantly improved results compared to supervised machine classifiers: Naïve Bayes and Maximum Entropy, when the latter are utilized in isolation.

In the paper of Wang *et al*. (2016 a), a sentiment computation methodology known as public sentiment discriminator (PSD) is developed and it is based on fuzzy clustering and sentiment orientations. PSD considers rich sentiment with six dimensions (love, joy, angry, sad, fear, and surprise) and contains majorly three models for fuzzy module: Fuzzy Sentiment Term Generation Model, Fuzzy Microblog Sentiment Computing Model and Fuzzy Public Sentiment Detection Model. Experiments reveal that the PSD can attain similar accuracy and F1-measure but more rational results while comparing it with conventional machine learning techniques. An onto-fuzzy logic-based system for SA of tweets is proposed by Joshi et.al (2016). This system is an integration of three do-mains: sentiment analysis, fuzzy logic & ontology. The following algorithms are employed in the system: Naïve String Searching Algorithm, Loopy Belief Algorithm, Enhanced Boosting Classification Algorithm and Fuzzy Logic. The fuzzy logic module takes count and value of positive words and negative words in single tweet (using lexicons) as input and evaluate the polarity score of that tweet. The normalized polarity score is checked to determine the polarity classes: positive, negative or neutral. In this work, fuzzy logic has been applied vaguely, there is no discussion about use of fuzzy sets. Though the sub categories of polarity classes are unique, that is, fuzzy granularity has been applied. The positive polarity category further sub categorizes into good, better, best and negative polarity category is further sub categorizes into bad, worse, worst. An unsupervised fuzzy clustering approach for SA of twitter dataset can determine the sentiment expressed in social media (Suresh, 2016). Such cluster-based techniques produce essentially accurate experimental results without manual processing, linguistic knowledge or training time. The experimental analysis, proves that the proposed method yields good quality results in the area of social media.

In NLP, one of the key issues is how to convert the unstructured text into a numerical form. Bag of Words (BoW) model can be applied for this purpose, it conducts exact word matching, which can be regarded as a hard mapping from words to numerical term. Fuzzy BoW is an enhanced version, which solves the BoW model issues of extreme sparsity, high dimensionality, and inability to capture high-level semantic meanings behind text data (Zhao and Mao, 2017). It adopts a fuzzy map-ping based on semantic correlation among words quantified by cosine similarity measures. In FBoW, word embeddings are utilized to measure semantic similarity among words and construct fuzzy membership functions of basis terms in BoW. This concept was introduced for document categorization but it can be utilized in the field of sentiment classification. Liu *et al*. (2017) proposed a technique for ranking products via online reviews based on SA and intuitionistic fuzzy set theory. HowNet sentiment dictionary is employed for computing the semantic orientation of words. The semantic orientation of sentence is computed using a proposed algorithm, it contains a clear logic and a simple operational process, which is a valuable attempt for refining more valuable information. This approach converts the identified sentiment orientations into intuitionistic fuzzy numbers by applying intuitionistic fuzzy weighted averaging operator (IFWA). The proposed process lays a good foundation for further conducting studies on ranking products through online reviews.

Ravi *et al*. (2017) presented a novel hybrid model on fuzzy formal concept analysis and SA. The hybrid model executes aspect- and concept-level sentiment analysis on complaints. In the model, following six steps were executed in aspect level SA: tokenization, lemmatization, parsing, parse graph generation, aspect-oriented opinion word extraction, and aspect-level sentiment score computation in order to identify opinions expressed on available aspects in the given sentence. This model is novel in the field of SA for handling complaints and thus it is a useful tool for manager or service provider in many ways. Since the financial company are more interested to study about negative sentiment than positive; the model generates association rules that are rank ordered according to the negative sentiment score. The authors have mentioned some examples of polarity expressions that are not existing in popular lexical resource-SenticNet 3.0. One of the disadvantages of this model is that it selects the aspects manually, it would be better to automate the aspect extraction techniques. Also, it lacks the incorporation of fuzzy linguistic hedges like ‘very’, ‘slightly’, etc. to handle ambiguity or vagueness. In this study, TF-IDF value decides the degree of membership value; but it will be better if any other relevant information is considered to apprehend relative relevance of an aspect.

In a recent work (Montoro *et al*., 2018), a classification model is built using a list named: Affective Norms for English Words (ANEW). It is a set of English words with emotion measures: valence, arousal and dominance for each term is used to build a classification model. This fuzzy based-model is created using k-means clustering, Principal Component Analysis (PCA) and fuzzy trapezoidal membership function. It categorizes tweets into five fuzzy opinion classes (very negative, negative, neutral, positive and very positive). The authors have explained the sentiment classification process wonderfully through visual and graphical representation. A clear and transparent view of opinion classification is depicted, for each tweet the degree of multiple opinion classes is computed, for example a tweet is 70.88% Negative and 12.90% Neutral. The limitation of model is that it does not cover the case of the use of negations in any of the words of the ANEW study and it lacks some new words, because ANEW study was published in 1999. Alharbi and Alhalabi, (2018) developed SA system by integrating multiple inputs of different forms, which can be extracted from the text and its associated attachments using fuzzy logic. The sentiment class outputs of the proposed system are: Very Positive, Positive, Neutral, Negative, Very Negative. The features that are extracted from text are word polarity scores from two sentiment lexicons- SentiWordNet and SentiStrength. These features are fed as input to Takagi-Sugeno Kang (TSK) fuzzy logic control system for the sentimental classification. A limitation of the proposed approach is for the precision, good results were obtained for all the categories, except of the "very negative" as the proposed approach labelled some of the "neutral" category as negative.

Roustakiani *et al*. (2018) created an algorithm for SA with aim of increasing accuracy of existing SA systems by integrating appraisal theory and fuzzy logic. The polarity of words is identified by using a dictionary; further fuzzy logic is used to ascertain the severity of the positive and negative aspects of the text. The proposed algorithm determines the rate of positive or negative sentiment and its results can be used for managing customer complaints and offers, sales, forecasts, etc. The concept of fuzzy logic and how it is been applied has not been explained clearly in this work. Krishna *et al*. (2018) proposed a new model for SA, that employs machine learning algorithms and fuzzy methodologies to classify the sentiment on textual reviews posted on Twitter. The proposed process is an automatic process that is capable of extracting opinions, feelings and, hidden emotions from text data. In this model, different product features are extracted and fuzzy sets are created based on the degree of polarity of these features. A new SA technique for feature extraction and polarity classification based on fuzzy ontology is constructed by Ali *et al*. (2018). The fuzzy ontology module is employed to present the relations between concepts semantically in the domain of transportation. The features in the documents are extracted using semantic knowledge. The polarity of these features is calculated by assigning their opinionated words in the document into SentiWordNet lexicon. The experiments on social media datasets demonstrate that fuzzy ontology with learning algorithms- logistic regression and multi-layer perceptron is more effective than classifiers without ontology. Classifiers without ontology are unable to learn features of aspects; this is because fuzzy logic is able to learn many implicit aspects of transportation.

Sentiment Analysis can be applied to different languages. Sharmista and Ramaswami, 2018 developed a SA system for analysing online reviews about mobile products written in the Tamil language. Their model performs sentence level SA, each feature is extracted by applying POS tagger. The model employs modified version of machine learning algorithm: SVM, named as Fuzzy SVM to categorize the reviews into five classes: Positive, Most Positive, Negative, Most Negative and None. The model has incorporated fuzzy based semantic knowledge to handle any kind of ambiguity or uncertainty in the data. The proposed architecture contains word sense disambiguation, supported by lexicon, and natural language parsing techniques. Another work in field of SA, analysed the Arabic language content postposteded on social media (Rattrout and

Ateeq, 2019). The authors in this work created a hybrid system that integrates dictionary-based approach and fuzzy logic for classifying Arabic sentences based on their opinion values. The word level polarity is evaluated with the help of a dictionary- SentiWordNet. The fuzzy logic module has three inputs: word level polarity, Sentence Likes' Ratio and Account Orientation. The fuzzy logic finally classifies sentences into seven classes: very positive, positive, good, neutral, not good, negative, and very negative. Though this study has overcome the challenges of dealing with Arabic text in the SA field, it has not described the fuzzification method implemented in the fuzzy logic module.

In another recent work, Bedi and Khurana (2020) proposed a hybrid model- FLSTMD (Fuzzifier-LSTM-Defuzzifier) that combines fuzzy logic with deep neural network- LSTM. The reviews to be processed for SA, are converted into integral values by using word embeddings. Word embedding is a concept where each word is transformed into a vector of real number values. The advantage of this model is that it has more processed output by mapping fuzzy output to numerous sentiment classes: Extremely positive, Positive, Neutral, Negative, and Extremely negative, for classifying the reviews. Since each output has different intensity levels this reduces ambiguity to a certain extent. Bernabé-Moreno *et al*. (2020 presented an approach that automatically extracts a polarity dictionary from the stock market, without human intervention. The proposed system identifies price changes of particular stocks over time, using them as a navigating polarity value. The TF–IDF technique is applied to compute the TF–IDF value for each term obtaining the signed guiding polarities. GloVe (Global Vectors for words representation) algorithm is applied to create the global vectors and compute the neighborhood embeddings; the TF-IDF values are disseminated within the neighborhood. In the last step, the terms are mapped to fuzzy linguistic labels and provide a supporting indicator to indicate how reliable the scores are. In the proposed approach, there are no human biases in the overall process but the polarity values are context-dependent, i.e. specific to the stock market domain.

Most of the SA approaches use machine learning classifiers like Naïve Bayes (Bhoir and Kolte, 2015; Songpan 2017; Tripathy *et al*., 2016), Support Vector Machine (SVM) (Li *et al*., 2013; Tripathy *et al*., 2016), and Maximum Entropy (Tripathy *et al*., 2016). But there has been little use of classifiers based on fuzzy sets. The importance of fuzziness comes into play while dealing with natural language due to the presence of ambiguity in language. The concept of fuzzy sets was formulated by Zadeh (Zadeh 1975). Fuzzy sets can be applied to decide the degree of a positive or negative word, with the help of fuzzy memberships, for evaluating sentiment (Zhao *et al*., 2014). In (Jusoh *et al*., 2013), the proposed model for SA shows that not all positive or negative words can be treated as equal; some words are more positive or negative compared to other words. The concept of fuzzy helps us to deal with real-world problems. SA is performed for product reviews to classify them as positive, negative, or neutral with the help of a fuzzy model. In (Liu *et al*., 2017), intuitionistic fuzzy set theory is applied to convert the sentiment orientations into fuzzy numbers. There are SA approaches that incorporate the effect of different linguistic hedges with fuzzy logic to compute the sentiment (Vashishtha and Susan, 2018; Liu *et al*., 2017). While the SentiWordNet lexicon is used by (Vashishtha and Susan, 2018), the Feature Orientation dictionary is used in (Liu *et al*., 2017) to calculate the fuzzy value of each word, and further, these values are used to evaluate the results. Triangular fuzzy sets can be applied to hotel reviews with the help of three quantifiers: "most", "half of" and "few". Short sentiment summaries are created from these fuzzy quantified sentences (Dundar *et al*., 2016). Fuzzy rule-based systems for SA (Jefferson and Cocea, 2017; Liu and Cocea, 2017; Vashishtha and Susan, 2019) have shown better results than commonly used Naïve Bayes, Decision Trees, and SVM. The aim is to provide polarity classification degree values. There exist fuzzy logic-based systems for sentiment classification of online customer reviews (de Sousa *et al*., 2015; Nadali *et al*., 2010) using fuzzy inference systems based on rules.

2.4 Phrase-level Sentiment Analysis

Phrase-level sentiment analysis has been of great interest for the past decade because of its practical utility in social sentiment analysis. Turney presented an unsupervised algorithm for the classification of reviews into two classes: recommended or not recommended (Turney, 2002). He presented phrase extraction patterns; then the semantic orientation of a phrase is computed using the PMI-IR algorithm. PMI-IR is Pointwise Mutual Information (PMI) and Information Retrieval (IR), it measures the similarity of pairs of words or phrases. Reference Word Pairs are used for predicting the sentiment of phrases, and the average semantic orientation of the review is used to classify the review. Another work that used Turney's phrase patterns and PMI-IR algorithm integrated several new information sources as features for the Support Vector Machine (SVM) classifier (Mullen & Collier, 2004). These new features: semantic differentiation with WordNet, topic proximity, and syntactic-relation features, boosted the accuracy of the hybrid SVM classifier.

There are numerous phrase-level SA approaches that identify contextual polarities of phrases for sentiment classification. In (Wilson *et al*., 2005), initially, the polarity of the phrase is detected: neutral or polar and then the polarity of the polar expressions is disambiguated. The valence adjustment method is used (Kennedy & Inkpen, 2006; Polanyi & Zaenen, 2006), this method adds contextual valence shifters: negations, intensifiers, and diminishers to positive and negative words for sentiment analysis. Another work introduced an unsupervised SA approach based on snippets for unknown phrases by utilizing sentiment lexicon and the search engine-Bing (Peng & Shih, 2010). The sentiment of the unknown phrase is computed by using top-N relevant snippets returned by Bing, followed by occurrence-based and distance-based approaches. Some researchers (Peng & Shih, 2010), have resolved polarity disambiguation in phrase-level SA by utilizing the SentiWordNet lexicon (Baccianella *et al*., 2010) and a word sense disambiguation algorithm (Adapted Lesk). The SentiWordNet scores, for the correct sense word identified by Lesk; negation and polarity modifications are used as classification features in AdaBoost-MH and Naïve Bayes classifier for SA. PUMA, an automated, phrase-based approach to extract user opinions in-app reviews was developed in 2016 (Vu *et al*., 2016). This approach extracts phrases using part-of-speech (POS) templates, measures the similarity of phrases using vector representation of words from word2vec (Mikolov *et al*., 2013), and groups similar phrases using the soft clustering method. Sentiment analysis can also be carried out using n-gram techniques. In Wang *et al*. (2012), the authors examined various features: n-grams, emotion lexicons, POS, and n-gram positions of tweets; and applied machine learning techniques: LIBLINEAR and Multinomial Naive Bayes. Tripathy *et al*. (2016) utilized various machine learning approaches like Maximum Entropy (ME), SVM, Naive Bayes (NB), and Stochastic Gradient Descent (SGD) for document-level SA with the n-gram method.

An n-gram sentiment lexicon is introduced in (Dey *et al*., 2018), this is an enhanced lexicon that not only contains unigrams but also combinations of unigrams with intensifiers, diminishers, and negations. This lexicon is utilized by the rule-based document level SA approach that uses the ratio of the count of positive and negative sentences as a feature. We generally observed that the results of n-gram based approaches for SA are better as compared to other methods. This motivated us to use n-gram based techniques for SA in our work. Recent works in the field of opinion mining have proved that sentiment classification based on fuzzy setscan handle the uncertainty or vagueness in a pretty skillful way. This skill is essential because human language is sensitive and difficult to comprehend. People try to express their emotions via text, and picking the correct words or phrases that have the correct sentimental value is challenging.

Fuzzification of sentiment for phrase-level SA has been applied in several works (Dalal & Zaveri, 2014; Gupta & Abhinav, 2013; Phan *et al*., 2019; Srivastava & Bhatia, 2013). In Gupta and Abhinav (2013) authors proposed a fuzzy approximation technique to compute the opinion intensity of movie review. This review intensity is combined with probabilistic latent semantic analysis to form a regression model that predicts weekly movie revenue. The strength and polarity of the opinionated phrases are also affected by the presence of linguistic hedges such as ‘very’, ‘slightly’, or ‘not’. Zadeh established the concept of fuzzy linguistic variables and linguistic hedges that change the meaning and membership function of an atomic word (Zadeh, 1975). Feature-based sentiment classification methods have been developed to integrate the effect of several linguistic hedges by using fuzzy functions to match the effect of modifiers, concentrators, and dilators (Dalal & Zaveri, 2014; Srivastava & Bhatia, 2013). Phan *et al*. (2019) proposed a method that detects polarity of only specific kinds of tweets that contain fuzzy sentiment phrases (FSPs).The first step of this method is to detect FSPs, then extract a set of features related to FSPs based on the syntactic, lexical, semantic, and polarity sentiment of the words. The second step is to classify tweets using the Multilayer Perceptron model. A recent paper applied fuzzy sets, fuzzy entropy, k-means clustering for keyphrase extraction (Vashishtha and Susan, 2021 a). The key phrases are created with help of n-gram technique and fuzzy linguistic hedges for sentiment computation.

2.5 Neuro-fuzzy Sentiment Analysis

In 1993, Jang proposed ANFIS, a fuzzy inference system implemented in the framework of adaptive networks (Jang, 1993). It uses a hybrid learning procedure to construct an input-output mapping based on both human knowledge (in the form of fuzzy if-then rules) and designated input-output data pairs. ANFIS have been applied for classification (Žunić *et al*., 2016), prediction (Cosma and Acampora, 2016; Kulkarni *et al*., 2012; Kulkarni *et al*.; Mazumder *et al*., 2018), subjectivity analysis (Rustamov *et al*., 2013; Rustamov, 2018; Kamil *et al*., 2018; Padmaja and Hedge, 2019) and SA (Katta and Hedge, 2019), (Mazumder *et al*., 2018), (Kamil *et al*., 2018; Padmaja and Hedge, 2019; Acampora and Cosma, 2014; Çakıt *et al*., 2019). Twitter data has been analyzed for political issues: using a hybrid system of ANFIS and SVM, where the SVM classifier optimizes the fuzzy rules (Katta and Hedge, 2019), and using Principal Component Analysis (Mazumder *et al*., 2018) with ANFIS to predict political sentiment. Another hybrid classifier of ANFIS with Genetic Algorithm (GA) has shown good results in twitter SA (Padmaja and Hedge, 2019; Acampora and Cosma, 2014). GA extracts the important and semantic features from each tweet and optimizes the fuzzy principles in ANFIS. A study utilized supervised soft computing approaches: Fuzzy Time Series (FTS), artificial neural network (ANN)-based FTS, and ANFIS to estimate the human emotional states expressed in tweets on social network (Çakıt *et al*., 2019). A combination of ANFIS and Hidden Markov Models has been deployed for SA in (Rustamov *et al*., 2013) and subjectivity analysis in (Rustamov, 2018). During the preparatory investigation in this study, it was found that triangular, gaussian and generalized bell MFs have lower root mean square error (RMSE). Machine learning approaches have proved to perform well for sentiment classification, some of the popular algorithms are Naïve Bayes (Parveen and Pandey, 2016), Support Vector Machines (SVM) (Katta and Hedge,2019), (Valdivia *et al*., 2017; Valdivia *et al*., 2018), Long short-term memory (LSTM) and ANFIS. Out of all these options, neuro-fuzzy networks perform generally well in classification tasks since they combine learning with reasoning (Susan *et al*., 2013).

2.6 Social Media based Sentiment Analysis

In past few years, we can notice tremendous progress has been achieved in the task of sentiment classification of social media posts. Among social media posts, tweets are most popular. Most of the researchers have classified tweets according to the sentiment contained in tweets. The different approaches for performing SA of social media posts can be categorized as supervised, semi-supervised and unsupervised approach. In social media, to keep track of user opinion behaviour, historical information about users can be used to develop a content-based supervised model to predict the sentiment. These models are developed in Chen, Wang, and Wenjie (2018) using re- current neural network in order to explore the expression styles of users which give useful information to marketing companies. Models have been developed in Liu, Cheng, Li, and Li (2015) for the sentiment classification of tweets specific to a topic. These classifiers are supervised and developed on common features and mixed labelled data from various topics. Finding the most significant features that contain class- specific information is a subject of investigation in several works (Susan & Keshari, 2019). Many authors have used machine learning techniques like Naïve Bayes (Neethu & Rajasree, 2013; Jain & Katkar, 2015; Parveen & Pandey, 2016; Yan, Yang, & Wang, 2017; Saleena, 2018; Barnaghi, Ghaffari, & Breslin, 2016; Hamdan, Béchet, & Bellot, 2013) and Support Vector Machines (SVM) (Neethu & Rajasree, 2013; Saleena, 2018; Hamdan *et al*., 2013) for Sentiment Analysis from tweets. Windasari, Uzzi, & Satoto, 2017 used n-gram unigram and Term Frequency- Inverse Document Frequency (TF-IDF) as feature extraction methods and applied these features to SVM algorithm for classifying tweets (Windasari, Uzzi, & Satoto, 2017).

Most of the machine learning techniques for emotion classification use the following features: term presence, term frequency, negation, n-grams and part-of-speech (Mejova, 2009). The unsupervised techniques for sentiment classification have the edge that they can adapt to dynamically changing topics and opinions in social media. In microblogging services, we can observe trending topics related to different events and domains. A model based on Latent Dirichlet Allocation (LDA) has been proposed to find emerging topics and investigate the problem of public sentiment variations. This model not only computes the sentiment of tweets but also ranks the most popular and representative tweets among the emerging topics (Tan *et al*., 2014). Sports events invoke immense flow of emotions among fans on twitter. One such event was the FIFA World Cup 2014. The sentiment of users, players, teams, etc. was observed to change over time during a critical match or any other event. These emotions can be analyzed and classified either using supervised classifier (Barnaghi *et al*., 2016) or statistical analysis (Lucas *et al*., 2017).

A recent work investigates Sentiment Analysis of twitter data regarding Artificial Intelligence (AI) assistants (Park & Seo, 2018). This work focuses on the sentiment about these AI assistants to ascertain which assistant is statistically better than the other with the help of VADER lexicon (Gilbert & Hutto, 2014) and T-test, Kruskal-Wallis test, and Mann-Whitney test (Park & Seo, 2018). In another re- cent work (Montoro, Olivas, Peralta, Romero, & Serrano-Guerrero, 2018), a list named: Affective Norms for English Words (ANEW) that is a set of English words with emotion measures: valence, arousal and dominance for each term is used to build a classification model. This fuzzy based-model is built using k-means clustering, Principal Component Analysis (PCA) and fuzzy trapezoidal membership function and finally the twitter text-data is classified into five fuzzy opinion categories (very negative, negative, neutral, positive and very positive).

Fuzzy logic-based systems can deal with vagueness and ambiguity (Zadeh, 2015; Zadeh, 1996). One important contribution of fuzzy logic is the technique for computing with words, i.e. words can be transformed into numerical values for further computation. Fuzzy logic provides us a desirable way to deal with linguistic problems (Ross, 2004). Tsukamoto fuzzy rule- based system has been used in (Liu & Mihaela, 2017; Jefferson, Liu, & Cocea, 2017) for Sentiment Analysis. The input attribute of this system uses trapezoid fuzzy membership function to convert numerical values into fuzzy linguistic terms. This system delivers two outputs: dual output with values for both the positive and the neg- ative class and an output indicating different intensities of sentiment (Jefferson *et al*., 2017). Siddiqua *et al*. integrated a rule-based classifier based on emoticons and sentiment-bearing words with supervised Naïve Bayes classifier to classify sentiment of tweets. This Naïve Bayes classifier is trained with the help of several sentiment lexicons (Siddiqua, Ahsan, & Chy, 2016).

In 1975, Mamdani and Assilian's influential work (Mamdani & Assilian, 1975) introduced the first rule-based controller powered by a fuzzy inference mechanism. Such a system is generally called fuzzy-rule-based system (FRBS). Mamdani FRBS have been developed by researchers (Dutu, Mauris, & Philippe, 2018; Márquez, Peregrín, & Herrera, 2007) for different application problems. Inspired by the Mamdani FRBS, we have developed our fuzzy rule based unsupervised sentiment classification system using the mamdani rule system. Márquez *et al*. (2007) proposed a mamdani fuzzy rule system that learns a linguistic rule base, the para- metric aggregation connectors of the inference and defuzzification in a single step to increase the accuracy. Several authors have worked with fuzzy rule-based systems customized for different application areas (Chang, Chen, & Churn-Jung, 2008; López *et al*., 2015; Sanz, Fernandez, Bustince, & Herrera, 2013; Ishibuchi & Tomoharu, 2001; Ishibuchi & Yamamoto, 2005). A linguistic cost- sensitive fuzzy rule-based classification method can handle imbalanced huge data with good precision and without increasing the execution time (López *et al*., 2015). The effects and specifications of rule weight in fuzzy rule-based classification systems has been discussed in (Ishibuchi & Tomoharu, 2001; Ishibuchi & Yamamoto, 2005).

A fuzzy logic based approach developed by (Vashishtha & Susan, 2018) plots the dynamic mood swings from tweets over time. This approach analyzes the tweets of cricket fans by determining the polarity of tweets and plotting their mood versus time. Few survey papers about twitter Sentiment Analysis describe the various supervised, unsupervised and hybrid techniques for text classification (Martínez-Cámara, Martín-Valdivia, Urena-López, & Montejo-Ráez,

2014) while another paper compares the machine learning based, lexicon based and graph-based classification methods (Giachanou & Crestani, 2016).

2.7 Multimodal Sentiment Analysis and Speech Emotion Recognition

2.7.1 Multimodal Sentiment Analysis

In current times, the opportunity to apprehend people's opinions has embossed the expanding interest both within the scientific society for the new research challenges and in the business world due to the notable benefits in market analysis, financial sector, market prediction, etc. The power of sentiment analysis has been realized during the past decade. Different modes of communication used by human to express their sentiment, other than text, like speech is gaining popularity; hence demanding multimodal sentiment analysis.

One of the initial works that tried multimodal sentiment analysis integrated a subset of audio-visual features with text; besides creating a YouTube dataset (Morency *et al*., 2011). A total of five features were fused together: polarized words, look away, pauses, smile and pitch. Wöllmer *et al.* built a multimodal sentiment analysis system that incorporates a decision-level fusion of text, audio-video features (Wöllmer *et al*., 2013). The linguistic text features are extracted using Bag of Words with TF-IDF, openSMILE tool is used for extracting speech features; text features are classified using SVM (Vapnik 1995), while for audio features Bidirectional Long short-term memory (BLSTM) classifier is incorporated. The overall sentiment classification is performed by applying unimodal scores into BLSTM. Another system that implements the BLSTM classifier (Poria *et al*., 2017; Poria *et al*., 2018) uses *word2vec* vectors for text features, openSMILE tool for extracting speech features on the CMU-MOSI dataset. Results show that multimodal features with BLSTM perform better than multimodal and unimodal features with the SVM classifier.

Majumder *et al.* introduced a hierarchical (HFusion) and a context-aware hierarchical (CHFusion) fusion-based approaches, the former implements a neural network with softmax output, while the latter implements a recurrent neural network (RNN), specifically GRU (Gated Recurring Units) to model semantic dependency among the utterances (Majumder *et al*., 2018). Experiments on CMU-MOSI dataset reveal that CHFusion is the best performer followed by HFusion and Early Fusion (concatenation of multimodal features into a single vector passed as input to SVM). Zadeh *et al.* introduced a tensor fusion network that learns intra-modality and inter-modality dynamics end-to-end in multimodal sentiment analysis (Zadeh *et al*., 2017). Inter-modality dynamics means the interactions between spoken words (text), acoustic (speech) acoustic and visual actions that change the notion of the asserted sentiment, while intra-modality is the unimodal interactions. This tensor network, a fully connected deep neural network using LSTM, achieves good experimental results for both binary and five-class sentiment classification on the CMU-MOSI dataset.

Fuzzy logic-based systems can handle uncertainty and ambiguity (Zadeh 1996; Zadeh 2015). Linguistic problems can be solved using fuzzy logic (Ross 2004). Very few works have incorporated fuzzy logic with speech classification, examples of fuzzy classifiers that include auditory cues are: (Susan and Sharma 2012; Scherer *et al*., 2013). These fuzzy classifiers demonstrate superior performance to the nearest neighbour classifier and multiple layer perceptron (MLP) neural networks. Another work in speech emotion recognition system uses a neural fuzzy network (NEWFM) for the classification of emotion-related audio signals (Zhang and Lim 2015). This NEWFM is based on a function with weighted fuzzy membership and takagi-sugeno defuzzification. In text-based sentiment analysis, several works have used fuzzy logic for better classification of sentiment.

Recently a fuzzy logic based system has been built to classify tweets into five fuzzy classes (very negative, negative, neutral, positive and, very positive) using functions such as Principal Component Analysis (PCA), k-means clustering and Fuzzy trapezoidal membership (Montoro *et al*., 2018). There are three types of fuzzy inference models: Mamdani, Takagi-Sugeno or Sugeno and Tsukamoto. Tsukamoto fuzzy rule-based system has been used in (Liu *et al*., 2017; Jefferson *et al*., 2017) for sentiment analysis. These works use the trapezoid fuzzy membership function to transform numerical values into fuzzy linguistic terms. Fuzzy logic-based approach can analyse real-time tweets to plot the sentiment of tweets over time (Vashishtha and Susan 2018). The sentiment of these tweets depicts the dynamic mood swings of cricket fans while watching a cricket match. Márquez *et al.* developed an evolutionary learning model comprised of Mamdani fuzzy rules. This model acquires rules and aggregation connector parameters in one step to improve the accuracy (Márquez *et al*., 2007).

2.7.2 Speech Emotion Recognition

The discipline of automatically recognizing human emotion and affective states from the speech is generally termed Speech Emotion Recognition (SER). SER can be carried out by using only speech features or a fusion of audio and text features. Numerous research studies have adopted only audio cues and have obtained good results. Shi *et al.* used a back-propagation neural network and a Multi-feature SER decision tree algorithm for recognizing the emotions: happy, angry, sad, and surprise (Shi and Song 2010). Both the algorithms employed multi-acoustic features: MFCC, pitch, energy, an average amplitude of energy, and standard deviation of the third formant. The Hidden Markov Model tool kit (HTK) used for speech, speaker, and emotion recognition is based only on MFCC features (Revathy 2015). The HTK training models are built for emotions such as boredom, disgust, joy, fear, sad and neutral. Another work that used only MFCC features for SER identified emotions- happy, sad and angry- from speech signals; made decisions based on standard deviation (Likitha *et al*., 2017). These works have employed only audio features for speech recognition, while in our proposed work we include textual features as well. The benefit of textual features is that it enhances the speech recognition process by incorporating linguistic features: words and sentiment scores.

Sentiment Analysis (SA) is a field closely related to emotion detection. In SA we can detect positive and negative feelings (Vashishtha and Susan 2018; Vashishtha and Susan 2019) and in some systems, neutral feelings as well, from the given user input text. But determining the emotions from only text features is a challenging task. This task was attempted by Jayswal *et al.*, by extracting emotional keywords from text using natural language processing (Gosai *et al*., 2018). Emotions cannot be detected accurately using only text features, acoustic features are also required. Concatenation of text and audio features into a single vector is one of the simplest methods in hybrid SER models. Houjeij *et al.* proposed a hybrid approach by integrating both speech and text features for detecting five emotions- happy, sad, angry, afraid, or neutral (Houjeij *et al*., 2012). WordNet Affect was used to obtain emotional tags from text and the following speech features: audio spectrum roll-off, audio spectrum centroid, MFCCs, ZCR (zero-crossing rate), and log-attack time were extracted from speech signals. These feature vectors were concatenated and applied to SVM for classification.

Another work (Bhaskar *et al*., 2015) that implemented an SVM classifier for SER, adopted pitch, energy, formants, intensity, and ZCR as speech cues; while the text cues were represented using Term Frequency and SentiWordNet lexicon (Baccianella *et al*., 2010). These hybrid SER models (Houjeij *et al*., 2012; Bhaskar *et al*., 2015) employing the concatenation method do not incorporate the fuzziness involved in natural language while speaking or in written form, while our proposed model successfully incorporates fuzzy logic by applying Mamdani FIS to linguistic and acoustic features. A decision tree SVM model with fisher feature selection has been employed for SER (Sun *et al*., 2019). This model is capable of filtering out unnecessary speech features and retaining relevant required speech features for detecting the correct emotional state. Neural network architectures such as Convolutional neural network (CNN) and Long short-term memory (LSTM) have also been deployed for only audio (Zhao *et al*., 2019) and both audio-textual emotion recognition (Cai *et al*., 2019). The supervised algorithms for SER (Sun *et al*., 2019; Zhao *et al*., 2019) have mostly utilized only audio features for identifying emotions in speech, while in our proposed work we include linguistic features as well. Multimodal SER models based on deep learning have also proved to deliver good results in this field (Yoon *et al*., 2018; Sahu 2019). In (Sahu 2019), the term frequency-inverse document frequency (TFIDF) is used as text feature, and pitch, harmonics, speech energy, pause, and central moments are extracted as speech features. Various machine learning models: Random Forest, Gradient Boosting, SVM, Logistic Regression, MLP, Naive-Bayes, and LSTM are deployed on these features. These supervised machine learning models perform SER well but they require extra training time and computational cost overhead.

The concept of fuzzy logic has been employed in many studies for solving convoluted real-world problems. Few real-world issues also require feature selection techniques for choosing the optimal subset of features that enhances classification accuracy (Sawhney *et al*., 2018; Jain *et al*., 2018). The semantic part of speech contains linguistic information which can be interpreted differently according to the tone of the speech and the context of other words; thus, there exists fuzziness in audio and text cues. There are various research articles based on SER that have embedded fuzzy logic in different forms. A fuzzy inference system based on fuzzy associative memory used MFCC speech features for SER (Ton-That *et al*., 2019). This FIS performed better than Naïve Bayes and SVM. But it only encompasses speech features; for better recognition text features can be analysed also. Esau *et al*. presented an adaptable emotion model based on the fuzzy hypercube (Esau *et al*., 2005). The fuzzy hypercubes are used for the fuzzy classification of emotional states.

Fuzzy nearest neighbor classifiers can also be used for classifying emotions: happiness, anger, surprise, and sadness (Ming *et al*., 2015) and for identifying the speaker (Susan and Sharma 2012) based on speech features. A fuzzy model was developed (Bakhtiyari and Husain 2014) that represents a closer analogy between human brain detection of emotions and machine models. Elbarougy *et al*. constructed a model comprising of three layers: acoustic features, semantic primitives, and emotion dimensions for SER based on a dimensional approach (Elbarougy and Akagi 2012). For the estimation of the emotion dimensions (valence, activation, and dominance), a fuzzy inference system was used. While in our proposed work, the FIS is employed for resolving the fuzziness present in audio samples, by analysing both text and audio features. A fuzzy rule-based system: PROSBER was developed using fuzzy grid algorithm for emotion recognition based on prosody in natural speech (Austermann *et al*., 2005 a; Austermann *et al.*, 2005 b). PROSBER autonomously selects the most significant speech features for each emotion class: happiness, sadness, anger, fear, and neutral, based on the training of audio clips. This is an advantage of using this system because it chooses the appropriate speech features.

Fuzzy systems have been designed for detecting emotions like suicide ideation by performing textual classification of tweets that exhibit suicidal intent (Sawhney *et al*., 2019). In order to combat the ambiguity and vagueness present in natural language in speech form or in written form, fuzzy inference systems have proved to work well. In literature, generally, there are three types of fuzzy inference models: Mamdani, Tsukamoto, and Sugeno. In a previous work of the authors, a Sugeno fuzzy model was built for multimodal sentiment classification of reviews using text and speech features (Vashishtha and Susan 2020 a). Mamdani fuzzy rule-based system (FRBS) was developed by (Mamdani and Assilian 1975). The Mamdani FRBS has been developed by researchers (Vashishtha and Susan 2019) for sentiment classification using text features.

2.8 Deep Learning-based Sentiment Analysis

Deep learning technology is in boom nowadays due to the existence of huge amounts of data. In the field of sentiment analysis, many researchers have deployed different deep learning architectures, to extract correct sentiment from available data. Deep neural network models such as long short-term memory (LSTM) and tree- LSTM have been proven to be effective for sentiment analysis. Wang *et al.*, (2019) developed a capsule tree-LSTM model, introducing a dynamic routing algorithm as an aggregation layer to build sentence representation by assigning different weights to nodes according to their contributions to prediction. Another popular deep learning network- Convolutional Neural Network (CNN) has been extensively used in this field. Phan *et al.*, (2020) established a feature ensemble model that translates tweets into tweet embedding by extracting numerous features using CNN. Fuzzy sentiment phrases are formulated but do not consider the influence of slang and sarcasm in tweets containing fuzzy sentiment. The ensemble model of LSTM and CNN, is able to capture the temporal information of the data, and the other one to extract the local structure (Minaee *et al.*,2019). The experiments of the ensemble model on the movie reviews dataset- IMDB and Stanford Sentiment Treebank2 (SST2), shows that it outperforms both individual models. CNN architecture can be used to generate subword-level representations for detecting sentiment in multilingual sentences (Lal *et al.,* 2019). The generated representations are used as inputs to a Dual Encoder Network which consists of two different BiLSTMs.

A deep learning modified neural network (DLMNN), created by Sasikala and Sheela, (2020), performs sentiment analysis of product reviews by incorporating ANFIS. This network also predicts online products by deploying the improved Adaptive Neuro-Fuzzy Inferences System (IANFIS). Bedi and Khurana, (2020) developed a hybrid model- FLSTMD (Fuzzifier-LSTM-Defuzzifier) that combines fuzzy logic with a deep neural network- LSTM. This model reduces ambiguity to a certain extent by providing different intensities of output (sentiment classes). Domain adaptation tasks such as cross-domain sentiment classification aim to utilize existing labelled data in the source domain and unlabelled or few labelled data in the target domain to improve the performance. Xi *et al.*, (2020) designed a Category Attention Network (CAN), and further developed a model -CAN-CNN to integrate CAN and a Convolutional Neural Network (CNN). This model makes an attempt at interpretability to learn the transferred category attribute words for better optimization.

Multilingual sentiment analysis on social media based on deep neural architectures is gaining a lot of interest in this field (Lal *et al.*, 2019, Agüero-Torales *et al.*, 2021). A cross-lingual graph-based model (transfer-learning), from a rich source language with embeddings of supervised training, on Amazon reviews to a dual-channel neural architecture, was developed by Dong and De Melo, (2018). Cross-lingual deep neural transfer learning technique by LASER with (low-resource) language corpus and BiLSTM, can predict the sentiment of texts in other (high-resource) languages (Kanclerz *et al.*, 2020). Liu *et al.*, (2020) generated word vector representation improvement based on the gate mechanism, which obtains the time-series relationship of different sentences in the multilingual comments through an RCNN. The local features, of the specific aspects in the sentence and the long-distance dependence in the whole comment, are fetched through a hierarchical attention BiLSTM. The deep neural network-BiLSTM has been widely used for detecting sentiments in multilingual social media text (Choudhary *et al.*, 2018; Lal *et al.*, 2019; Liu *et al.*, 2020; Kanclerz *et al.*, 2020; Jamatia *et al.*, 2020).

RESEARCH OBJECTIVES

1. Development of fuzzy technique for highlighting of high sentiment cognition words
2. Development of fuzzy technique for highlighting of high sentiment cognition phrases
3. Design and implementation of neuro-fuzzy networks to detect sentiment
4. Analysis of sentiment from social media posts based on fuzzy logic
5. Design of Multi-modal Speech Emotion Recognition by fuzzy inferencing

CHAPTER 3

DEVELOPMENT OF FUZZY TECHNIQUE FOR HIGHLIGHTING OF HIGH SENTIMENT COGNITION WORDS

In Sentiment Analysis (SA), highlighting the correct words which contribute towards sentiment cognition is very difficult. Simulating this task of shortlisting the words by human observers is challenging due to the complexity of the human mind's processing. In this chapter[1] we have proposed a novel method of selecting words that aid in sentiment cognition using a combination of fuzzy entropy, k-means clustering and sentiment lexicon SentiWordNet. The use of fuzzy entropy is proposed in our work as an innovative step to extract sentiment-carrying words from online movie reviews. We have addressed this challenging task of simulating the human cognition of words by developing a model that recognizes sentiment based on fuzzy scores derived from SentiWordNet in an automatic manner. Experiments on two benchmark movie review datasets-IMDB and the polarity dataset by Pang and Lee, with training by LSTM neural networks, yield high accuracy for our approach in comparison to other state-of-the-art-methods of SA.

:[1] The contents of this chapter are published in "Sentiment Cognition From Words Shortlisted by Fuzzy Entropy," in IEEE Transactions on Cognitive and Developmental Systems, vol. 12, no. 3, pp. 541-550, Sept. 2020, doi: 10.1109/TCDS.2019.2937796.

This work attempts to create a bridge between the human mind's processing of words during reading and Natural Language Processing (NLP) based sentiment analysis system. We propose a sentiment analysis system based on Long Short-Term Memory (LSTM) with the application of fuzzy logic for shortlisting the words. The fuzzy measure- fuzzy entropy is used to extract cognitive words which is fed into LSTM for training. These cognitive words build up a visual attention model that emulates the way human mind perceives text.

The rest of the chapter is organized as follows. Section 3.2 describes the fuzzy entropy measure. The proposed methodology and experimental setup of our model is discussed in section 3.3 and 3.4 respectively. In section 3.5, we discuss the results. An unsupervised methodology for fuzzy interpretation of word polarity scores is described in section 3.6. Section 3.7 summarizes the findings of the chapter.

3.1 FUZZY ENTROPY

A fuzzy set A can be represented as $A = \{(x, \mu_A(x)\}, x \in U$ where x is the element from the universal set U and $\mu_A(x)$ is the membership of element x and there are total n elements. Different fuzzy measures can be applied on a fuzzy set: cardinality, energy, entropy, specificity, etc. (Pedrycz and Gomicide, 2007). In our chapter, we have applied entropy measure of fuzziness. Luca and Termini (1993) defined the fuzzy entropy of a fuzzy set as follows:

$$d(f) = H(f) + H\overline{(f)} \tag{3.1}$$

where $\overline{(f)}$ is defined as 1-(f), the complement of the fuzzy set. This entropy is based on Shannon Entropy given by:

$$H(p) = -\sum_{i=1}^{n} p_i \log p_i \tag{3.2}$$

where p_i is the probability of the ith element and i =1,2…n. The equation for Shannon Fuzzy entropy, H_A, is obtained by substituting p_i with membership $\mu_A(x_i)$ in (3.2) and using (3.1).

$$H_A = -\sum_{i=1}^{n} [\mu_A(x_i) \log \mu_A(x_i) + (1 - \mu_A(x_i)) \log(1 - \mu_A(x_i))] \tag{3.3}$$

Different forms of fuzzy entropy have been proposed in literature; these measures satisfy the basic properties of fuzzy entropy (Qing and Li, 2004; Deshmukh, 2011; Susan and Hanmandlu, 2013). The properties of fuzzy entropy are (Luca and Termini, 1993):

1. $H_A = 0\ if\ A\ is\ crisp\ set\ (\mu_A(x_i) = 0\ or\ 1\ if\ \forall\ x_i \in A)$
2. $H_A\ maximum\ if\ (\mu_A(x_i) = 0.5\ \ \forall\ x_i \in A)$
3. $H_A\ \geq\ H_A*, where\ H_A*\ is\ the\ entropy\ of\ A*, a\ sharpened\ version\ of\ A$
4. $H_A\ =\ \overline{H}_A, where\ \overline{H}_A\ is\ the\ entropy\ of\ complement\ set\ \bar{A}$
5. $H_A\ increases\ monotonically\ in\ the\ interval\ [0, 0.5]\ and$

   $decreases\ monotonically\ in\ the\ interval\ [0.5, 1]$
6. $H_A\ is\ a\ non-linear\ curve$
7. $H_A\ is\ a\ concave\ function\ of\ \mu_A(x_i)$

These properties can be observed graphically in Fig. 3.1. For this purpose, we have evaluated the values of corresponding to different values of $\mu_A(x_i)$, as shown in Table 3.1.

Table 3.1 Entropy values H_A Versus μ_A Values

$\mu_A(x_i)$	$H_A(\mu_A(x_i))$
0	0
0.1	0.3251
0.2	0.5004
0.3	0.6109
0.4	0.6730
0.5	**0.6931**
0.6	0.6730
0.7	0.6109
0.8	0.5004
0.9	0.3251
1	0

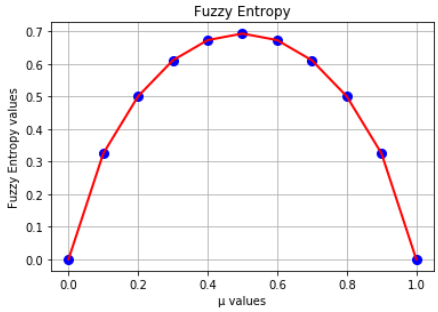


Fig. 3.1. *Graph of Fuzzy Entropy.*

In Fig. 3.1, it is apparently clear that fuzzy entropy value is 0 when $\mu_A(x_i) = 0$ or 1, it attains maximum value at $\mu_A(x_i) = 0.5$. It is symmetric because it attains same value for complement set, for example H_A (0.3) = 0.6109 and H_A (1-0.3) = H_A (0.7) = 0.6109. The sixth property of fuzzy entropy is its non-linearity. It is clearly visible in Fig. 3.1 that the slope of the graph is not constant. The third property states that $H_A \geq H_A*$ for the sharpened version of A. Sharpened means crisper, the entropy value is high when there is more fuzziness and is low when there is less fuzziness (crisp). We have applied this property in our work, to select the significant words in the text. In sentiment analysis, the words that carry more emotion have got either very positive or very negative sentiment scores. So, we deal with extreme boundaries of $\mu_A(x_i)$ that correspond to low fuzzy entropy values. We remove the words which carry high fuzzy entropy values as they are insignificant and they don't contribute to the evaluation of sentiment of the text. In doing so, we take the advantage of properties 1-7 listed above since the non-linearity of the fuzzy entropy curve ensures that very few words are shortlisted. The symmetricity of the curve ensures that the minimum strength of positive and negative shortlisted scores remains the same.

3.2 Proposed Approach

In this chapter, a supervised model which grades words for sentiment quotients using fuzzy entropy has been proposed. K-means clustering is used to segregate the important words versus non-significant words and the significant words are finally trained using LSTM classifier. The shortlisted words in the text are extracted in an automatic manner without any human intervention. Our model emulates the way the human brain thinks, by giving attention to significant words only. This section discusses the methodology adopted in this chapter, starting with fuzzy logic, related to sentiment analysis.

A fuzzy set A can be represented as $A = \{(x, \mu_A(x)\}, x \in U$ where *x* is the element from the universal set U and $\mu_A(x)$ is the membership of element *x*. In our approach, the Universe of Discourse is the set of all the words in each review. First of all, tokenization of review into sentences and sentences into words takes place. Words are tagged using NLTK POS Tagger (Bird *et al.*, 2009) and lemmatized. Only those words that are nouns, adjectives, verbs or adverbs are used further. If the lemmatized word exists in SentiWordNet, its scores are obtained. Each word has a positive and negative score ((3.6) and (3.7)) and can be interpreted as a fuzzy membership pertaining to the fuzzy sets Pos and Neg ((3.4) and (3.5)).

$$Pos = \{(a, \mu_{Pos}(a)\}, a \in X_i \tag{3.4}$$

$$Neg = \{(a, \mu_{Neg}(a)\}, a \in X_i \tag{3.5}$$

where *a* is the word. The membership functions $\mu_{Pos}(a)$ and $\mu_{Neg}(a)$ associated with the lemmatized word *a* are defined below. These are average positive and negative scores.

$$\mu_{Pos}(a) = \frac{\sum_{synsets} [syn.pos\_score()]}{length(synsets)} \tag{3.6}$$

$$\mu_{Neg}(a) = \frac{\sum_{synsets} [syn.neg\_score()]}{length(synsets)} \tag{3.7}$$

where $syn.pos\_score()$ and $syn.neg\_score()$ are the scores obtained from SentiWordNet; synset is the set of synonyms of each word present in SentiWordNet. We are assuming that the strength of human cognition can be measured using scores from SentiWordNet since human annotators have compiled this lexicon. The difference of positive and negative score is delta score $\mu_d(a)$, that is represented as:

$$\mu_d(a) = [\mu_{Pos}(a) - \mu_{Neg}(a)] \tag{3.8}$$

The normalization of delta score is fuzzy score *m* shown in (3.9), where $\mu_{Min}(a)$ and $\mu_{Max}(a)$ are maximum and minimum value of delta score in each sentence of review. Substituting the fuzzy score value m in the fuzzy entropy formula in (3), H_i is formulated in (3.10) and calculated for each word in a sentence.

$$m = \frac{\mu_d(a) - \mu_{Min}(a)}{\mu_{Max}(a) - \mu_{Min}(a)} \tag{3.9}$$

$$H_i = -[m \log(m + eps) + (1-m)) \log((1-m) + eps)] \tag{3.10}$$

The fuzzy entropy for each word is calculated for all sentences in a review. k-means clustering (MacQueen, 1967) is applied to these entropy values to segregate them into two clusters: low entropy and high entropy. Centroids (c1 and c2) of each cluster are calculated, and their mean, threshold, are computed, as shown in (3.11).

$$threshold = \frac{c1 + c2}{2} \tag{3.11}$$

We are interested in the entropy values less than this *threshold*, the words corresponding to these entropy values are selected. These are the significant words that contribute to evaluating the sentiment of the review. This *threshold* cuts the fuzzy entropy graph into two halves, the lower entropy values are selected as shown in Fig. 3.2. Suppose the computed *threshold*=0.4, entropy values less than 0.4 are selected, these values are marked in bold. k-means clustering is applied to all fuzzy entropy values of a review, it divides these values into two clusters. The cluster with values less than the threshold is selected as depicted in Fig. 3.3 Algorithm1 describes the steps to extract shortlisted words from a review. These shortlisted words are trained on the LSTM network to evaluate the sentiment of all reviews. The overall process flow for the Sentiment Cognition model is presented in Fig 3.4.

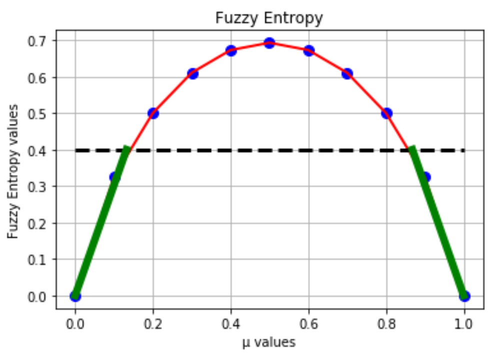


Fig. 3.2. *Graph of Fuzzy Entropy with threshold=0.4.*

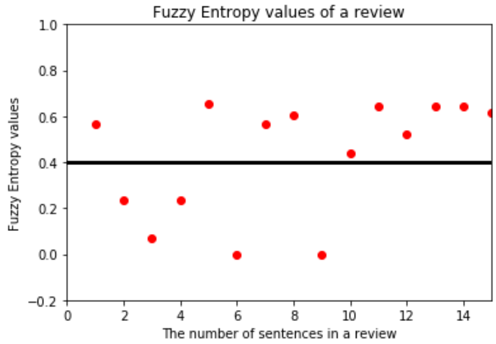


Fig. 3.3. *k-means clustering applied on entropy values of a review.*

Algorithm1 GetShortlistedWords

1: function **GetShortlistedWords**(review)
2: call **GetFuzzyScore**(review)

  1: function **GetFuzzyScore**(review)
  2: sentenceTokens = nltk. sent_tokenize(review)
  3: for each sentenceToken in sentenceTokens do
  4: processed_sentenceToken: Pre-Process sentenceToken
  5: wordTokens = [nltk.word_tokenize(processed_sentenceToken)]
  6: for each wordToken in wordTokens do
  7: POS Tagging wordToken
  8: lemmatizedWordToken = Lemmatize wordToken
  9: if ((lemmatizedWordToken is Noun) OR (lemmatizedWordToken is Adjective) OR (lemmatizedWordToken is Verb) OR (lemmatizedWordToken is Adverb)) AND (lemmatizedWordToken exists in SentiWordNet)
  10: Calculate average scores for wordToken using (6) and (7)
  avgPosScore = average Positive score;
  avgNegScore = average Negative scores;
  11: deltaScore = avgPosScore - avgNegScore (difference of above scores) using (8)
  12: fuzzyScore = Normalize deltaScore using (9)
  13: return listFuzzyScore = fuzzy Score for each word in each sentence of the review
  14: end function

3: for each fuzzyScores in listFuzzyScore do
4: call **GetFuzzyEntropy**(fuzzyScores)
  1: function **GetFuzzyEntropy**(fuzzyScores)
  2: For each fuzzyScore in fuzzyScores do
  3: Calculate Fuzzy Entropy for fuzzyScore using (10)
  4: Return fuzzyEntropy
  5: End function
5: fuzzyEntopies.Append = fuzzyEntropy

6: return fuzzyEntropies

7: call **GetThresholdFromClustering**(fuzzyEntropies)
- 1: Function **GetThresholdFromClustering**(fuzzyEntropies)
- 2: Create two clusters of low and high fuzzy Entropies
- 3: Calculate centroid of low and high clusters
- 4: Calculate threshold = Mean of Centroids using (11)
- 5: Return Threshold

8: call **GetFuzzyEntropy**(fuzzyScores)
9: FuzzyEntropyScores.Append = fuzzyEntropy
10: for each fuzzyEntropyScore in FuzzyEntropyScores
11: if fuzzyEntropyScore < threshold
12: ShortlistedWordTokens.Append = wordToken corresponding to that fuzzyEntropyScore
13: return ShortlistedWordTokens
14: end function

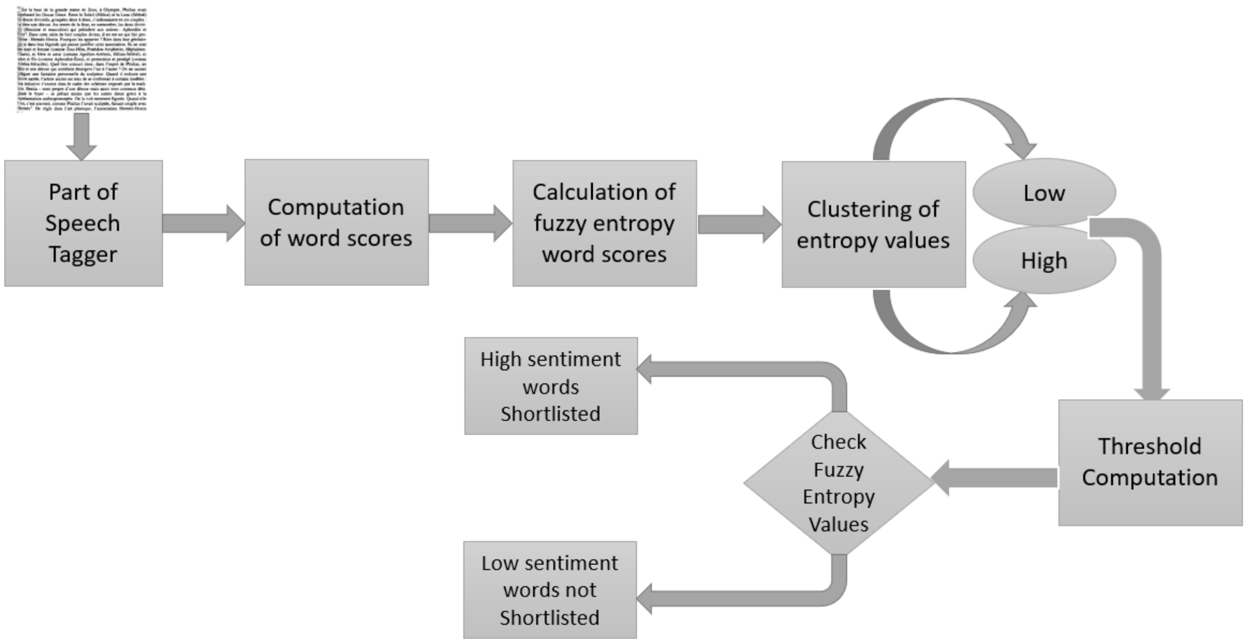


Fig. 3.4. *Overall Process flow of Sentiment Cognition Model.*

3.3 EXPERIMENTAL SETUP

In our approach, we have used two movie review datasets. The first one is the Movie Review Dataset IMDB. Each movie review has several sentences. The IMDB dataset has a training set of 25,000 labeled instances and a testing set of 25,000 labeled instances; the dataset has positive and negative labels balanced in training and testing set. Many researchers have experimented with the IMDB dataset for sentiment analysis. The second is polarity dataset v2.0 by Pang and

Lee[2] (Pang and Lee, 2004). This dataset contains 1000 positive and 1000 negative processed movie reviews. The train/test split of this dataset is 70-30 and we report 10-fold cross-validation results. The code for our work is available online[3]. Our model is implemented on Keras, which is a popular python library, to create an LSTM network for the classification of tasks. Text-preprocessing is implemented using NLTK (natural language processing toolkit) python library (Bird *et al.*, 2009). Both datasets were preprocessed to a dictionary size of 5,000.

An LSTM sequential model with a linear stack of layers is built up. The first layer is the embedding layer with a vocabulary size of 5,000 words with a zero-padded maximum sequence of 500 words per review and an embedding size of 32. In the second layer, 100 units of LSTM are added. The last layer is a fully-connected layer from the concatenated input to a single output. The layer is followed by a simple sigmoid activation function to limit the output between 0 and 1. The final yield is a single output.

The model is compiled using adam as an optimizer and binary cross-entropy as a loss function. This calculates loss with two classes (0 and 1). In our experiments, 0 represents negative sentiment and 1 represents positive sentiment. The loss is calculated on the single and final output of the dense layer. The best network used a batch size of 64 and 3 epochs.

3.4 RESULTS AND DISCUSSION

3.4.1 COMPARISON WITH OTHER SENTIMENT ANALYSIS TECHNIQUES

This section is about the discussion of the results of our proposed approach described in section IV. The comparison of our model and other models on the IMDB movie review dataset is shown in Table 3.2. Beineke *et al.*, 2004 have used Naïve Bayes (NB) model for sentiment classification. They have extracted a pair of derived features that are linearly combined to predict the sentiment. This method achieved the lowest accuracy of 65.9% as evident from Table 3.2. Pang *et al.*, 2002 have considered the aspect of sentiment classification based on a categorization study, with positive and negative sentiments. They have undertaken the experiment with three different machine learning algorithms, such as NB, SVM, and Maximum Entropy. They have used bag-of word features framework to implement the machine learning algorithms. As per their analysis, NB algorithm shows poor results among the three algorithms

[2] http://www.cs.cornell.edu/people/pabo/movie-review-data/,

[3] https://github.com/SrishtiVashishtha/Sentiment-Cognition-from-Words-Shortlisted-by-Fuzzy-Entropy

and SVM algorithm yields the best result with an accuracy of 82.9%. Matsumoto *et al.*, 2005 have used the syntactic relationship among words as a basis for document level sentiment analysis.

The combined unigram+bigram as word feature gives an accuracy of 84.6%. Both the traditional and the proposed LSTM models (Rahman *et al.*, 2016) plateaued at an accuracy of just over 80%, with a maximum of under 85%. Mullen and Collier have implemented an SVM classifier for Sentiment Analysis, few selected words are sent, and an accuracy of 86% was (Mullen and Collier, 2004). Mass *et al.*, 2011 proposed a hybrid supervised-unsupervised model when they created the IMDB review sentiment dataset. Their highest reported accuracy was 88.89% from their full model with additional unlabelled reviews and bag of words vectors. Tripathy *et al.*, (2016) experimented with multiple variations of machine learning models (NB, SVM, Maximum entropy, etc.) on the IMDB dataset. Their approach converted text reviews into numeric matrices using a count vectorizer and TF-IDF, which were then sent as input to machine learning algorithms for the purpose of classification. The best accuracy achieved in Tripathy *et al.*'s model was 88.94% from their combined Unigram-Bigram-Trigram model using SVM. Johnson and Zhang, (2016) proposed a model that implements an SVM classifier on bag of words obtained from text, followed by a deep neural architecture- LSTM.

Minaee *et al.*, (2019) developed a two-layer bi-LSTM, which gets the Glove embedding of words in a review, and predicts the sentiment for that. Minaee *et al.*, (2019) also deployed the word embeddings on a pre-trained Glove model and used a convolutional neural network (CNN) with 4 filter sizes (1,2,3,4), and 100 feature maps for each filter. The hidden representation is then followed by two fully-connected layers and fed into a SoftMax classifier. Kumar *et al.*, (2019) generated a set of Hybrid features by concatenating Machine Learning features with Lexicon features (Positive-Negative word count, Connotation). The effectiveness of the proposed work was demonstrated by applying four different classifiers such as SVM, KNN, Maximum Entropy and Naïve Bayes; among these Maximum Entropy with correlation demonstrated the best output. Our proposed Fuzzy Sentiment Cognition model has achieved the highest accuracy of 89.80% in the IMDB dataset.

We have performed a comparative analysis of our approach with different state-of-the-art methods of sentiment analysis on the polarity dataset by Pang and Lee, 2004 in Table 3.3. The SVM classifiers used in the comparison methods have gaussian (rbf) kernel with C=1.0 and auto-mode for gamma. Naïve Bayes classifier (Beineke *et al*., 2004) achieves 71.17% and LSTM (Rahman *et al.*, 2016) gives an accuracy of 69.45%. Our proposed model has achieved the highest accuracy of 78.5% in this dataset.

We have performed a t-test to determine if the results are statistically different. For t-tests, we have compared the predicted output of our method with other comparison methods. In t-tests, the larger the t score, the more difference there is between groups. The smaller the t score, the more similarity there is between groups. A p-value is a probability that the results from the sample data occurred by chance. Low p-values are good; They indicate your data did not occur by chance. Since our t-scores are generally high and p-values are generally low, this implies the results of these comparison methods are statistically different compared to our method's results. Our t-test results are shown for the IMDB dataset in Table 3.2 and for the polarity dataset by Pang and Lee in Table 3.3.

Very few works have considered downsizing the sample set based on cognitive principles. We have implemented our model by generating cognitive words as features using fuzzy entropy measures, and these cognitive shortlisted words are fed as input to the LSTM network. The application of fuzzy logic provides fuzzy reasoning and LSTM layers are able to use the information for a better interpretation of sentiment while classifying the reviews. LSTM is a more complicated function that learns to control the flow of information so as to prevent the vanishing gradient and allow the recurrent layer to more easily capture long-term dependencies. It is also proficient in learning sequence characteristics due to its ability to acquire syntax features of linguistics. The accuracy of our proposed model surpasses the previously published models.

Table 3.2 Comparison of our method compared to other sentiment techniques on IMDB dataset

Models	**IMDB Accuracy (in %)**	**t-test Statistic Score**	**p-value**
(Beineke *et al*., 2004) **Naïve Bayes**	65.9	13.03	$9.31e^{-39}$
(Pang *et al.*, 2002) **Unigram SVM**	82.9	-23.33	$8.389\ e^{-120}$
(Matsumoto *et al.*, 2005) **Unigram+Bigram**	84.6	-23.89	$1.821\ e^{-125}$
(Rahman *et al.*, 2016) **LSTM**	85	-18.04	$1.574\ e^{-22}$
(Mullen and Collier, 2004) **SVM**	86	-227	0
(Mass *et al.*, 2011) **Bag of Words**	88.89	-16.47	$8.10e^{-61}$
(Tripathy *et al.*, 2016) Unigram+Bigram+Trigram **SVM**	88.94	-23.96	$3.388\ e^{-125}$
(Johnson and Zhang, 2016) **SVM** on BOW + **LSTM**	88.7%	-18.33	$6.24e^{-92}$
(Minaee *et al*., 2019) **BiLSTM Model**	89%	-20.27	$5.12e^{-104}$
(Minaee *et al*., 2019) **CNN Model**	89.3%	-14.59	$3.86e^{-125}$
(Kumar *et al*., 2019) **Hybrid Model- Max Entropy**	83.93%	-19.53	$1.57e^{-113}$
Fuzzy Sentiment Cognition model	89.80		

Table 3.3 Comparison of our method compared to other sentiment techniques on Pang and Lee dataset

Models	**Pang and Lee Accuracy (in %)**	**t-test Statistic Score**	**p-value**
(Beineke *et al*., 2004) **Naïve Bayes**	71.17	3.543	0.0004
(Pang *et al*., 2002) **Unigram SVM**	68	12.37	$3.31\ e^{-33}$
(Matsumoto *et al*., 2005) **Unigram+Bigram**	61.5	17.51	$2.37\ e^{-61}$
(Rahman *et al*., 2016) **LSTM**	69.5	-0.01	0.984
(Mullen and Collier, 2004) **SVM**	49.7	43.66	$5.7e^{-250}$
(Mass *et al*., 2011) **Bag of Words**	68.5	14.08	$8.46\ e^{-42}$
(Tripathy *et al*., 2016) Unigram+Bigram+Trigram **SVM**	49.6	44.23	$3.5e^{-254}$
(Minaee *et al*., 2019) **BiLSTM Model**	75.3	2.68	$2.01e^{-201}$
(Minaee *et al*., 2019) **CNN Model**	76	12.7	$6.34e^{-194}$
(Vashishtha and Susan, 2020b) **Fuzzy Cardinality- SWN**	63.5	17.21	0.00
(Vashishtha and Susan, 2020b) **Fuzzy Cardinality AFINN**	65.45	16.44	0.00
Fuzzy Sentiment Cognition model	78.5		

3.4.2 SENTIMENT COGNITION FROM SHORTLISTED WORDS

3.4.2.1 Positive review example

We here demonstrate sentiment cognition from shortlisted words using our approach. The shortlisted words in a positive movie review are marked in bold as shown in Fig. 3.5. We can clearly observe words like "great", "hard", "shines", "best", etc. are conveying positive emotion, thereby this review is classified as positive.

Bobbie Phillips, who in her own right has amassed a **great** list of credits as a **hard** working Hollywood actress, **shines** in this third installment of UPN and Village Roadshow is Chameleon series. Bobbie **delivers** this **role** to the viewers with **great** panache. The action and **stunts** were the **best** in the series.

Fig. 3.5. *A positive movie review*

3.4.2.2 Negative review example

Similarly, we next demonstrate the evaluation of negative sentiment cognition from shortlisted words. Fig. 3.6 depicts the shortlisted words marked in bold. The words "shame", "bad", "forget" and "not" convey negative emotion, hence this review is classified as negative.

It **is** a **shame**. There is an **interesting** idea here, but it gets **completely** lost in a confusion of Commodore 64 style computer effects and **bad** storytelling. It should be a fairly straightforward hunter/hunted kind of story, but the director and/or the writer **seem** like they **forgot** what the movie was **supposed** to be when they were about three days into shooting. **Not** much to **recommend** here, even the title **seems** to propel it into obscurity.

Fig. 3.6. *A negative movie review*

3.4.2.3 Challenge review example

There are some reviews where evaluating sentiment cognition from shortlisted words becomes difficult. Like in Fig. 3.7, the shortlisted words are “like”, “liked”, “not” and “artistic” are marked in bold. Here, some simple emotional words are left out due to the presence of strong emotional words (“not”). The review is classified as negative.

I usually ignore whatever rating IMDb has when looking up a movie because I think I might **like** it anyway or whatever and I should at least give it a chance, but this time I wish I would paid attention. I know some people **liked** it, and I am **not** trying to say that they should **not**. The actors were either positively wooden or way over the top, and the film quality was awful, fuzzy and grainy and bland and **not** in an **artistic** way at all.

Fig. 3.7. *A challenging movie review*

3.4.3 SENSITIVITY ANALYSIS FOR THRESHOLD SELECTION

In this section, we discuss the sensitivity analysis of threshold selection. This threshold is based on k-means clustering of fuzzy entropy values. The following are different cases of shortlisted words under different thresholds, used to slice the fuzzy entropy graph for a movie review from the IMDB dataset. All the values below the threshold are selected. Fig. 3.8 depicts a sample review for the experiment.

the movie had extremely rich colour, in the way the house was decorated, in the clothes that the characters wore, in the open-skied daylight scenes that is in contrast to most horror movies, which, typically makes use of desaturated tones and gloomy environs (think Dark Water, which is another show i like) that gives this film a sense of aesthetics and joy when it was not in its, more, gripping moments.

Fig. 3.8. *A sample movie review*

3.4.3.1 Threshold is too low (0.1-0.2)

When the threshold is too low (0.1-0.2), very few words are shortlisted, hence polarity of review can't be determined. The shortlisted words are shown in bold in Fig. 3.9 and the corresponding fuzzy entropy graph with threshold 0.2 is shown in Fig. 3.10.

the movie had extremely rich colour, in the way the house was decorated, in the clothes that the characters wore, in the open-skied daylight scenes that is in contrast to most horror movies, which, typically makes use of desaturated tones and gloomy environs (think Dark Water, which is another show i like) that gives this film a sense of aesthetics and **joy** when it was **not** in its, more, gripping moments.

Fig. 3.9. *Shortlisted words in movie review when threshold is 0.2*

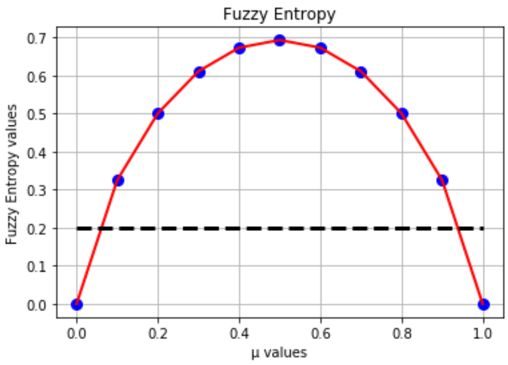


Fig. 3.10. *Fuzzy entropy graph when threshold is 0.2*

3.4.3.2 Threshold is too high (0.6)

When the threshold is 0.6, more shortlisted words are obtained. Not all of these words are important, some extra words have been shortlisted which will predict the wrong polarity of the sentence. The shortlisted words are shown in bold in Fig. 3.11 and the corresponding fuzzy entropy graph with a threshold of 0.6 is shown in Fig. 3.12. We can clearly see that when the threshold is too high the graph is divided unequally, and almost all fuzzy entropy values are selected.

> the movie had **extremely rich** colour, in the way the house was **decorated**, in the clothes that the **characters** wore, in the open-skied daylight scenes that is in contrast to most horror movies, which, **typically makes use** of desaturated tones and **gloomy** environs (**think** Dark Water, which is another show i like) that gives this film a **sense** of **aesthetics** and **joy** when it was **not** in its, more, **gripping** moments.

Fig. 3.11. *Shortlisted words in movie review when threshold is 0.6*

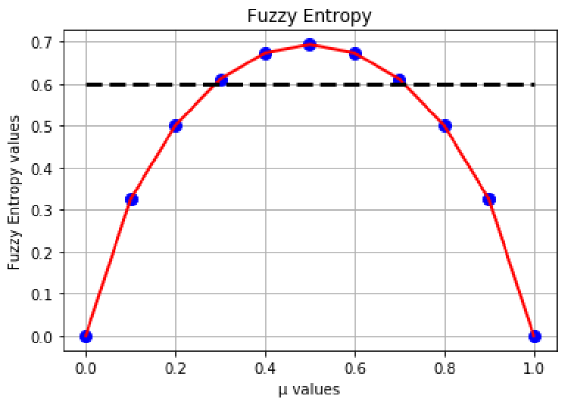


Fig. 3.12. *Fuzzy entropy graph when threshold is 0.6*

3.4.3.3 Threshold is medium (0.3-0.5)

When the threshold is 0.4, some shortlisted words are obtained which are appropriate to determine the polarity of the sentence correctly. The shortlisted words are shown in bold in Fig. 3.13 and the corresponding fuzzy entropy graph with a threshold 0.4 is shown in Fig. 3.14. We can observe this threshold divides the graph almost equally, the values lower than the threshold are the significant words.

the movie had **extremely rich** colour, in the way the house was **decorated**, in the clothes that the **characters** wore, in the open-skied daylight scenes that is in contrast to most horror movies, which, typically makes use of desaturated tones and **gloomy** environs (think Dark Water, which is another show i like) that gives this film a sense of **aesthetics** and **joy** when it was **not** in its, more, **gripping** moments.

Fig. 3.13. *Shortlisted words in movie review when threshold is 0.4*

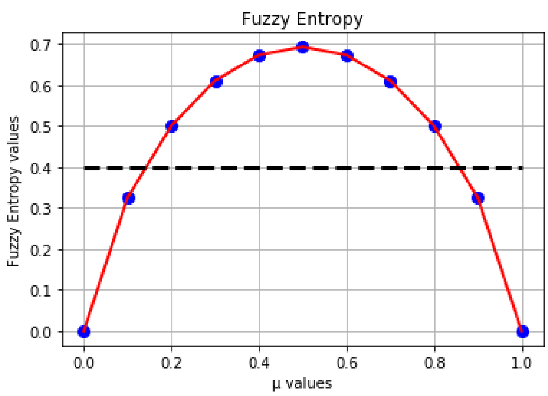


Fig. 3.14. *Fuzzy entropy graph when threshold is 0.4*

3.4.4 COMPARISON OF OUR MACHINE METHOD WITH THE HUMAN MIND

The experiment was validated by word annotations by 15 human observers. Each observer's shortlisted words for the sample review were noted down. We followed the protocol for Manual Annotation Task. The annotator was asked to glance through the paragraph and write with memory the significant words they remember. An ethical clearance certificate was obtained from all the collaborators stating that their inputs could be utilized for research purposes without revealing individual identities. Their motive for shortlisting words was not based on sentiment analysis. The compiled list of all shortlisted words for the positive review with their frequencies is shown in Fig. 3.15. We can infer that many people perceived negative words as well, like "horror" or "Dark" but our algorithm can correctly identify the positive sentiment words like "extremely", "rich", "joy", etc.

We compared the shortlisted word list by machine (our fuzzy method) with each of the human annotators and computed the Precision, Recall, and F-scores. These are presented in Table 3.4. We also predicted the sentiment for these shortlisted words using our fuzzy method (machine). We can clearly observe that the machine shortlisted words depicts positive sentiment and out of 15 only 6 predicted positive sentiment, rest predicted negative. Hence our fuzzy method (machine) predicts the correct sentiment at a higher rate than human annotators. In Table 3.4, Ma stands for Machine and HA stands for Human Annotators.

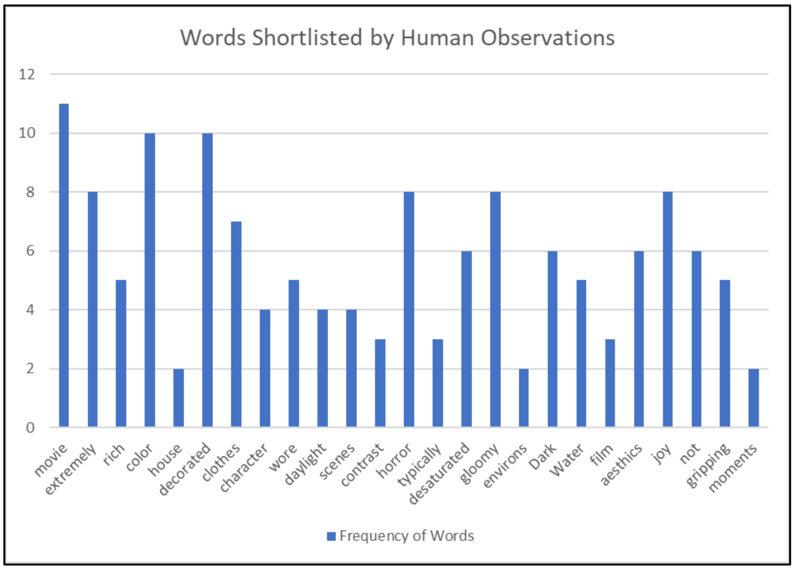


Fig. 3.15. *Shortlisted words by human observations*

Table 3.4 Comparison of our machine method compared with human mind

	Ma	HA1	HA2	HA3	HA4	HA5	HA6	HA7	HA8	HA9	HA10	HA11	HA12	HA13	HA14	HA15
Precision		0.56	0.22	0.44	0.44	0.44	0.33	0.33	0.33	0.22	0.22	0.33	0.56	0.44	0.56	0.44
Recall		0.56	0.2	0.44	0.4	0.4	0.38	0.38	0.33	0.22	0.29	0.3	0.5	0.36	0.44	0.4
F-Score		0.56	0.21	0.44	0.42	0.42	0.35	0.35	0.33	0.22	0.25	0.32	0.53	0.4	0.42	0.42
Sentiment	**Pos**	Pos	Neg	Pos	Pos	Pos	Neg	Pos	Pos	Neg	Neg	Neg	Neg	Neg	Neg	Neg

3.5 Fuzzy Interpretation of Word Polarity Scores

We have proposed an unsupervised algorithm to analyze the effect of fuzzification of word polarity sentiment scores. These word scores are obtained by deploying two lexicons: SentiWordNet and AFINN. Experiments are conducted on three benchmark datasets: polarity movie dataset by Pang-Lee, IMDB and hotel reviews dataset. The key highlights are: i) proposed an unsupervised fuzzy logic-based approach for sentiment analysis of textual reviews, ii) the proposed model formulated fuzzy cardinality as the measure for the evaluation of word polarity scores, iii) our model has two versions based on the sentiment lexicon deployed in the model, iv) comparison of our fuzzy cardinality approach with other non-fuzzy state-of-the-art approaches reveals the superiority of the proposed fuzzy approach.

3.5.1 Proposed Approach

Our proposed unsupervised approach for sentiment analysis of textual reviews has four major steps. The steps include tokenization, formulation of a bag of words model, formulation of fuzzy sentiment score and assigning polarity. We have two versions depending on the type of lexicon being applied. The two lexicons are SentiWordNet (Baccianella *et al.*,2010) and AFINN (Nielsen, 2011). Following is the description of the steps:

3.5.1.1 Tokenization & Lemmatization

Reviews are stored in a document. To work on documents, we need to first break down the document into sentences. Splitting up paragraphs into sentences is termed Sentence Tokenization. Tokenizing a sentence is the process of splitting a sentence into a list of words.

In other words, a tokenizer parses a sentence into a list of tokens (words). The output of the tokenization process will be stored in a dynamic list. Each word of the sentence is lemmatized.

3.5.1.2 Bag of Words

Bag of words term in SA refers to those keywords which are important for mining reviews, opinions, etc. Technically it is a model that transforms documents into vector (numerical) form, where each word in the document is assigned some score in the range of 0 to 1. This can be interpreted as a fuzzy membership pertaining to the fuzzy sets Pos and Neg. In our proposed model we have used NLTK Part of Speech (POS) Tagger to extract words which are nouns, adjectives, verbs or adverbs. The fuzzy sets *Pos* and *Neg* are represented as:

$$Pos = \{(a, \mu_{Pos}(a)\}, a \in X_i \tag{3.12}$$

$$Neg = \{(a, \mu_{Neg}(a)\}, a \in X_i \tag{3.13}$$

where a is the word, X_i is i^{th} set of Bag of words. If the total number of reviews is n; then a bag of words is created for each review. The membership functions μ_{Pos} and μ_{Neg} are renamed as μ_{swnPos} and μ_{swnNeg} for SentiWordNet lexicon (Baccianella *et al.*,2010) and; μ_{afPos} and μ_{afNeg} for AFINN lexicon (Nielsen, 2011).

$$\mu_{swnPos}(a) = \frac{\sum_{synsets} [syn.pos\_score()]}{length(synsets)} \tag{3.14}$$

$$\mu_{swnNeg}(a) = \frac{\sum_{synsets} [syn.neg\_score()]}{length(synsets)} \tag{3.15}$$

Eqs. (3.14) and (3.15) represent the fuzzy membership functions of a word for SentiWordNet lexicon where *syn.pos_score()* and *syn.neg_score()* are the scores obtained from SentiWordNet; synsets is the set of synonyms of each word present in SentiWordNet. The second version uses the AFINN lexicon. In (3.16), the score, μ_{af}, of each word is computed using AFINN and divided by five because the score by AFINN is between -5 and +5. Then the range of scores is checked, if it is greater than or equal to zero it is a positive score, μ_{afPos}, otherwise, it is a negative score, μ_{afNeg}. In (3.17) and (3.18) the computation of fuzzy membership functions of a word for AFINN lexicon is represented.

$$\mu_{af}(a) = \frac{af.score(a)}{5} \tag{3.16}$$

$$if\,(\mu_{af}(a) >= 0) then (\mu_{afPos}(a) = \mu_{af}(a)) \tag{3.17}$$

$$if\,(\mu_{af}(a) < 0) then (\mu_{afNeg}(a) = -\mu_{af}(a) \tag{3.18}$$

3.5.1.3 Formulation of proposed Fuzzy Cardinality measure

The fuzzy interpretation of word polarity scores for textual reviews using both lexicons is computed in the last section. We have proposed the Fuzzy Cardinality (Jang *et al.*, 1997) measure for evaluating the sentiment score of each review. The fuzzy sets *Pos* and *Neg* contain the positive and negative scores of words in a review. The positive cardinality of the *Pos* set and negative cardinality of the *Neg* set is calculated by summing all the elements in respective sets. This measure shows the strength of fuzzy sets. Following are the definitions of positive and negative Cardinality:

$$Pos\_cardinality = \sum_{j=1}^{l} \mu_{Pos}(a), a \in X_i \tag{3.19}$$

$$Neg\_cardinality = \sum_{j=1}^{l} \mu_{Neg}(a), a \in X_i \tag{3.20}$$

where l is the length of a review, a is the word, X is the set of Bag of Words and it belongs to the *i*th Bag of words.

3.5.1.4 Assigning Polarity

The binary polarity classification of a review is either positive or negative. A comparison of positive and negative cardinality (3.19 and 3.20) is performed to assign polarity to each textual review. If positive cardinality of a review is greater than or equal to the negative cardinality, it implies it is a positive review; otherwise, it is a negative review.

$$Polarity(i) = \begin{cases} P, Pos\_Cardinality \geq Neg\_Cardinality \\ N, otherwise \end{cases} \tag{3.21}$$

In (3.21), *Polarity(i)* is the polarity of i^{th} review and the labels for positive and negative polarity is *P* and *N*. Thereby using above comparison all the reviews are divided into two classes: Positive (*P*) and Negative (*N*). The process flow of proposed fuzzy approach is presented in Fig. 3.16.

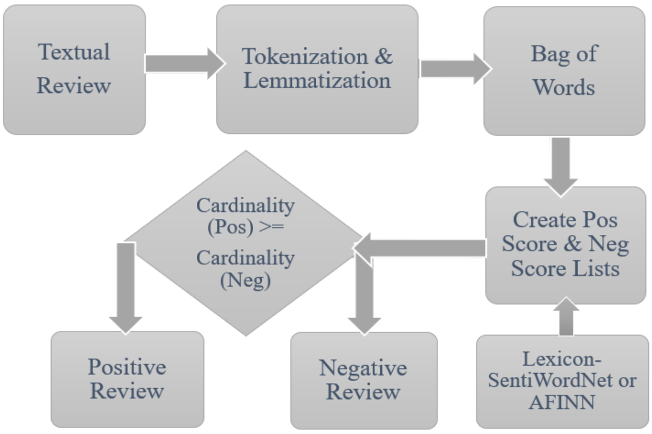


Fig. 3.16. *Process flow of proposed Fuzzy approach*

3.5.2 Experimental Setup And Implementation

The proposed unsupervised methodology for SA has been implemented in Python. The experiments are conducted on three datasets that contain online reviews by users. There are two movie datasets: polarity dataset v2.0 by Pang and Lee (Pang and Lee, 2004) and IMDB. The third dataset provides reviews of a single hotel[4]. These reviews were provided by the travellers who stayed in this hotel. All the datasets are freely accessible via the internet. The first dataset contains 2000 reviews, the IMDB dataset contains 50,000 reviews and the hotel review dataset has 38932 reviews. The dataset distribution of all the datasets according to each sentiment class is represented in Table 3.5.

[4] http://www.kaggle.com/harmanpreet93/hotelreviews

Two sentiment lexicons: SentiWordNet and AFINN are applied to compute the word polarity scores. Table 3.6 shows the positive and negative scores of a few words using the SentiWordNet lexicon. The positive words have a positive score higher than the negative score and it is vice-versa for negative words. There are some words that have zero value for both positive and negative scores, these are neutral words containing no sentiment, for example "Hotel" and "Staff". "Nice", "Helpful", "Clean" and "Beautiful" are positive words while "Filthy", "Difficult", "Disgusting" and "Accident" are negative words.

Table 3.5 Dataset Distribution for each class.

Dataset	Positive	Negative	Total
Pang-Lee Movie	1000	1000	2000
IMDB Movie	25000	25000	50000
Hotel Reviews	26521	12411	38932

Table 3.6 SentiWordNet Scores

Word	Pos Score	Neg Score
Hotel	0.0	0.0
Staff	0.0	0.0
Nice	0.15	0.0
Helpful	0.25	0.0
Clean	0.0278	0.0
Beautiful	0.3125	0.0
Filthy	0.0417	0.25
Difficult	0.0	0.3125
Disgusting	0.0625	0.3125
Accident	0.0	0.125

Table 3.7 AFINN Scores

Word	Score	Score/5	Pos Score	Neg Score
Hotel	0.0	0.0	0.0	-
Staff	0.0	0.0	0.0	-
Nice	3.0	0.6	0.6	-
Helpful	2.0	0.4	0.4	-
Clean	2.0	0.4	0.4	-
Beautiful	3.0	0.6	0.6	-
Filthy	-2.0	-0.4	-	0.4
Difficult	-1.0	-0.2	-	0.2
Disgusting	-3.0	-0.6	-	0.6
Accident	-2.0	-0.4	-	0.4

Table 3.7 depicts the scores of these same words computed using the AFINN lexicon. The words “Hotel” and “Staff” are neutral words in both the lexicons. Since the scores of “Nice”, “Helpful”, “Clean” and “Beautiful” are greater than zero their polarity is positive. Whereas the words “Filthy”, “Difficult”, “Disgusting” and “Accident” are negative words because their scores are less than zero. The positive and negative words have different scores in both lexicons, for example, the word “Nice” has a 0.15 positive score and 0 negative scores in the SentiWordNet lexicon while in the AFINN lexicon it has a 3.0 score. Hence, both lexicons can be interpreted in different ways to evaluate the polarity of a word and the score for the same word is different.

We have compared our two versions of the fuzzy approach with two non-fuzzy approaches for SA. The first comparison method is the Cavalcanti *et al.*, 2011 approach, where the sentiment score of each review is calculated by adding up the sentiment score of each word, *a*, and dividing it by length of each review, *l*, as shown in (3.22). Here the sentiment score of each word is the difference of the positive and negative score of each word computed using the SentiWordNet lexicon, as given in (3.14) and (3.15).

$$score(X_i) = \frac{\sum_{a \in X_i} [\mu_{Pos}(a) - \mu_{Neg}(a)]}{l}, a \in X_i \tag{3.22}$$

In the Cavalcanti *et al*. approach, the scores greater than or equal to zero denote the positive sentiment (P) and negative scores denote the negative sentiment (N). The second comparison method is Hutto and Gilbert's 2014 approach, this approach had created a VADER lexicon and its tool. This method computes various scores for the given input textual sentence. We apply the VADER approach to a review; the compound score is generated. The range of compound scores is checked for different polarity classes: positive and negative. The implementation of our proposed approach is available here[7].

3.5.2 Results And Discussion

An unsupervised fuzzy approach has been presented to evaluate the sentiment of online textual reviews. There are two versions of our fuzzy cardinality approach: SentiWordNet and AFINN. These versions are compared with two unsupervised non-fuzzy approaches: Cavalcanti *et al*., 2011 approach and Hutto and Gilbert's, 2014 approach. All the experiments are conducted on the three datasets of online reviews. The comparison of our versions of fuzzy cardinality versions with other methods, in terms of accuracy, is depicted in Table 3.8.

Table 3.8 Comparison of our versions of the Fuzzy Cardinality Approach with other methods.

Dataset	**Accuracy**			
	Cavalcanti *et al.*, 2011	**Hutto and Gilbert, 2014**	**Fuzzy Cardinality SentiWordNet Approach**	**Fuzzy Cardinality AFINN Approach**
Pang-Lee Movie	54.8%	63%,	63.5%	65.45%
IMDB Movie	52.87%	69.43%	64.13%	70.06%
Hotel Reviews	64.54%	76.1%	72.74%	76.2%

Experiments on Pang-Lee movie datasets reveal that our fuzzy cardinality AFINN version has achieved the highest accuracy of 65.45%, followed by the SentiWordNet version of 63.5%; comparable accuracy of 63% is gained by Hutto and Gilbert's approach and Cavalcanti *et al.*'

[7]https://www.github.com/SrishtiVashishtha/Fuzzy-Interpretation-of-Word-Polarity-Scores-for-Unsupervised-Sentiment-Analysis

s approach has acquired the least accuracy of 54.8%. For the IMDB movie dataset, our fuzzy cardinality AFINN version has achieved the highest accuracy of 70.06% and Hutto and Gilbert's approach achieved the second-highest accuracy of 69.43%. Our fuzzy cardinality AFINN version has gained the highest accuracy of 76.2% in the Hotel reviews dataset, which is comparable to that of Hutto and Gilbert's approach. Our fuzzy cardinality SentiWordNet version has scored higher accuracy compared to Cavalcanti *et al.*' s approach in all datasets.

From the results in Table 3.8, we can conclude that our fuzzy approach based on the AFINN lexicon has scored the highest accuracy in all the datasets, Cavalcanti *et al.*' s approach has gained the lowest accuracy in all the datasets. Whereas the results of Hutto and Gilbert's approach are comparable to our fuzzy cardinality versions.

3.6 CONCLUSION

Sentiment Analysis (SA) is a language processing task which is applied to extract the opinion expressed by humans in online reviews, social media, blogs, mobile applications, etc., and classify this opinion into polarities like positive or negative. In this chapter, we have proposed a model that extracts the high sentiment cognition words from the text in an automated way using fuzzy entropy measures for fuzzy reasoning. The fuzzy scores are extracted from the SentiWordNet lexicon that has been compiled with the help of human annotations. The clustering technique is applied to these fuzzy entropy values to divide them into two clusters of significant and non-significant words. We are interested in the cluster with low fuzzy entropy values since they correspond to the important words which contribute to the evaluation of the sentiment of text. Our approach emulates the way a human mind works by glancing at only a few key important words in the text, to determine the polarity of review. These shortlisted words are trained using an LSTM network. Our approach is implemented on two movie review datasets: IMDB and the polarity dataset by Pang and Lee. We have compared both datasets with other state-of-the-art approaches to SA. It has been observed that our model has achieved the highest accuracies of 89.8% in IMDB and 78.5% in Pang and Lee datasets, as compared to other methods. We propose to incorporate eye movement data in our future work.

In this chapter, a fuzzy logic-based technique is deployed to online reviews to compute the fuzzy sentiment score. Two sentiment lexicons- SentiWordNet and AFINN are used to compute the sentiment score of words. The key highlights are: i) proposed an unsupervised approach based on fuzzy logic for sentiment analysis of textual reviews, ii) the proposed model uses fuzzy cardinality as the measure for the evaluation of word polarity scores, iii) our model has two versions based on the sentiment lexicon deployed in the model, iv) our fuzzy cardinality approach is compared to non-fuzzy state-of-the-art methods.

Our proposed fuzzy methodology is better than non-fuzzy methods. This is because fuzzy deals with ambiguity in real-world problems. Our approach calculates the strength of average positive and negative scores of each word in each review and these scores are fuzzy. Thus, the strength of fuzzy sets gives better results than simple average scores. The application of fuzzy logic with NLP provides us with results that match human interpretation for SA. Our approach can be applied to all textual dataset that contains online or social media content, like Twitter datasets, product reviews dataset, any other customer review datasets, etc. The limitation of our work is that the scores of words are dependent on lexicons; some words which do not exist in lexicons cannot be processed further.

CHAPTER 4

DEVELOPMENT OF FUZZY TECHNIQUE FOR HIGHLIGHTING OF HIGH SENTIMENT COGNITION PHRASES

Sentiment Analysis is a process that aids in assessing the performance of products or services from user-generated online posts. In the present time, there are various websites that allow customers to post reviews about movies, products, events, or services, etc. This has led to the cumulative aggregation of a lot of reviews written in natural language. Prevailing factors such as the availability of online reviews and raised end-user expectations have motivated the evolution of opinion mining systems that can automatically classify customers' reviews. It is observed that in Sentiment Analysis (SA), highlighting the significant keyphrases which contribute to correct sentiment cognition is a tedious task. In this chapter[1], we have proposed an unsupervised sentiment classification system that comprehensively formulates phrases, and computes their senti-scores (sentiment scores) and polarity using the SentiWordNet lexicon and fuzzy linguistic hedges. Further, it extracts the keyphrases significant for SA using a fuzzy entropy filter and k-means clustering. We have deployed document level SA on online reviews using

:[1] The contents of this chapter are published in "Highlighting Keyphrases using Senti-Scoring and Fuzzy Entropy for Unsupervised Sentiment Analysis," in *Expert Systems with Applications* 169 (2021): 114323.

n-gram techniques, specifically combination of unigram, bigram and trigram. Experiments on two benchmark movie review datasets- polarity dataset by Pang and Lee and IMDB dataset, achieve high accuracy for our approach in comparison to the other state-of-the-art-approaches for phrase-level SA.

SA can be applied to text in two ways: word level and phrase level. Word level focuses on extracting and computing only words from the text and at phrase-level: extraction and computation of phrases-combination of words is done (Cambria and White, 2014). In text-based SA, the system can determine sentiment expressed by examining the words (Vashishtha and Susan, 2020 b), phrases, and dependencies among them. Further, these words and phrases can classify the given text into different sentiment classes. In this chapter, we have deployed the phrase level SA to classify online reviews into positive and negative polarity. There are primarily two types of machine learning techniques, generally used in SA, supervised and unsupervised learning techniques. In the supervised learning technique, the dataset is labeled and subsequently trained to obtain a reasonable output which helps in proper decision making. Unlike supervised learning, unsupervised learning processes do not need any label data; hence they cannot be processed at ease. This study demonstrates an unsupervised phrase-level SA which comprehensively formulates phrases, and computes their senti-scores (sentiment scores) and polarity using SentiWordNet lexicon and fuzzy linguistic hedges. These computed phrases are filtered out by the fuzzy measure- fuzzy entropy, and k-means clustering; finally, the senti-scores and polarity of selected phrases are used to determine the sentiment of the review.

The key highlights of this work are as follows:

1. An unsupervised phrase-level SA approach has been proposed to perform sentiment analysis on online reviews using n-gram techniques, specific a combination of unigram, bigram, and trigram.
2. Phrases are constructed comprehensively using part-of-speech (POS) Tagger, a list of concentrators, dilators, and negators. Their senti-scores and polarity are computed using SentiWordNet lexicon and fuzzy linguistic hedges.
3. Document-level SA on online reviews is executed by extracting high sentiment-bearing keyphrases filtered out by fuzzy entropy and k-means clustering, and finally computing the sentiment of the review.

4. The performance of our fuzzy technique is evaluated using the parameters of accuracy and f-score. The results indicate higher scores as compared to the state-of-the-art.

The rest of the chapter is organized as follows. Section 4.1 describes the proposed fuzzy approach for phrase-level SA. In Section 4.2, the experimental setup of our approach is discussed. The results are demonstrated in Section. 4.3. The overall conclusions are drawn in Section 4.4.

4.1 PROPOSED METHODOLOGY

4.1.1 MOTIVATION

There are millions of online reviews on the internet for various topics, events, products or services. Analyzing these reviews for SA is in demand for various organizations but it is a challenging task. Several works have tackled this issue by searching out different phrase patterns in text. They include all the extracted phrases for detecting the sentiment, some of these phrases are not important, and thereby the wrong sentiment is detected. This motivated us to extract only important phrases i.e. keyphrases. Extracting important phrases from a text document means highlighting the phrases that are significant for sentiment classification for determining the correct sentiment of the ext. Few works have attempted to extract key phrases, and most of them have implemented a non-fuzzy approach. The task of understanding human language requires a clear understanding of the knowledge of discourse. The computer needs to comprehend how things work in the real-world domain; this effort although very progressive has a limitation. There is an intelligence gap between a human and a machine. Fuzzy logic can be used to make the machine understand this intelligence gap in a better way because it deals with uncertainty, vagueness, or ambiguity factors that are present in human language (Zadeh, 2015). Thus, the incorporation of fuzzy logic module in sentiment analysis will improve the classification. In our work, we have used fuzzy linguistic hedges to calculate the senti-scores of keyphrases, and the fuzzy measure- fuzzy entropy, with k-means clustering for filtering out the high sentiment bearing keyphrases; only these selected phrases are used for sentiment classification. Most of the works in literature have implemented supervised classification for phrase-level SA, but we have proposed and implemented a novel unsupervised classification approach that doesn't require any training.

4.1.2 SENTI-SCORING OF WORDS

The first step in our proposed methodology is text pre-processing. It involves cleaning of text, since we are performing Sentiment Analysis, a data mining task, on online movie reviews, we need to format the text. This includes changing abbreviations, words like "can't", "don't", etc. are transformed to "can not", "do not"; some words like "we're", "I'd" are transformed to "we are," I would", etc. It also removes URL links "@" mentions, "#" hashtags, etc. The next step is to tokenize the text document into a list of tokens (words), followed by the lemmatization of words. NLTK POS (part-of-speech) tagger (Loper and Bird, 2002) is used to label each word with its POS. We need to know how to compute the senti-score of each word. In our approach, the Universe of Discourse, for fuzzy set, is the set of all the words in each review. Only those words that are adjectives or adverbs, called *character words*, are used further. Each word has one or more synonyms. Each synonym *s* has a positive and negative score ((4.3) and (4.4)) derived from the SentiWordNet lexicon (Baccianella *et al*., 2014) and can be interpreted as a fuzzy membership pertaining to the fuzzy sets *Pos* and *Neg* ((4.1) and (4.2)).

$$Pos = \{(s, \mu_{Pos}(s)\}, s \in X_i \tag{4.1}$$

$$Neg = \{(s, \mu_{Neg}(s)\}, s \in X_i \tag{4.2}$$

where s is the synonym of a lemmatized word and X_i is set of synonyms. The membership functions $\mu_{Pos}(s)$ and $\mu_{Neg}(s)$ associated with each synonym of the lemmatized word, a, are defined below.

$$\mu_{Pos}(s) = syn.pos\_score() \tag{4.3}$$

$$\mu_{Neg}(s) = syn.neg\_score() \tag{4.4}$$

$$\begin{aligned} &if \ldots \mu_{Pos}(s) > \mu_{Neg}(s), then \ldots \mu_{Pos}(s) \in P \\ &if \cdots \mu_{Pos}(s) < \mu_{Neg}(s), then \ldots \mu_{Neg}(s) \in N \end{aligned} \tag{4.5}$$

where $syn.pos\_score()$ and $syn.neg\_score()$ are the scores of each synonym obtained from SentiWordNet. *P* and *N* are two sets denoting the scores of the positive and negative synonyms respectively that are initialized to null sets for each lemmatized word. For each synonym of the lemmatized word, *a*, the positive and negative scores (obtained in (4.3) and (4.4)) are compared to check whether the synonym is positive or negative as depicted in (4.5). The positive synonym is assigned score, *p*, while the negative synonym is assigned a score, *n* in (4.5). Let the count of positive synonyms be *count_pos* which is equal to the number of elements in *P*, and let the count of negative synonyms be denoted by *count_neg* which is equal to the number of elements in *N*. These are computed for each lemmatized word *a*. In our proposed work we are utilizing the n-gram technique, thus this lemmatized word, *a*, is a unigram. The senti-score, *u*, of the lemmatized word that indicates whether the unigram is positive, negative, or neutral is computed by (4.6), (4.7), or (4.8) depending on the values of *count_pos* and *count_neg*. The polarity, *pol,* of the lemmatized word is computed in (4.9); for a positive word, it is +1, for negative word it is -1 and, for a neutral word it is 0.

$$\begin{aligned}&if \ldots count\_pos > count\_neg\\&then \ldots u = \frac{\sum_{p \in P} p}{count\_pos}\end{aligned} \tag{4.6}$$

$$\begin{aligned}&if \ldots count\_neg > count\_pos\\&then \ldots u = \frac{\sum_{n \in N} n}{count\_neg}\end{aligned} \tag{4.7}$$

$$\begin{aligned}&if \ldots count\_pos = count\_neg\\&then \ldots u = 0\end{aligned} \tag{4.8}$$

$$\begin{aligned}&if \ldots count\_pos > count\_neg, then \ldots pol = 1\\&if \ldots count\_pos < count\_neg, then \ldots pol = -1\\&if \ldots count\_pos = count\_neg, then \ldots pol = 0\end{aligned} \tag{4.9}$$

4.1.3 FORMULATION OF PHRASES

We have formulated phrases using the n-gram technique where a n-gram has length of 1,2 or 3 i.e. unigram, bigram or trigram. Unigram, in our case, is either an adjective or adverb. The computation of senti-score for unigram has been discussed in section 3.2. These unigrams are usually preceded by modifiers. In linguistics, modifiers are words that change the meaning of the unigram, these modifiers can modify the sentiment of unigram also. Usually these are adverbs or adjectives like '*very'*, *'more or less*', '*fairly*', etc. Modifiers, also known as linguistic hedges, have the capability of changing the membership function for a basic term (Ross, 2004). Linguistic hedges have been utilized for fuzzy sentiment analysis in (Srivastava and Bhatia, 2013; Dalal and Zaveri, 2014; Vashishtha and Susan, 2018). Modifiers are of three types: concentrators, dilators and negators. Concentrators increase the senti-score of unigrams, while dilators decrease the senti-score and negators flip the polarity of unigrams. The concentrators, dilators and negators used in our work are listed below.

Concentrator = {'very', 'extremely', 'absolutely', 'highly', 'incredibly', 'positively', 'significantly'}

Dilator= {'quite', 'hardly', 'somewhat', 'almost', 'more or less', 'slightly', 'approximately', 'occasional'}

Negator= {'not', 'never', 'not so', 'neither', 'nor', 'nothing', 'although'}

Table 4.1. Formulation and Senti-Scoring of Phrase Patterns.

N-gram	word3	word2	word1	Senti- Score	Polarity
Unigram			Adj or Adv	u: Positive word: $\frac{\sum_{p\in P} p}{count\_pos}$ or Negative word: $\frac{\sum_{n\in N} n}{count\_neg}$ or Neutral word: 0	pol: Positive word: Polarity=+1 or Negative word: Polarity=-1 or Neutral word: Polarity=0
Bigram_1		Concentrator	Adj or Adv	if $u < 1$: $b_1 = u^{0.5}$ if $u > 1$: $b_1 = u^2$	$pol(b_1) = pol(u)$
Bigram_2		Dilator	Adj or Adv	if $u < 1$: $b_2 = u^2$ if $u > 1$: $b_2 = u^{0.5}$	$pol(b_2) = pol(u)$
Bigram_3		Negator	Adj or Adv	if $u < 1$: $b_3 = u^{1.5}$ if $u > 1$: $b_3 = u^{0.75}$	$pol(b_3) = -pol(u)$
Trigram_1	Concentrator	Concentrator	Adj or Adv	if $b_1 < 1$: $t_1 = b_1^{0.5}$ if $b_1 > 1$: $t_1 = b_1^2$	$pol(t_1) = pol(b_1)$
Trigram_2	Dilator	Concentrator	Adj or Adv	if $b_1 < 1$: $t_2 = b_1^2$ if $b_1 > 1$: $t_2 = b_1^{0.5}$	$pol(t_2) = pol(b_1)$
Trigram_3	Negator	Concentrator	Adj or Adv	if $b_1 < 1$: $t_3 = b_1^{1.5}$ if $b_1 > 1$: $t_3 = b_1^{0.75}$	$pol(t_3) = -pol(b_1)$

$Trigram_4$	Concentrator	Dilator	Adj or Adv	if $b_2 < 1$: $t_4 = b_2^{0.5}$ if $b_2 > 1$: $t_4 = b_2^2$	$pol(t_4) = pol(b_2)$
$Trigram_5$	Dilator	Dilator	Adj or Adv	if $b_2 < 1$: $t_5 = b_2^2$ if $b_2 > 1$: $t_5 = b_2^{0.5}$	$pol(t_5) = pol(b_2)$
$Trigram_6$	Negator	Dilator	Adj or Adv	if $b_2 < 1$: $t_6 = b_2^{1.5}$ if $b_2 > 1$: $t_6 = b_2^{0.75}$	$pol(t_6) = -pol(b_2)$
$Trigram_7$	Negator	Negator	Adj or Adv	if $b_3 < 1$: $t_7 = b_3^{1.5}$ if $b_3 > 1$: $t_7 = b_3^{0.75}$	$pol(t_7) = pol(b_3)$

We have formulated three bigrams and seven trigrams; these are depicted in Table 4.1. In Table 4.1, word1 depicts the *character_word*, which is either an adjective or adverb; word2 is the word occurring just before word1, similarly word3 is the word occurring just before word2 in the text document. If word2 is a concentrator or dilator or negator then bigrams are formed. Example of bigrams are 'very good', 'slightly good', and 'not bad' for concentrator, dilator and, negator respectively. Further we ,investigated if $Bigram_1$ is preceded by the concentrator, dilator or negator then the trigrams- $Trigram_1$, $Trigram_2$ or $Trigram_3$ are created. Similarly, we checked for $Bigram_2$ thereby generating $Trigram_4$, $Trigram_5$ or $Trigram_6$. $Trigram_7$ is formulated when a pattern of two negator words co-occur with a *character_word*. Some examples of trigrams are: 'extremely very bad, 'not very good', 'not hardly worth', 'nothing nothing good', etc. We have checked the logical checks for consistency. There is no conflict and redundancy in phrase patterns. In our work, all the eleven patterns are unique and any two rules do not yield conflicting or same results. We have also checked the completeness. There are no missing rules. We have created unigrams, bigrams and trigrams by checking all feasible patterns.

4.1.4 SENTI-SCORING OF PHRASES

The phrases obtained in the last section undergo the process of senti-scoring to assign each phrase a senti-score and polarity. The senti-score, *u*, and polarity, *pol*, of *character_word*, has been calculated in section 4.1.2 in (4.6-4.9). The phrases formed due to the presence of modifiers are bigrams and trigrams. The formulae for computing the senti-scores and polarity of phrases are given in (4.10- 4.19), where *u* is the senti-score of the unigram (*character_word*), b_1-b_3 are bigram senti-scores, t_1-t_7 are trigram senti-scores and *pol* depicts the polarity. The detailed formulation and senti-scoring of all phrase patterns are presented in Table 4.1. As observed from the senti-scores computation in Table 4.1, the concentrators tend to increase the senti-scores of unigrams by power of 2, but if the unigram scores are in range (0,1) then the scores increase by a power of reciprocal of 2 i.e. 0.5 (Ross, 2004). Similarly, dilators decrease the senti-scores of unigrams by power of 0.5, but if the unigram scores are in range (0,1) then the scores decrease by a power of reciprocal of 0.5 i.e. 2 (Ross, 2004). In the case of negators, it flips the polarity and tends to change the senti-score of unigrams in such a way that it is more than the dilator's effect and less than the concentrator's effect. Hence, we have introduced a novel modified score of negators, which is a power of 0.75 (0<0.75<1) of senti-score; but if the unigram scores are in range (0,1) then the modified score is a power of 1.5 (2<1.5<0.5) of senti-score. The reason behind choosing these constant values is that the modified score of the negator is greater than the dilator modified score and less than the concentrator modified score. We have performed experiments on alternative constant values, in our rules, for 0.75 constant in the range of (0,1) and 1.5 constant in the range of (0.5,2.0). The results of sensitivity analysis for both datasets - Pang-Lee and IMDB, used in our experiments, reveal that the highest accuracy is achieved at the values of 0.75 and 1.5 in the two graphs, respectively. The sensitivity analysis graphs are shown in Fig. 4.1 and Fig. 4.2 for the 0.75 constant and the 1.5 constant, respectively.

$$\begin{aligned}&if \ldots u<1, then \ldots b_1=u^{0.5}, pol(b_1)=pol(u)\\&else \ldots b_1=u^2\end{aligned} \tag{4.10}$$

$$\begin{aligned}&if \ldots u<1, then \ldots b_2=u^2, pol(b_2)=pol(u)\\&else \ldots b_2=u^{0.5}\end{aligned} \tag{4.11}$$

$$\begin{aligned}&if \ldots u<1, then \ldots b_3=u^{1.5}, pol(b_3)=-pol(u)\\&else \ldots b_3=u^{0.75}\end{aligned} \tag{4.12}$$

$$\begin{aligned}&if \ldots b_1<1, then \ldots t_1=b_1^{0.5}, pol(t_1)=pol(b_1)\\&else \ldots t_1=b_1^2\end{aligned} \tag{4.13}$$

$$\begin{aligned}&if \ldots b_1<1, then \ldots t_2=b_1^2, pol(t_2)=pol(b_1)\\&else \ldots t_2=b_1^{0.5}\end{aligned} \tag{4.14}$$

$$\begin{aligned}&if \ldots b_1<1, then \ldots t_3=b_1^{1.5}, pol(t_3)=-pol(b_1)\\&else \ldots t_3=b_1^{0.75}\end{aligned} \tag{4.15}$$

$$\begin{aligned}&if \ldots b_2<1, then \ldots t_4=b_2^{0.5}, pol(t_4)=pol(b_2)\\&else \ldots t_4=b_2^2\end{aligned} \tag{4.16}$$

$$\begin{aligned} &if \ldots b_2 < 1, then \ldots t_5 = b_2^{2}, pol(t_5) = pol(b_2) \\ &else \ldots t_5 = b_2^{0.5} \end{aligned} \tag{4.17}$$

$$\begin{aligned} &if \ldots b_2 < 1, then \ldots t_6 = b_2^{1.5}, pol(t_6) = -pol(b_2) \\ &else \ldots t_6 = b_2^{0.75} \end{aligned} \tag{4.18}$$

$$\begin{aligned} &if \ldots b_3 < 1, then \ldots t_7 = b_3^{1.5}, pol(t_7) = pol(b_3) \\ &else \ldots t_7 = b_3^{0.75} \end{aligned} \tag{4.19}$$

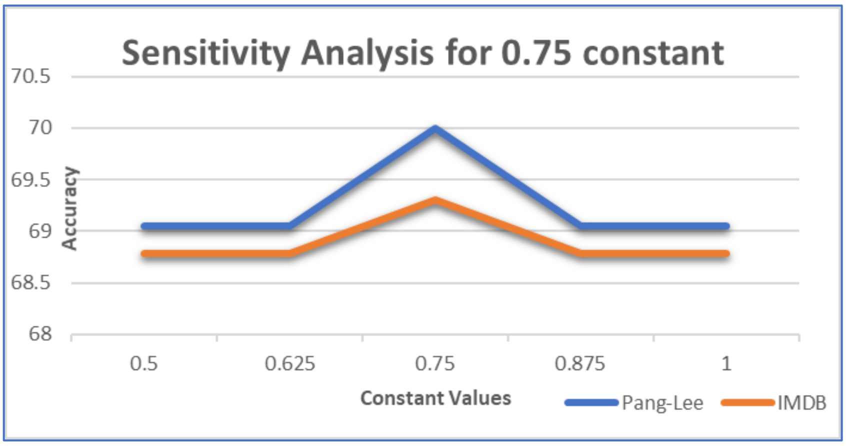


Fig. 4.1. Sensitivity analysis with accuracy values for 0.75 constant.

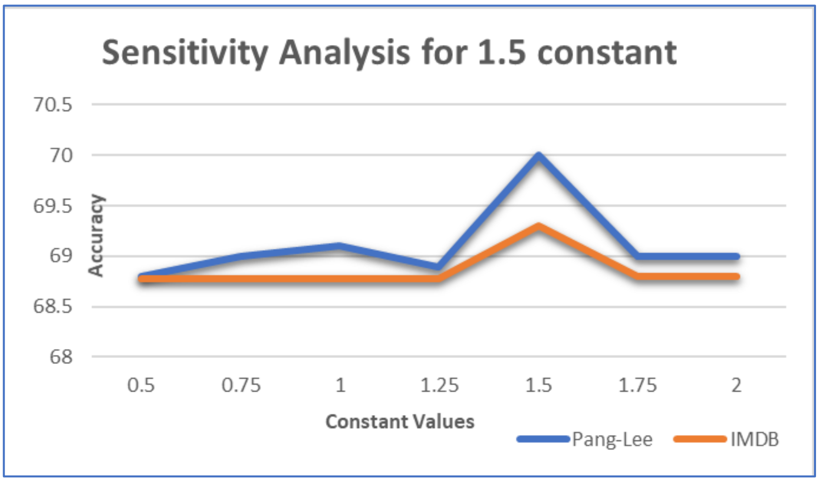


Fig.4.2. Sensitivity analysis with accuracy values for 1.5 constant.

4.1.5 FUZZY ENTROPY FILTER FOR EXTRACTING KEYPHRASES

The next step is to select the keyphrases which are significant for sentiment classification. We have applied the fuzzy entropy filter for this purpose that was used in our recent work for shortlisting single words that have high sentiment quotient (Vashishtha and Susan, 2019 b). In order to compute the entropy values for each phrase, normalization of senti-scores of phrases is performed. This is necessary so that senti-scores of unigrams, bigrams, and trigrams are normalized to [0,1], this processing is called scaling. Scaling can be performed in two ways- the same scale for all phrases or a different scale for each n-gram. The results for both versions are discussed in section 4; the same scale version performs better than the different scale version. The normalized senti-score is *m* as shown in (4.20), where $\mu_{Min}(ph)$ and $\mu_{Max}(ph)$ are maximum and minimum value of senti-score of a phrase *ph* in each sentence of the review.

$$m = \frac{\mu_d(ph) - \mu_{Min}(ph)}{\mu_{Max}(ph) - \mu_{Min}(ph)} \tag{4.20}$$

Different forms of fuzzy entropy have been proposed in the literature (De Luca and Termini, 1972; Susan and Hanmandlu, 2013). We have applied the entropy measure of fuzziness. De Luca and Termini (1972) defined the fuzzy entropy of a fuzzy set as

$$H_{ph} = -[m\log(m) + (1-m)\log(1-m)] \tag{4.21}$$

Here *m* is the normalized senti-score value; H_{ph} is fuzzy entropy value for each phrase *ph* in a review. The fuzzy entropy measure followed by k-means clustering has been utilized in our recent work in (Vashishtha and Susan, 2019 b) to shortlist significant words in a text. Extending the idea in our work on keyphrases, the fuzzy entropy value for each phrase is calculated for all reviews in the text document. k-means clustering (MacQueen, 1967) is applied to these fuzzy entropy values to segregate them into two clusters: low entropy and high entropy. Centroids (*c1* and *c2*) of each cluster are calculated, and their mean, *threshold*, is computed, as shown in (4.22).

$$threshold = \frac{c1 + c2}{2} \tag{4.22}$$

We are interested in the entropy values less than this *threshold*, the words corresponding to these entropy values are selected. These indicate the important keyphrases that contribute to evaluating the sentiment of review. The fuzzy entropy-based thresholding in (Vashishtha and Susan, 2019 b) is explained below for reference. This *threshold* cuts the fuzzy entropy graph into two halves; the lower entropy values are selected as shown in Fig. 4.3. Suppose the computed *threshold*=0.4, entropy values less than 0.4 are selected, these values are marked in bold in Fig. 4.3. k-means clustering is applied on all fuzzy entropy values of a review, it divides these values into two clusters. The cluster with values less than *threshold* is selected as depicted in Fig. 4.4.

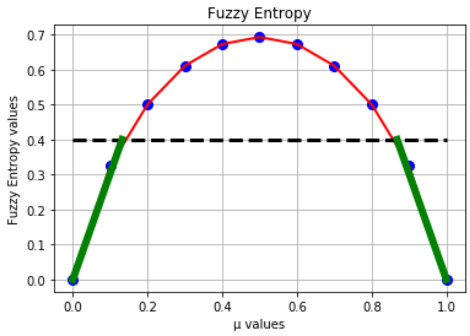


Fig. 4.3. Graph of fuzzy entropy with threshold=0.4.

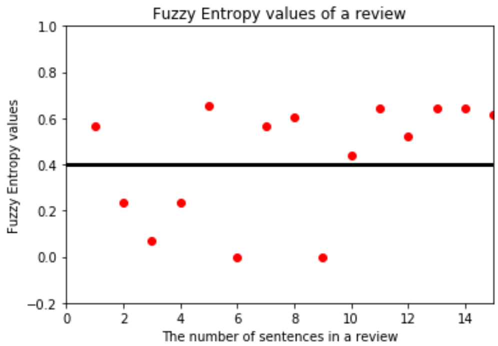


Fig. 4.4. k-means clustering applied on entropy values of a review.

4.1.6 Sentiment Evaluation

The final step of our methodology is to determine the sentiment of reviews using the extracted keyphrases, their scores and polarity, in an unsupervised manner. The sentiment of a review is calculated in three steps using (4.23-4.25). The i^{th} keyphrase has senti-score $\mu_{ph\_sentiscore}(i)$ and polarity *ph_polarity(i)*. The product of senti-score and polarity, *fuzzy_product* is computed in eq. (23), where *j* denotes the number of keyphrases shortlisted by fuzzy entropy filter in a review. An overall summation of these fuzzy products yields *fuzzy_sum* in eq. (24) that indicates the sentiment of a review as shown in (4.25); for positive sentiment the value is 1 and for negative sentiment it is 0.

$$fuzzy\_product(j) = \prod_{i=1}^{j} \mu_{ph\_sentiscore}(i) * ph\_polarity(i) \tag{4.23}$$

$$fuzzy\_sum = \sum_{j} fuzzy\_product(j) \tag{4.24}$$

$$sentiment = \begin{Bmatrix} 1, \; fuzzy\_sum \geq 0 \ldots Positive \\ 0, \; fuzzy\_sum < 0 \ldots Negative \end{Bmatrix} \tag{4.25}$$

The overall process flow for our proposed approach is shown in Fig. 4.5. Algorithm1 summarizes the steps to extract and highlight the keyphrases from a text document for the unsupervised Sentiment Analysis proposed in this chapter.

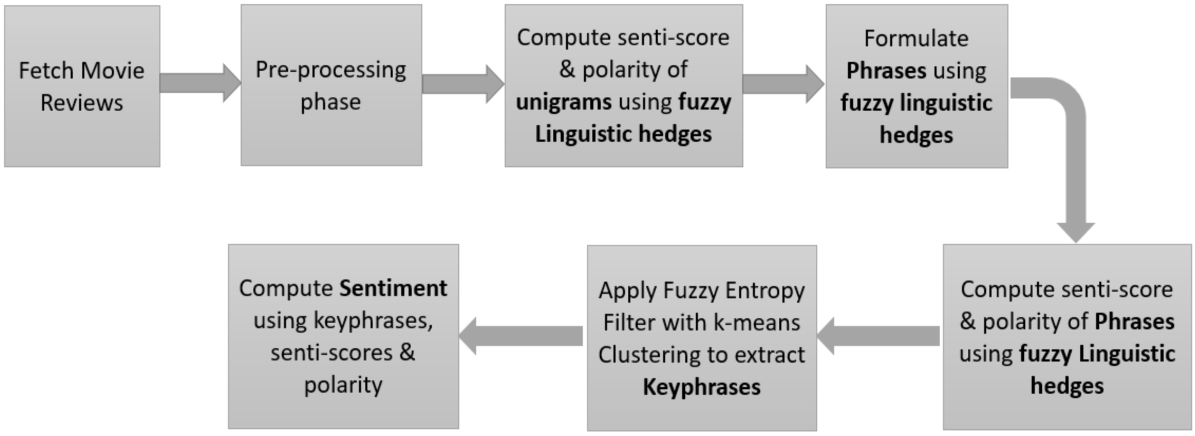


Fig. 4.5. Overall process flow for our unsupervised Sentiment Analysis system.

4.2 EXPERIMENTAL SETUP

In the experimental phase, the proposed system was executed in python with Intel Core i5 processor, 64-bit operating system and 8GB RAM. We have used two movie review datasets: polarity dataset v2.0 by Pang-Lee and the IMDB dataset. The first dataset includes 1000 positive and 1000 negative processed movie reviews. The IMDB dataset has a total of 50,000 reviews, with a training set of 25,000 labelled instances and a testing set of 25,000 labelled instances; we have merged both the sets since our approach is an unsupervised one. Our data now contains 25,000 positive and 25,000 negative labels. The dataset distribution of different datasets is shown in Table 4.2. We have utilized SentiWordNet lexicon (Baccianella *et al*., 2014) for computing the senti-scores of keyphrases for unsupervised Sentiment Analysis.

Table 4.2. Dataset Distribution of different datasets.

	Pang-Lee	IMDB
Positive	1000	25000
Negative	1000	25000
Total	2000	50000

4.3 RESULTS AND DISCUSSION

In this chapter we have proposed an unsupervised system for phrase-level sentiment analysis of online movie reviews. The senti-score of keyphrases is calculated based on the appearance of concentrators, dilators or negators with sentiment unigrams. Few step wise examples of such score calculation is shown in Table 4.3. For bigrams, all possible cases- positive and negative unigrams with concentrators, dilators or negators are explained and for trigrams different cases are explained in Table 4.3.

Table 4.3. Examples of phrases with score computation.

n-gram	Phrases	Steps for score Calculations	Senti- score & polarity computed from fuzzy linguistic hedges		Reference Equation
			Senti-score	Polarity	
Bigram Concentrator	Very cool	1. Senti-score of unigram- good	0.325	+1	eq. (1-9)
		2. score<1, Increase the score by 0.5	$(0.325)^{0.5} = 0.5701$	+1	eq. (10)
Bigram Concentrator	Extremely bad	1. Senti-score of unigram- bad	0.6923	-1	eq. (1-9)
		2. score<1, Increase the score by 0.5	$(0.6923)^{0.5} = 0.8320$	-1	eq. (10)
Bigram Dilator	Occasional good	1. Senti-score of unigram- bad	0.619	+1	eq. (1-9)
		2. score<1, Decrease the score by 2	$(0.619)^{2} = 0.3832$	+1	eq. (11)
Bigram	Almost	1. Senti-score of unigram- unlimited	0.5	-1	eq. (1-9)

Dilator	Unlimited	2. score<1, Decrease the score by 2	$(0.5)^2 = 0.25$	-1	eq. (11)
Bigram Negator	Nothing spectacular	1. Senti-score of unigram-spectacular	0.1667	+1	eq. (11)
		2. score<1, Modify the score by 1.5	$(0.1667)^{1.5} = 0.0681$	-1	eq. (12)
Bigram Negator	Not Horrible	1. Senti-score of unigram-horrible	0.625	-1	eq. (11)
		2. score<1, Modify the score by 1.5	$(0.625)^{1.5} = 0.4941$	+1	eq. (12)
Trigram Concentrator Concentrator	Very very sad	1. Senti-score of unigram- sad	0.667	-1	eq. (1-9)
		2. score<1, Increase the score by 0.5	$(0.667)^{0.5} = 0.8165$	-1	eq. (10)
		3. score<1, Increase the score by 0.5	$(0.8165)^{0.5} = 0.9036$	-1	eq. (13)
Trigram Dilator Concentrator	Somewhat very close	1. Senti-score of unigram- close	0.1875	+1	eq. (1-9)
		2. score<1, Increase the score by 0.5	$(0.1875)^{0.5} = 0.4330$	+1	eq. (10)
		3. score<1, Decrease the score by 2	$(0.4330)^2 = 1875$	+1	eq. (14)
Trigram Negation Concentrator	Not extremely surprised	1. Senti-score of unigram-	0.25	+1	eq. (1-9)
		2. score<1, Increase the score by 0.5	$(0.25)^{0.5} = 0.5$	+1	eq. (10)
		3. score<1, Modify the score by 1.5	$(0.5)^{1.5} = 0.3536$	-1	eq. (15)
Trigram Negation Negation	Although not wonderfully	1. Senti-score of unigram: wonderfully	0.5	+1	eq. (1-9)
		2. score<1, Modify the score by 1.5	$(0.5)^{1.5} = 0.3536$	-1	eq. (12)
		3. score<1, Modify the score by 1.5	$(0.3536)^{1.5} = 0.2102$	-1	eq. (19)

The Fuzzy Keyphrase Sentiment model proposed in this chapter has extracted many keyphrases, a few of them are mentioned here: "*Very cool*", "*Extremely bad*", "*Occasional good*", "*Almost Unlimited*", "*Nothing spectacular*", "*Not Horrible*", "*Very very sad*", "*Somewhat very close*", "*Not extremely surprised*", "*Although not wonderfully*".

There are different versions possible of our proposed system depending on the n-gram combinations and normalization scaling (explained in section 3.4). Scaling can be performed in two ways- same scale for all phrases or different scale for each n-gram i.e. unigram, bigram and trigram. There are three n-gram combinations: unigram and bigram; bigram and trigram; and unigram, bigram and trigram. Thereby a total of six versions exists for our proposed system that are investigated and compared in the subsequent sections.

4.3.1 EXPERIMENTS ON PANG-LEE DATASET

The accuracy and f-score for different versions of our system on the Pang-Lee dataset are depicted in Table 4.4. The values in bold are the highest accuracy and f-score values among all versions. The all n-gram combination with same scale has achieved the highest accuracy of 70% and highest f-score of 0.701. The comparison of our best version system and other phrase-level SA based models for the Pang-Lee movie review dataset are shown in Table 4.5. Srivastava and Bhatia (2013) selected only adjective words as unigrams; fuzzy linguistic hedges are applied on these unigrams to generate bigrams and trigrams. They have created a linguistic resource: Opinion Words Lexicon (OWL) and utilized it to compute scores of n-grams. This model achieved lowest f-score and lowest accuracy of 49 % and 49.6% for unigram-bigram and unigram-bigram-trigram combinations respectively. Polanyi and Zaenen (2006) used valence points adjustment approach- a simple hedge adjustment method with General Inquirer (GI) lexicon. All positive sentiment terms from GI are given an initial value of 2 and negative terms are given an initial value of -2. If this term is preceded by a concentrator (intensifier) in the same phrase its value is increased by 1, while if it is preceded by a dilator (diminisher) its value is decreased by 1. Positive or negative valanced words with a negation, flips the valence to a Negative or Positive valence. This approach achieves an accuracy and f-score of 60.3% and 0.4123 respectively if the count of positive and negative terms is compared, and 61.25 % and 0.4154 respectively if sum of scores is compared. In the next method, Turney's (Turney, 2002) phrases are extracted and their scores are computed using SentiWordNet lexicon (Baccianella *et al*., 2014). The scores of all phrases are summed up and compared; if the sum is greater than 0 it's a positive sentiment, otherwise it's a negative sentiment. This method produces an accuracy of 65.25% and f-score of 0.6502. Dalal and Zaveri proposed phrase level SA that uses fuzzy linguistic hedges to compute the scores of phrases (Dalal and Zaveri, 2014). The phrases are constructed using concentrators that increase and dilators that decrease the scores. The inverter hedges don't modify the score but only reverse the sentiment polarity. Scores of all phrases and feature nouns are used to compute the sentiment. This approach yields an accuracy of 65% and f-score value is 0.6501. Our proposed system with all n-gram combination and same normalization scaling achieves the highest accuracy of 70% and highest f-score of 0.701 among all phrase level SA methods using Pang-Lee dataset.

Table 4.4. Different versions of our system on Pang-Lee dataset.

N-gram	Different Scale		Same Scale	
	Accuracy (%)	**f-score**	**Accuracy (%)**	**f-score**
Uni + Bigram	62.0	0.619	68.95	0. 689
Bi + Trigram	54.15	0.541	54.7	0.546
Uni + Bi + Trigram	62.25	0.622	**70**	**0.701**

Bold values are the highest accuracy and f-score values among all versions.

Table 4.5. Comparison of state-of-the-art vs our system on Pang-Lee dataset.

Method	Accuracy (%)	f-score macro
(Srivastava and Bhatia, 2013) Uni+ Bi	49	0.3044
(Srivastava and Bhatia, 2013) Uni+Bi+Tri	49.6	0.3057
(Polanyi and Zaenen, 2006) count	60.3	0.4123
(Polanyi and Zaenen, 2006) sum	61.25	0.4154
(Turney, 2002)	65.25	0.6502
(Dalal and Zaveri, 2014)	65	0.6501
Fuzzy Keyphrase Sentiment Model	**70**	**0.701**

Bold values are the highest accuracy and f-score values among all methods.

4.3.2 EXPERIMENTS ON IMDB DATASET

IMDB dataset contains 50,000 reviews. Results of the experiments conducted on this dataset for different versions of our system is shown in Table 4.6. Bold values are the highest accuracy and f-score values among all versions. The unigram-bigram-trigram combination with same scale has achieved the highest accuracy of 69.3% and highest f-score of 0.691. The comparison of our best version with other sentiment analysis techniques is depicted in Table 4.7. Here also, Srivastava and Bhatia (2013) attained the lowest accuracy of 48% and 48.75% for unigram-bigram and unigram-bigram-trigram approaches respectively. Polanyi and Zaenen (2006) achieves 62.45% and 63.36 % accuracy for both cases, and f-score is 0.4421 and 0.4451 respectively. Turney's (Turney, 2002) extracted patterns with phrase level SA technique gained an accuracy of 67.42% and f-score value is 0.4488. The last method (Dalal and Zaveri, 2014) produced 64.73% accuracy and 0.6452 f-score. Hence, our system outperforms all other phrase level SA techniques on the IMDB dataset with 69.3 % accuracy and 0.691 f-score.

Comparison of our proposed unsupervised Sentiment Analysis system that extracts and highlights the keyphrases from movie reviews with other phrase level SA state-of-the displays good results. Our system yields better results by using senti-scores of phrases, formulated from sentiment lexicon, fuzzy linguistic hedges and fuzzy entropy, as features for sentiment classification.

Table 4.6. Different versions of our system on IMDB dataset.

N-gram	Different Scale		Same Scale	
	Accuracy (%)	**f-score**	**Accuracy (%)**	**f-score**
Uni + Bigram	66.72	0.663	68.69	0.686
Bi + Trigram	66.72	0.663	55.81	0.536
Uni + Bi + Trigram	66.98	0.665	**69.3**	**0.691**

Bold values are the highest accuracy and f-score among all versions.

Table 4.7. Comparison of state-of-the-art vs our system on IMDB dataset.

Method	Accuracy (%)	f-score macro
(Srivastava and Bhatia, 2013) Uni+ Bi	48	0.3011
(Srivastava and Bhatia, 2013) Uni+Bi+Tri	48.75	0.3129
(Polanyi and Zaenen, 2006) count	62.45	0.4421
(Polanyi and Zaenen, 2006) sum	63.36	0.4451
(Turney, 2002)	67.42	0.4488
(Dalal and Zaveri, 2014)	64.73	0.6452
Fuzzy Keyphrase Sentiment Model	**69.3**	**0.691**

Bold values are the highest accuracy and f-score values among all methods.

4.4 CONCLUSION

Sentiment Analysis is the evaluation and study of public's opinions, attitudes, and emotions toward an entity. The entity can represent individuals, events or topics. These topics are most likely to be covered by reviews. Public sentiment regarding any social issue can be analysed easily using SA. The proposed unsupervised sentiment classification system comprehensively formulates phrases, compute their senti-scores (sentiment scores) and polarity using fuzzy linguistic hedges. Further, it extracts the keyphrases, significant for SA, using fuzzy an entropy filter with k-means clustering. The key contributions are i) An unsupervised phrase-level SA approach has been proposed to perform sentiment analysis on online reviews using n-gram techniques, specifically a combination of unigram, bigram and trigram, ii) Phrases are constructed comprehensively using part-of-speech (POS) Tagger, list of concentrators, dilators and negators. Their senti-scores and polarity are computed using SentiWordNet lexicon and fuzzy linguistic hedges, iii) Document-level SA on online reviews is executed by extracting high sentiment bearing keyphrases filtered out by fuzzy entropy and k-means clustering, and finally computing the sentiment of the review and iv) The performance of our fuzzy technique is evaluated using the parameters of accuracy and f-score.

The results indicate higher classification scores as compared to the state-of-the-art. According to our insights, till now, keyphrase extraction in Sentiment Analysis and computation of sentiment scores of keyphrases using fuzzy measures have not been investigated by any researcher. The advantage of this unsupervised algorithm is that it easily handles the uncertainties, ambiguities, or vagueness that exist while interpreting the sentiment; and also mimics the logic of human thought. Moreover, it doesn't require any training thus it saves computational complexity and time.

In the future, we would employ a supervised sentiment classification system with keyphrases and their senti-scores as features. We would compute senti-scores for emoticons used in the reviews that were not considered for automatic score calculation. Some jargons, slang, misspellings and oddly spelled words such as "ROFL", "LOL", "verrrrryyyy", "funnyyyyy", "omggggg" that probably emphasizes certain feeling are also ignored in this work. We will tackle such acronyms in our future work.

CHAPTER 5

DESIGN AND IMPLEMENTATION OF NEURO-FUZZY NETWORKS TO DETECT SENTIMENT

In this chapter[1], we focus on segregating tweets that contain neutral sentiment. We classify tweets into two classes: neutral and not-neutral; the latter class includes both positive and negative polarity. This type of classification will be acknowledged for applications that aim to search for neutral content posted by the users. Nowadays, there is a new trend, tasks like personality profiling and behavior profiling are executed by organizations for visa applications, military services, and intelligence agencies for tracking anti-nationalists or terrorist activities. Such tasks require SA on an online social network that can classify neutral and non-neutral content about a person or group. In this study we have proposed MultiLexANFIS which is an Adaptive Neuro-Fuzzy Inference System (ANFIS) that incorporates inputs from multiple lexicons- VADER (Hutto and Gilbert, 2014), AFINN (Nielsen, 2011) and SentiWordNet (Baccianella *et al*., 2010), to perform sentiment analysis of social media posts. The key contributions of this research work are: 1) a novel neuro-fuzzy system: MultiLexANFIS that combines sentiment

---

:[1] The contents of this chapter are published in " Neuro-Fuzzy Network incorporating Multiple Lexicons for Social Sentiment Analysis" in *Soft Computing* (2021): 1-21.

scores from multiple lexicons to classify tweets, 2) a novel set of 64 rules for Sugeno-type fuzzy Inference System, 3) single lexicon-based ANFIS variants to classify tweets, and 4) comparison of fuzzy approaches with non-fuzzy approaches reveals the superiority of our proposed neuro-fuzzy system.

This chapter is organized as follows. Section 5.1 describes several recent works related to ANFIS and Sentiment Analysis. The proposed MultiLexANFIS system is explained in section 5.2. The experimental setup and implementation details are discussed in Section 5.3. The results are demonstrated in Section 5.4. Finally, Section 5.5 draws the overall conclusion.

5.1 PROPOSED ANFIS ARCHITECTURE FOR SOCIAL SENTIMENT ANALYSIS

In this chapter a novel ANFIS based system: MultiLexANFIS that integrates inputs from multiple sentiment lexicons to compute the sentiment of social media posts has been proposed. We have also proposed single lexicon based ANFIS variants that require input from one sentiment lexicon at a time. Our dataset comprises of tweets, we have applied different ANFIS variants on tweets to categorize them into two classes: neutral and not-neutral. We have merged the positive and negative polarity into one class: not-neutral.

5.1.1 MOTOVATION

In the last decade, fuzzy systems and their hybrid derivations have proved that they can replicate the typical human reasoning ability in a computationally productive way. Fuzzy models have been used to construct cognitive models based on rules, expert systems, classifiers, and universal approximators. Adaptive Neuro-Fuzzy Inference System (ANFIS) is a neuro-fuzzy system that encompasses the power of the artificial neural networks and fuzzy logic. Our ANFIS model combines natural language processing (NLP) with fuzzy logic, which boosts the performance of the SA model. The importance of fuzziness comes into play while dealing with natural language due to the presence of ambiguity in language. The amalgamation of fuzzy logic concepts into SA helps in dealing with the fuzziness of natural language in a very efficient and automatic manner. There are neuro-fuzzy systems in the field of SA or opinion mining, but these systems are unable to deal with the uncertainty and vagueness present in the natural language.

Moreover, these systems usually categorize the data into positive or negative classes; our aim is to categorize data into neutral on a non-neutral class that would aid organizations that are hunting for neutral content about persons, events, services, etc. To the best of our knowledge, this novel sentiment classification problem- segregation of neutral and not-neutral classes, has not been handled by any research work. We have designed two types of ANFIS variants- single lexicon and multiple lexicons. The single variant is based on the type of sentiment lexicon chosen. This variant has been designed to figure out which lexicon is best suited for ANFIS based sentiment classification in the field of social SA. Since a single lexicon variant is not able to conclude which fuzzy membership is best suited for ANFIS based SA, we designed a multiple lexicon ANFIS variant- MultiLexANFIS. To solve the above confusion of choosing the best membership function (MF) for a given lexicon and dataset combination, MultiLexANFIS architecture has been proposed. This system integrates all the lexicons into a single unit and produces lower errors compared to other models and it takes less execution time in contrast to single lexicon based ANFIS. In our work, the proposed novel rules are domain-independent. We can apply these rules to any textual data that employs lexicons to extract negative and negative scores for each textual document or sentence or review.

5.1.1.1 Motivation behind segregating neutral and not-neutral classes:

The motivation behind segregating neutral and not-neutral classes is that in some applications the neutral content carries more importance than others. It has been demonstrated that a mix of training samples containing neutral samples is better compared to a training set of the same size that includes only positive and negatives samples (Koppel and Schler, 2005; Koppel and Schler, 2006). The use of neutral training samples in learning not only promotes better discrimination between positive and negative samples but also increases the classification accuracy (Koppel and Schler, 2006). There exists a bias in social media content posted by users. Social theories claim that bias is a deep-seated feature of human behavior, which is described by a lack of proper balance and neutrality; one of the instances of bias is when someone supports one side too strongly or too often (Guerra *et al.*, 2011). We have designed an ANFIS based architecture that can be used by organizations which hunt for neutral and unbiased content. Currently, there is a new trend, tasks like personality profiling and behavior profiling are deployed by organizations based on the neutrality of content. Organizations that deal with visa applications require the candidates to have neutral behavior for their country, those candidates who are too biased can pose a threat for that country. Such system can also be used by military services and intelligence agencies for tracking anti-nationalists or terrorist activities.

5.1.1.2 Comparison of our proposed ANFIS model with existing ANFIS works in literature

Katta and Hedge (2019) proposed a hybrid ANFIS that integrates Non-linear SVM for SA of political Twitter. In the feature extraction phase, unigrams and bigrams models are used. Only the words or pair of words are fed as input to ANFIS; whereas in our proposed model we evaluate the positive and negative sentiment scores of each tweet and send them as input to ANFIS. This will give better results because sentiment scores extracted from lexicons would provide better information for sentiment classification. In (Katta and Hedge, 2019), the authors used only one fuzzy MF: generalized bell; while in our proposed model we have explored three fuzzy MF: triangular (trimf), generalized bell (gbellmf), and gaussian (gaussmf), in order to select the best performing MF. The authors (Katta and Hedge, 2019) have conducted experiments and compared their model's different hybrid ANFIS-SVM variants but not with other state-of-the-art approaches, whereas in this work we have conducted experiments and compared them with one fuzzy-rule based classifier and four non-fuzzy state-of-the-art methods.

Kamil *et al*., (2018) applied ANFIS for executing three different classification problems. In the pre-processing phase, the text is converted into an array of words which is sorted in increasing order of alphabets. Each word is assigned a unique code. The words are not lemmatized, hence neither verbs in different tenses nor nouns as singular or plural are combined. In our approach, the words are lemmatized. The advantage of lemmatization is that it checks the context of the word to ascertain which is the correct meaning of that word. The authors in (Kamil *et al*., 2018) do not use lexical knowledge, whereas we use lexicon to extract the positive and negative polarity scores that contribute to SA. Sentiment lexicons permit the creation of important and high-quality features for supervised SA. Moreover, they can be applied to diverse domains. The incorporation of sentiment lexicon makes our model domain independent. In this work, the membership degree of each term is computed by Pruned ICF (Inverse-Class Frequency), whereas in our model, three popular fuzzy MF is applied for fuzzification of sentiment scores.

Padmaja and Hedge, (2019) developed a hybrid classifier based on ANFIS and Genetic Algorithm (GA) to categorize the tweets into binary sentiment classes- positive or negative class, while our proposed model classifies tweets into neutral or not-neutral sentiment classes.

The ANFIS classifier integrates the advantages of both artificial neural networks and fuzzy logic controllers for generating the fuzzy rules and the GA optimizes the fuzzy principles in the ANFIS classifier. After pre-processing of raw tweets, features are extracted; here positive word count, negative word count, and tag count are considered as features. The positive and negative words are attained by considering only adjectives, whereas in our proposed model the positive and negative score of tweets is computed by considering adjectives, nouns, verbs, and adverbs. The inclusion of other parts of speech words enhances the sentiment evaluation process. Moreover, they have not mentioned how the words are classified as positive or negative. In our ANFIS model, we have deployed sentiment lexicons to evaluate the positive and negative scores of tweets. The authors (Padmaja and Hedge, 2019) have compared only one existing methodology Latent Dirichlet Allocation Possibilistic Fuzzy C-Means (LDA-PFCM), and conducted experiments on only one dataset of 479 tweets. Our proposed model has been compared with five existing SA approaches and the experiments are conducted on seven datasets.

5.1.2 PROCESSING OF TWEETS

In this section, we will describe the overall procedure of tweet processing. Initially, the Twitter data is collected; many publicly available datasets consist of positive, negative, and neutral tweets. We have used a total of seven publicly available datasets: The Apple Twitter Dataset, The Nuclear Twitter Dataset, The Sanders Twitter Dataset, SemEval 2017, SemEval 2016, SemEval 2015, and The Stanford Twitter Sentiment Test Set (STS-Test) (Go *et al.*, 2009). All the datasets contain three types of tweets: positive, negative, and neutral. We have combined the positive and negative tweets into one class: not-neutral; our aim is to categorize the tweets by the ANFIS system into two classes: neutral and not-neutral. We changed the labels of positive and negative tweets to not-neutral; thus, we have now two types of classes: neutral and not-neutral. These labels, 1: neutral and 0: not-neutral are the class labels in ANFIS. The various steps in tweet processing are outlined below.

5.1.2.1 Data pre-processing

The next step is data pre-processing, which is very essential because raw tweets are in unstructured form. In data pre-processing, we eliminate URLs, '@' symbol, and hashtag '#'

symbol. We have re-phrased commonly used phrases (like “can’t”) with their grammatical form (“can not”). This has been re-phrased because words like ‘not’ contain polarity scores, if these phrases are not replaced these words would be ignored, and thus it can lead to incorrect results.

5.1.2.2 Feature extraction using Sentiment Lexicon

We have to compute the sentiment of this pre-processed tweet, which requires numerical data as a feature of the tweet. So that we can use this feature as input in ANFIS. Sentiment lexicons are employed for this purpose, these lexicons assign scores to words or to the whole tweet. The three sentiment lexicons are VADER, AFINN, and SentiWordNet. These sentiment lexicons are popularly applied in SA applications.

VADER (Valence Aware Dictionary for sEntiment Reasoning) is a sentiment lexicon that is especially attuned to sentiment in microblog-like contexts (Hutto and Gilbert, 2014). It works fast and yields high accuracy. It computes the scores for the whole tweet, *s*, using VADER lexicon’s *polarity_scores(s)* method and gives positive (*TweetPos*) and negative (*TweetNeg*) scores of a tweet as output. AFINN lexicon consists a list of English terms manually rated for valence with an integer between -5 (negative) and +5 (positive) (Nielsen, 2011). This list also contains slang and obscene words; hence it can deal with modern-day tweets very well. AFINN lexicon has the AFINN method which fetches the score of each word, *a*, (5.1), if it is greater than 0 it is a positive word (5.2), and *if* less than 0 it is a negative word (5.3). The positive score of the tweet (*TweetPos*) is calculated by adding up all positive words (5.4); similarly, the negative score (*TweetNeg*) is calculated for each tweet (5.5).

$$afn = afinn.score(a) \tag{5.1}$$

$$if\,(afn > 0)then(af_{Pos} = afn)) \tag{5.2}$$

$$if\,(afn < 0)then(af_{Neg}(a) = -afn) \tag{5.3}$$

$$TweetPos = \sum_{a=1}^{m} af_{Pos}(a) \tag{5.4}$$

$$TweetNeg = \sum_{a=1}^{m} af_{Neg}(a) \tag{5.5}$$

where *a* is a word in a tweet, *m* is the number of selected words and X_i is the set of total words.

One of the most commonly used lexicons by researchers for SA is SentiWordNet lexicon. It allocates three numerical scores to each word: positive, negative and objective; these scores are in the range of 0.0 to 1.0 (Nielsen, 2011). The tweets are processed further for this lexicon: we remove the stopwords and punctuations, perform lemmatization, each word is assigned Part of Speech (POS) tagger and application of Word Sense Disambiguation (WSD) by Lesk (Banerjee and Pedersen, 2002). WSD is the method of ascertaining which meaning of a word is stimulated by the use of the word in a particular context. The SentiWordNet method acquires the scores of each word, *a*, from this lexicon using *syn.pos_score* () and *syn.neg_score* () Each word has a positive (5.8) and negative score (5.9) computed using WSD, that can be interpreted as a fuzzy membership belonging to the fuzzy sets *Pos* (5.6) and *Neg* (5.7). The words, which are having higher positive scores than negative scores in a tweet, are summed up to evaluate the positive score (*TweetPos*) of the tweet (5.10). Similarly, words which are having higher negative scores than positive scores in a tweet are summed up to evaluate the negative score (*TweetNeg*) of the tweet (5.11). These scores are computed for all tweets.

$$Pos = \{(a, \mu_{Pos}(a)\}, a \in X_i \tag{5.6}$$

$$Neg = \{(a, \mu_{Neg}(a)\}, a \in X_i \tag{5.7}$$

$$\mu_{Pos}(a) = syn.pos\_score() \tag{5.8}$$

$$\mu_{Neg}(a) = syn.neg\_score() \tag{5.9}$$

$$\begin{aligned} &if\,(\mu_{Pos}(a) > 0\ \&\ \mu_{Pos}(a) > \mu_{Neg}(a)) \\ &then(TweetPos = \sum_{a=1}^{m} \mu_{Pos}(a)) \end{aligned} \tag{5.10}$$

$$\begin{aligned} &if\,(\mu_{Neg}(a) > 0\ \&\ \mu_{Neg}(a) > \mu_{Pos}(a)) \\ &then(TweetNeg = \sum_{a=1}^{m} \mu_{Neg}(a)) \end{aligned} \tag{5.11}$$

where *a* is a word in a tweet, *m* is the number of selected words and X_i is the set of total words. The computed *TweetPos* and *TweetNeg* of each tweet represents the numerical score of that tweet; these scores are sent as input in ANFIS.

5.1.3 ANFIS PROCESSING OF TWEETS

ANFIS was introduced by Jang (1993), which takes crisp inputs and fuzzifies these inputs using membership functions. ANFIS is a technique in the field of artificial intelligence, which incorporates fusion learning that associates fuzzy logic and neural networks into one unit. It is based on specific "human features", because fuzzy logic has a human subjectivity feature, while the neural network has a feature of the human way of thinking based on some experience. ANFIS networks are appropriate for solving function problems in various fields. Generally, the problem, for which we are finding the solution, is the approximation of an unknown function. We next explain the working of a general ANFIS network constituted by M rules of Sugeno first-order type (Sugeno and Kang, 1988). The kth rule, k =1…M, has the following form:

if x_1 is $B_1^{(k)}$, and x_2 is $B_2^{(k)}$… x_n is $B_n^{(k)}$, then $y^{(k)} = \sum_{j=1}^{n} a_j^{(k)} x_j + a_0^{(k)}$ (5.12)

where $x = [x_1 x_2 ... x_n]$ is the input pattern and $y^{(k)}$ is the output associated with the k^{th} rule. Each rule is characterized by the membership functions (MFs) $\mu_{B_j^{(k)}}(x_j)$ of the fuzzy input variables $B_j^{(k)}$, j = 1…n, and by the coefficients $a_j^{(k)}$, j = 1…n, of the crisp output. There are numerous options available for the fuzzifying crisp inputs, the composition of input MFs, and in which manner the rule outputs are integrated. The next step is to find the AND product of the fuzzy memberships of all input variables in a rule. This product is a MF $\mu_{B^{(k)}}(x)$, also called the firing strength (or weight) of the rule. The M-rule fuzzy inference system for ANFIS networks can be mathematically represented as:

$$\tilde{y} = \frac{\sum_{k=1}^{M} \mu_{B^{(k)}}(x).y^{(k)}}{\sum_{k=1}^{M} \mu_{B^{(k)}}(x)} \quad (5.13)$$

where $\tilde{y}$ is the estimate of the actual value $y = f(x)$, for a given input x.

We have proposed two ANFIS based architectures: one is the single lexicon based ANFIS and the other, MultiLexANFIS, combines all the lexicons into a single integrated unit. Single lexicon based ANFIS takes two inputs, *TweetPos* and *TweetNeg*; and one output, sentiment label (1: neutral or 0: not-neutral), for one sentiment lexicon-dataset combination at a time. MultiLexANFIS system has six inputs, it includes the positive and negative score of a tweet from all the three lexicons; and one sentiment label as output for each dataset.

5.1.3.1 Single Lexicon based ANFIS

The fuzzy inference system (FIS) is of Sugeno first-order type with two inputs *TweetPos* and *TweetNeg* and one output *sentiment*. A common rule set with two inputs and three Membership Functions (MFs) for each input would yield nine rules ($3^2=9$). The MFs for the input have linguistic labels: *Low*, *Medium* and *High*. The labels *Low*, *Medium* and *High* are two fuzzy sets created using different fuzzy membership for input variables: positive score (*TweetPos)* and negative score (*TweetNeg)*. Mathematically, we can denote the inputs as x and y, MFs for input x as A_i, MFs for input y as B_i. The fuzzy if-then rules are as follows:

Rule1: If x is A_1 and y is B_1 then $f_1 = p_1 x+q_1 y+r_1$ (5.14)

Rule2: If x is A_2 and y is B_1 then $f_2 = p_2 x+q_2 y+r_2$ (5.15)

Rule3: If x is A_3 and y is B_1 then $f_3 = p_3 x+q_3 y+r_3$ (5.16)

Rule4: If x is A_1 and y is B_2 then $f_4 = p_4 x+q_4 y+r_4$ (5.17)

Rule5: If x is A_2 and y is B_2 then $f_5 = p_5 x+q_5 y+r_5$ (5.18)

Rule6: If x is A_3 and y is B_2 then $f_6 = p_6 x+q_6 y+r_6$ (5.19)

Rule7: If x is A_1 and y is B_3 then $f_7 = p_7 x+q_7 y+r_7$ (5.20)

Rule8: If x is A_2 and y is B_3 then $f_8 = p_8 x+q_8 y+r_8$ (5.21)

Rule9: If x is A_3 and y is B_3 then $f_9 = p_9 x+q_9 y+r_9$ (5.22)

The antecedent part of these rules is the same as in the recently published previous work of authors on unsupervised fuzzy rule-based sentiment analysis system (Vashishtha and Susan, 2019 a). In Mamdani FIS, a fuzzy system with two non-interactive inputs A and B (antecedents) and a single output

C (consequent) is expressed by a number of r linguistic IF–THEN propositions in the Mamdani form:

Rule R_j: IF A is A_1^j and B is B_1^j THEN C is C_1^j, j =1,2,..,r (5.23)

where A_{1j} and B_{1j} are the fuzzy sets representing the jth antecedent or premise pairs and C_{1j} is the fuzzy set representing the jth consequent.

In this work for single lexicon ANFIS, we transform the set of nine Mamdani fuzzy rules in (Vashishtha and Susan, 2019 a) to parametrized Sugeno form where the parameters are learnt by ANFIS. Mamdani FIS entails a substantial computational burden and is appropriate to human input, while Sugeno FIS is computationally efficient, has flexibility in system design and its well appropriate to mathematically analysis. Hence our proposed system in this work applies Sugeno FIS for better classification results.

The ANFIS architecture comprises of five layers (Jang, Sun and Mizutani, 1997). Our single lexicon ANFIS architecture is shown in Fig. 5.1. We represent the output of the ith node in layer l as O_{li}. Each layer is described next.

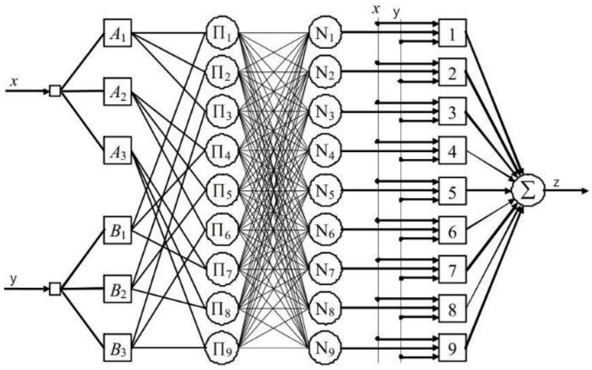


Fig. 5.1. Single Lexicon based ANFIS architecture with 9 rules

Layer 1: Every node i in this layer is an adaptive node with a node function

$$O_{l,i} = \mu_{A_i}(x), \text{ for } i = 1,2, \text{ or} \qquad (5.24)$$

$$O_{l,i} = \mu_{B_{i-2}}(y), \text{ for } i = 3,4, \qquad (5.25)$$

here O_{li} is the membership grade of a fuzzy set A (=A_1, A_2 or A_3) or B (=B_1, B_2 or B_3) and it specifies the degree to which the given input (x or y) satisfies the quantifier A. Here the MF for A can be any parametrized MF: triangular (trimf), generalized bell (gbellmf) and gaussian (gaussmf) described below.

$$\mu_t(x)=\begin{cases}0, x\le d\\ (x-d)/(e-d), d<x\le e\\ (f-x)/(f-e), e<x\le f\\ 0, x\ge f\end{cases} \tag{5.26}$$

$$\mu_{gb}(x)=\frac{1}{1+\left|\frac{x-w}{u}\right|^{2v}} \tag{5.27}$$

$$\mu_{ga}(x)=e^{-\frac{1}{2}\left(\frac{x-c}{\sigma}\right)^2} \tag{5.28}$$

A brief explanation of these MFs ensues. The triangular membership function has the simplest shape among others. It is defined by three parameters: d and f for feet, and e for the tip of the curve. The mathematical formula of the function is given in (5.26). Gbellmf is a symmetrical shape similar to a bell. As expressed by (5.27), this function employs three parameters: u determines the width of the bell like curve, v is a positive integer, while w sets the center of the curve in the universe of discourse. Gaussmf also has a smooth curve. However, as compared to all MFs mentioned above, it utilizes only two parameters: c for locating center and σ for determining the width of the curve as expressed mathematically by (5.28). The values of these parameters are attuned in the learning phase of ANFIS. All the parameters defined in this layer are called as premise parameters.

Layer 2: Every node in this layer is a fixed node labelled as Π, whose output is the product of all the incoming signals in (5.29). Each node output represents the firing strength of a rule.

$$O_{2,i}=w_i=\mu_{A_i}(x)\mu_{B_i}(y),\ i=1,2. \tag{5.29}$$

Layer 3: Every node in this layer is a fixed node labeled *N*. The i^{th} node calculates the ratio of the i^{th} rule's firing strength to the sum of all rules' firing strengths in (30). Outputs of this layer are called the normalized firing strengths.

$$O_{3,i}=\overline{w}=\frac{w_i}{w_1+w_2},\ i=1,2 \tag{5.30}$$

Layer 4: Every node *i* in this layer is an adaptive node with a node function:

$$O_{4,i}=\overline{w_i}f_i=\overline{w}_i(p_ix+q_iy+r_i) \tag{5.31}$$

where w_i is the normalized firing strength from layer 3 and $\{p_i, q_i, r_i\}$ is the parameter set of this node. Parameters in this layer are referred to as consequent parameters.

Layer 5: In this layer only one fixed node labeled ∑ is there, which calculates the overall output by summing up all incoming signals:

$$\textit{overall output} = O_{5,i} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{5.32}$$

5.1.3.2 Multiple Lexicon based ANFIS: MultiLexANFIS

MultiLexANFIS, combines all the three lexicons into a single integrated unit. MultiLexANFIS system has six inputs, it includes positive and negative score of a tweet from all the three lexicons; and one sentiment label as output for each dataset. The six inputs are: *vpos*, *vneg*, *apos*, *aneg*, *spos* and *sneg*. The *vpos* and *vneg* are positive and negative score of a tweet computed using VADER lexicon (Hutto and Gilbert, 2014); similarly, apos and aneg are evaluated using AFINN lexicon (Nielsen, 2011) and *spos* and *sneg* are calculated using SentiWordNet lexicon (Baccianella *et al*., 2010) for the same tweet. The output is sentiment label, which has value 1 for neutral tweet and 0 for not-neutral tweet. For ease of computations and for reducing the number of fuzzy rules, each input has been assigned two membership functions (MFs): *Low* and *High*. The labels *Low* and *High* are two fuzzy sets created using different fuzzy membership for input variables: *vpos*, *vneg*, *apos*, *aneg*, *spos* and *sneg*. Hence, there is a total of 64 rules (2^6=64). These rules are depicted in Table 5.1 and Table 5.2. These proposed novel rules are domain independent. We can apply these rules for any textual data that employs lexicons to extract positive and negative score for each textual document or sentence or review. MultiLexANFIS architecture has five layers similar to the single lexicon ANFIS discussed in last section, except for the fact that the number of inputs has increased from two to six and number of MFs for each input is two. MultiLexANFIS architecture is depicted in Fig. 5.2

Table 5.1. MultiLexANFIS Rules (1-32).

RULES	INPUT1	INPUT2	INPUT3	INPUT4	INPUT5	INPUT6	OUTPUT
1	Low	Low	Low	Low	Low	Low	O1
2	Low	Low	Low	Low	Low	High	O2
3	Low	Low	Low	Low	High	Low	O3
4	Low	Low	Low	Low	High	High	O4
5	Low	Low	Low	High	Low	Low	O5
6	Low	Low	Low	High	Low	High	O6
7	Low	Low	Low	High	High	Low	O7
8	Low	Low	Low	High	High	High	O8
9	Low	Low	High	Low	Low	Low	O9
10	Low	Low	High	Low	Low	High	O10
11	Low	Low	High	Low	High	Low	O11
12	Low	Low	High	Low	High	High	O12
13	Low	Low	High	High	Low	Low	O13
14	Low	Low	High	High	Low	High	O14
15	Low	Low	High	High	High	Low	O15
16	Low	Low	High	High	High	High	O16
17	Low	High	Low	Low	Low	Low	O17
18	Low	High	Low	Low	Low	High	O18
19	Low	High	Low	Low	High	Low	O19
20	Low	High	Low	Low	High	High	O20
21	Low	High	Low	High	Low	Low	O21
22	Low	High	Low	High	Low	High	O22
23	Low	High	Low	High	High	Low	O23
24	Low	High	Low	High	High	High	O24
25	Low	High	High	Low	Low	Low	O25
26	Low	High	High	Low	Low	High	O26
27	Low	High	High	Low	High	Low	O27
28	Low	High	High	Low	High	High	O28
29	Low	High	High	High	Low	Low	O29
30	Low	High	High	High	Low	High	O30
31	Low	High	High	High	High	Low	O31
32	Low	High	High	High	High	High	O32

Table 5.2. MultiLexANFIS Rules (33-64) Contd.

RULES	INPUT1	INPUT2	INPUT3	INPUT4	INPUT5	INPUT6	OUTPUT
33	High	Low	Low	Low	Low	Low	O33
34	High	Low	Low	Low	Low	High	O34
35	High	Low	Low	Low	High	Low	O35
36	High	Low	Low	Low	High	High	O36
37	High	Low	Low	High	Low	Low	O37
38	High	Low	Low	High	Low	High	O38
39	High	Low	Low	High	High	Low	O39
40	High	Low	Low	High	High	High	O40
41	High	Low	High	Low	Low	Low	O41
42	High	Low	High	Low	Low	High	O42
43	High	Low	High	Low	High	Low	O43
44	High	Low	High	Low	High	High	O44
45	High	Low	High	High	Low	Low	O45
46	High	Low	High	High	Low	High	O46
47	High	Low	High	High	High	Low	O47
48	High	Low	High	High	High	High	O48
49	High	High	Low	Low	Low	Low	O49
50	High	High	Low	Low	Low	High	O50
51	High	High	Low	Low	High	Low	O51
52	High	High	Low	Low	High	High	O52
53	High	High	Low	High	Low	Low	O53
54	High	High	Low	High	Low	High	O54
55	High	High	Low	High	High	Low	O55
56	High	High	Low	High	High	High	O56
57	High	High	High	Low	Low	Low	O57
58	High	High	High	Low	Low	High	O58
59	High	High	High	Low	High	Low	O59
60	High	High	High	Low	High	High	O60
61	High	High	High	High	Low	Low	O61
62	High	High	High	High	Low	High	O62
63	High	High	High	High	High	Low	O63
64	High	High	High	High	High	High	O64

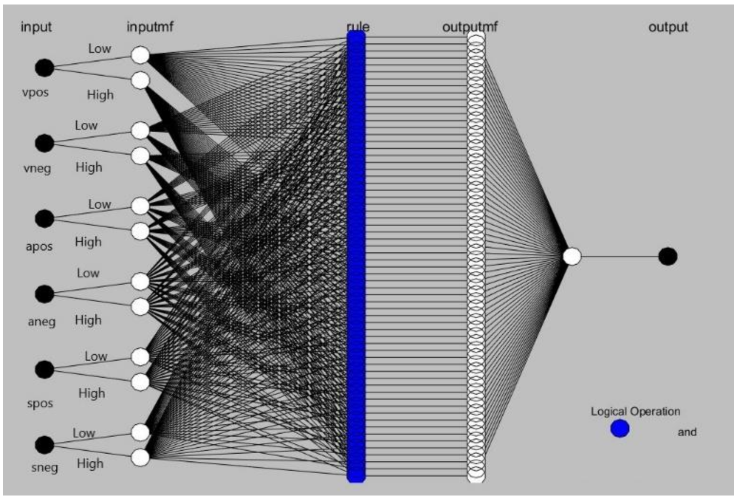


Fig. 5.2. MultiLexANFIS architecture with 64 rules

ANFIS learns by updating tunable parameters which are the premise parameters and the consequent parameters. The first and fourth layer have adaptive nodes, so they are trainable, whereas the nodes of rest of the layers are fixed. The optimization method adopted by ANFIS is, hybrid learning algorithm, composed of two passes. The first pass is known as forward pass; here ANFIS computes node outputs until the fourth layer where it uses least square approach to update the consequent parameters before computing the final output. In the second pass or backward pass, the error is transmitted backward until the first layer where ANFIS employs gradient descent to tune the premise parameters. In forward pass, the antecedent part is fixed while in backward pass the consequent part is fixed. The cost of training and computational complexity of the network is ascertained by the apt choice of the number and shape of membership functions. The performance of ANFIS is measured using root mean square error (RMSE). It is the most popularly used measure. It is a measure of the differences between values predicted by the model and the actual values, as shown in (32). It calibrates the quality of the fit between the predicted model and actual data.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(actual(t) - predicted(t))^2}{n}} \tag{32}$$

where *actual (t)* and *predicted (t)* are the actual and estimated values, respectively, and *n* is the total number of records in the testing data, *t* =1,2,3… n. The overall process flow of our proposed ANFIS system for social SA is summarized in Fig. 5.3.

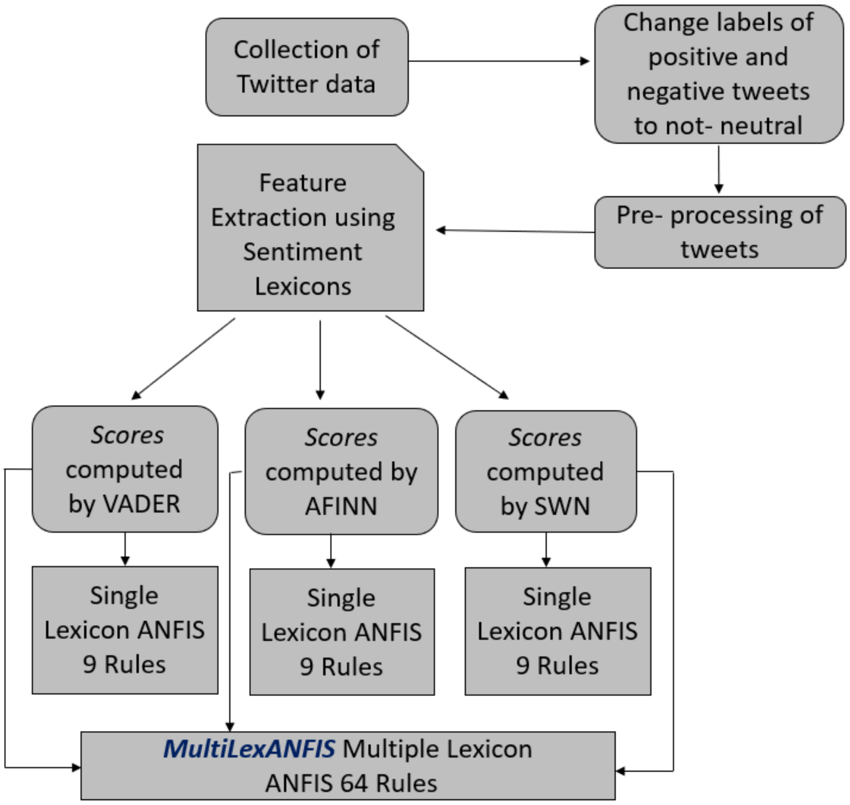


Fig. 5.3. Overall Process of our proposed ANFIS system for Social SA

5.2 EXPERIMENTAL SETUP AND IMLPLEMENTATION

In the experimental phase, the proposed ANFIS system was simulated in MATLAB with Intel Core i5 processor, 64-bit operating system and 8GB RAM. We have used a total of seven publicly available datasets: The Apple Twitter Dataset, The Nuclear Twitter Dataset, The Sanders Twitter Dataset, SemEval 2017, SemEval 2016, SemEval 2015 and The Stanford Twitter Sentiment Test Set (STS-Test) (Go *et al*., 2009). All the datasets contain three types of tweets: positive, negative and neutral. The dataset distribution of all datasets is represented in Table 5.3. We have combined the positive and negative tweets into one class: not-neutral; our aim is to categorize the tweets by ANFIS system into two classes: neutral and not-neutral. The training and testing set of these datasets are constructed using 70-30 % split ratio. Our proposed model uses inputs derived from sentiment lexicons. Sentiment Lexicons potray an important role in SA: they assign scores to words or tweets; which are further used in computing the sentiment value of each tweet. The three sentiment lexicons used in this study are: VADER (Hutto and Gilbert, 2014), AFINN (Nielsen, 2011) and SentiWordNet (Baccianella *et al*., 2010).

In our single lexicon based ANFIS variants, there are two inputs: positive and negative scores and one output for the sentiment. Class 1 indicates neutral tweets and Class 0 indicates not-neutral tweets. The number of fuzzy rules in grid-partitioning is number of MFs ^ number of input variables. In single lexicon based ANFIS variant, since each input has three membership functions (MFs), so in total ANFIS has 9 rules (3^2=9). In MultiLexANFIS there are six inputs (two from each of the three lexicons), the same output variable for sentiment and each input has two MFs. Hence it has a total of 64 rules (2^6=64). The optimal number of inputs and MFs are selected for both cases. In MultiLexANFIS, we have considered two MFs compared to three MFs in single lexicon ANFIS variants in order to reduce the curse of dimensionality. When the grid partitioning mode is applied, the number of fuzzy rules increments exponentially this scenario is known as the "curse of dimensionality". Grid partitioning creates rules by taking into account all viable combinations of membership functions of all inputs; when the count of inputs is fairly large this brings an exponential explosion. ANFIS with grid-partitioning was used to test different MFs: triangular, gaussian, and generalized bell. The optimization method used to train the ANFIS is hybrid learning, which integrates the gradient descent approach and the least squares estimate (LSE) to learn the parameters. For all the variants, the error tolerance is set to 0. While conducting experiments it was observed that training error in single lexicon ANFIS variants remain constant after 500 epochs and in MultiLexANFIS, the training error remains constant after 100 epochs. Hence, the number of epochs required by a single lexicon is 500 epochs while MultiLexANFIS require 100 epochs during the training phase.

Table 5.3. Dataset Distribution of different datasets.

	Sanders	**Nuclear**	**Apple**	**SemEval 2017**	**SemEval 2016**	**SemEval 2015**	**STS-Test (Go *et al.*, 2009)**
Positive	519	10	423	2375	5157	4377	182
Negative	572	19	1219	3972	1225	1745	177
Neutral	2333	161	2162	5937	2667	5593	139
Total	3424	190	3804	12284	9049	11715	498

5.3 RESULTS AND DISCUSSION

In this chapter we have developed two ANFIS systems for sentiment analysis of social media data: Single Lexicon based ANFIS and MultiLexANFIS, that incorporates inputs from multiple lexicons. These systems have been implemented on seven datasets. Since there are three sentiment lexicons, there are three: VADER, AFINN and SentiWordNet single lexicon based ANFIS systems. These systems have nine rules. Root Mean Square Error (RMSE) is the mostly used metric in most of the papers involving ANFIS models. Thus, we have used this as the performance metric in all experiments. All the RMSE values depicted in this work are test errors. Table 5.4 depicts the outcomes for VADER lexicon based ANFIS, the bold values highlight the lowest RMSE achieved for each dataset. Both trimf and gaussmf membership functions (MFs) have lower RMSE in a greater number of datasets compared to gbellmf. The least RMSE, 0.33174, is scored in the Nuclear dataset under gaussmf. While in the case of AFINN based ANFIS in Table 5.5, gbellmf has performed better than trimf and gaussmf; it has lowest RMSE for three datasets. The least RMSE, 0.39682, is scored in the Nuclear dataset under gbellmf. In Table 5.6, we can observe the results for SentiWordNet based ANFIS, here trimf and gbellmf have more cases of lower RMSE compared to gaussmf. Again, the Nuclear dataset has least RMSE of 0.36351under the trimf.

Our ANFIS system has been compared with various fuzzy and non-fuzzy methods. An unsupervised fuzzy rule-based system for sentiment analysis is a published previous work of authors (Vashishtha and Susan, 2019 a). This fuzzy rule-based classifier has nine rules implemented in a Mamdani fuzzy inference system. The non-fuzzy approaches are: i) VADER SA by (Hutto and Gilbert, 2014), ii) AFINN SA by Nielson (Nielsen, 2011), iii) SA using SentiWordNet lexicon by Cavalcanti et.al (Cavalcanti *et al*., 2011) and iv) SA using SentiWordNet lexicon by (Ortega *et al*., 2013). The VADER SA by (Hutto and Gilbert, 2014) is a VADER lexicon specific method; while the AFINN SA by Nielson (Nielsen, 2011) is an AFINN lexicon specific method and the rest two approaches are specific to SENTIWORDNET lexicon. The fuzzy method (Vashishtha and Susan, 2019 a) is implemented for all the three lexicons separately. These five methods are executed for all seven datasets. The lower values of RMSE exhibits better performance of the system. The execution time for single lexicon based ANFIS is about an average of 10-15 minutes.

Table 5.4. VADER Lexicon based ANFIS

Datasets	TRIMF	GBELL	GAUSS
	Root Mean Square Error (RMSE)		
Apple	0.46213	**0.45668**	0.45887
Nuclear	0.37404	0.35645	**0.33174**
Sanders	**0.43453**	0.43458	0.43683
SemEval 2017	0.47819	0.47825	**0.47814**
SemEval 2016	**0.45195**	0.45249	0.45197
SemEval 2015	**0.45257**	0.45301	0.45266
STS (Go *et al.*, 2009)	0.37445	0.36976	**0.36864**

Bold values are the lowest RMSE values for each dataset

Table 5.5. AFINN Lexicon based ANFIS

Datasets	TRIMF	GBELL	GAUSS
	Root Mean Square Error (RMSE)		
Apple	0.48046	0.50781	**0.46623**
Nuclear	0.4105	**0.39682**	0.40527
Sanders	**0.44414**	0.45402	0.44803
SemEval 2017	**0.47428**	0.47486	0.47449
SemEval 2016	0.45577	**0.45331**	0.45535
SemEval 2015	0.44567	0.44541	**0.44535**
STS (Go *et al.*, 2009)	0.40142	**0.38946**	0.39323

Bold values are the lowest RMSE values for each dataset

Table 5.6. SENTIWORDNET Lexicon based ANFIS

Datasets	TRIMF	GBELL	GAUSS
	Root Mean Square Error (RMSE)		
Apple	0.4878	**0.48503**	0.4866
Nuclear	**0.36351**	0.36638	0.36921
Sanders	0.45498	0.45341	**0.45318**
SemEval 2017	0.49336	**0.49243**	0.49294
SemEval 2016	**0.45204**	0.4526	0.45576
SemEval 2015	0.48483	**0.4844**	0.48484
STS (Go *et al.*, 2009)	**0.42304**	0.42831	0.42352

Bold values are the lowest RMSE values for each dataset

Comparison of VADER lexicon based ANFIS system with VADER specific approaches is shown in Table 5.8. It can be clearly observed the our ANFIS system outperforms all methods in all datasets. For instance, in Apple dataset RMSE of ANFIS is 0.45668 while other methods have 0.6127 and 0.6223. Table 5.9 presents the achievement of AFINN lexicon based ANFIS system with AFINN specific methods. Our ANFIS has the lowest RMSE compared to all non-fuzzy and fuzzy methods in all datasets. In SemEval 2017, ANFIS scored 0.47428 while others scored 0.6127 and 0.7046. Similar observations can be made for SentiWordNet lexicon based ANFIS with SENTIWORDNET specific approaches in Table 5.10. Here also, ANFIS surpasses all methods, for example in SemEval 2016, RMSE for ANFIS is 0.45204 and for others: 0.554, 0.7164 and 0.8115. Hence, we can deduce that all single lexicon based ANFIS systems have lowest RMSE for all datasets compared to existing approaches. But it is difficult to decide the best membership function (MF) for each lexicon among all datasets. In other words, we can't conclude which is best MF among trimf, gbellmf and gaussmf in VADER, AFINN or SentiWordNet lexicons. In VADER, there is a confusion which to choose among trimf or gaussmf; in AFINN, we can choose gbellmf as it performs slightly better; while in SentiWordNet it is hard to choose between trimf or gbellmf.

To solve the above confusion of choosing the best membership function (MF) for a given lexicon, a MultiLexANFIS architecture has been proposed. This system integrates all the lexicons into a single unit. Using each lexicon, two scores: *TweetPos*, positive score and

TweetNeg, negative score, are computed as discussed in Section 3. Hence, this integrated system has a total of six inputs, with two MFs for each input, 64 rules and one sentiment label as output.

Table 5.7 displays the performance of MultiLex ANFIS for all datasets. This system has been compared with single lexicon based ANFIS system as well, the best-cum-lowest RMSE-combination of lexicon-membership function for each dataset is represented in second column. In Apple, Nuclear, SemEval 2015 and STS datasets, MultiLexANFIS has 0.448, 0.320, 0.433 and 0.366 RMSE compared to 0.457 of VADER-gbellmf, 0.332 of VADER-gaussmf, 0.445 of AFINN-gaussmf and 0.369 of VADER-gaussmf respectively. This integrated ANFIS achieves lower RMSE compared to the single lexicon based ANFIS for four datasets, while for others the results are comparable. We can also note down that trimf: triangular membership function performs the best in MultiLexANFIS, gaussmf: gaussian membership function is at second position. Thus, the problem of confusion among the MFs for all datasets has been solved. Comparison amidst the single lexicon ANFIS systems reveals that VADER lexicon performs best. This can be observed in the second column of Table 5.7; among seven datasets, VADER lexicon has the lowest RMSE for five datasets. The execution time of MultiLexANFIS is about 30 minutes. The combined time complexity of three single lexicon based ANFIS will be 45 minutes (3x15), while the integrated ANFIS is 15 minutes faster. The training error graph for MultiLexANFIS for Apple dataset with gaussmf is shown in Fig. 5.4. The computed test error, 0.448, is depicted in Fig. 5.5.

One of the benefits of our proposed system is the determination of the best performing membership function for a given dataset and lexicon combination. It produces lower error compared to other models and it takes less execution time in contrast to single lexicon based ANFIS. Our model yields better results by using scores from sentiment lexicon as features for subjectivity classification (neutral and non-neutral) compared to fuzzy rule classifier proposed by the authors (Vashishtha and Susan, 2019 a), which uses the same scores for polarity classification (positive, negative or neutral).

Table 5.7. MultiLexANFIS: Multiple Lexicons Based ANFIS

Datasets	Least among Lex	TRIMF	GBELL	GAUSS
	Lex MF RMSE	RMSE	RMSE	RMSE
Apple	V Gbell- 0.457	0.475	0.452	**0.448**
Nuclear	V Gauss- 0.332	0.374	0.356	**0.320**
Sanders	V Trimf- 0.435	**0.440**	0.494	0.519
SemEval 2017	A Trimf- 0.474	**0.473**	0.474	0.473
SemEval 2016	V Trimf- 0.452	**0.455**	0.456	**0.454**
SemEval 2015	A Gauss 0.445	**0.433**	0.447	0.445
STS (Go *et al*., 2009)	V Gauss- 0.369	0.380	**0.366**	0.376

Bold values are the lowest RMSE value for each dataset. V stands for VADER and A stands for AFINN

Table 5.8. Comparison of VADER Lexicon Based ANFIS with VADER specific methods

Dataset	Methods	RMSE
Apple	VADER SA (Hutto and Gilbert, 2014)	0.6127
	Fuzzy Rule (Vashishtha and Susan, 2019 a)	0.6223
	ANFIS	**0.45668**
Nuclear	VADER SA (Hutto and Gilbert, 2014)	0.8013
	Fuzzy Rule (Vashishtha and Susan, 2019 a)	0.4168
	ANFIS	**0.33174**
Sanders	VADER SA (Hutto and Gilbert, 2014)	0.6353
	Fuzzy Rule (Vashishtha and Susan, 2019 a)	0.5593
	ANFIS	**0.43453**
SemEval 2017	VADER SA (Hutto and Gilbert, 2014)	0.6137
	Fuzzy Rule (Vashishtha and Susan, 2019 a)	0.6892
	ANFIS	**0.47814**
SemEval 2016	VADER SA (Hutto and Gilbert, 2014)	0.6309
	Fuzzy Rule (Vashishtha and Susan, 2019 a)	0.8078
	ANFIS	**0.45195**
SemEval 2015	VADER SA (Hutto and Gilbert, 2014)	0.5764
	Fuzzy Rule (Vashishtha and Susan, 2019 a)	0.689
	ANFIS	**0.45257**
STS (Go *et al*., 2009)	VADER SA (Hutto and Gilbert, 2014)	0.4251
	Fuzzy Rule (Vashishtha and Susan, 2019 a)	0.7525
	ANFIS	**0.36864**

Bold values are the lowest RMSE values for Bold values are the lowest RMSE value for each dataset. V stands for VADER and A stands for AFINN

Table 5.9. Comparison of AFINN Lexicon Based ANFIS with AFINN specific methods

Dataset	Methods	RMSE
Apple	AFINN SA (Nielsen, 2011)	0.6112
	Fuzzy Rule (Vashishtha and Susan, 2019 a)	0.6402
	ANFIS	**0.46623**
Nuclear	AFINN SA (Nielsen, 2011)	0.846
	Fuzzy Rule (Vashishtha and Susan, 2019 a)	0.4413
	ANFIS	**0.39682**
Sanders	AFINN SA (Nielsen, 2011)	0.6235
	Fuzzy Rule (Vashishtha and Susan, 2019 a)	0.5634
	ANFIS	**0.44414**
SemEval 2017	AFINN SA (Nielsen, 2011)	0.6127
	Fuzzy Rule (Vashishtha and Susan, 2019 a)	0.7046
	ANFIS	**0.47428**
SemEval 2016	AFINN SA (Nielsen, 2011)	0.6591
	Fuzzy Rule (Vashishtha and Susan, 2019 a)	0.8303
	ANFIS	**0.45331**
SemEval 2015	AFINN SA (Nielsen, 2011)	0.5649
	Fuzzy Rule (Vashishtha and Susan, 2019 a)	0.718
	ANFIS	**0.44535**
STS (Go *et al*., 2009)	AFINN SA (Nielsen, 2011)	0.4321
	Fuzzy Rule (Vashishtha and Susan, 2019 a)	0.8287
	ANFIS	**0.38946**

Bold values are the lowest RMSE values for each dataset

Table 5.10. Comparison of SENTIWORDNET Lexicon Based ANFIS with SENTIWORDNET specific methods

Dataset	Methods	RMSE
Apple	Cavalcanti (Cavalcanti *et al*., 2011)	0.7437
	Ortega (Ortega *et al*., 2013)	0.6449
	Fuzzy Rule (Vashishtha and Susan, 2019 a)	0.655
	ANFIS	**0.48503**
Nuclear	Cavalcanti (Cavalcanti et al., 2011)	0.9205
	Ortega (Ortega et al., 2013)	0.8553
	Fuzzy Rule (Vashishtha and Susan, 2019 a)	0.4168
	ANFIS	**0.36351**
Sanders	Cavalcanti (Cavalcanti et al., 2011)	0.795
	Ortega (Ortega et al., 2013)	0.6202
	Fuzzy Rule (Vashishtha and Susan, 2019 a)	0.5655
	ANFIS	**0.45318**
SemEval 2017	Cavalcanti (Cavalcanti et al., 2011)	0.6892
	Ortega (Ortega et al., 2013)	0.6693
	Fuzzy Rule (Vashishtha and Susan, 2019 a)	0.7176
	ANFIS	**0.49243**

SemEval 2016	Cavalcanti (Cavalcanti et al., 2011)	0.554
	Ortega (Ortega et al., 2013)	0.7164
	Fuzzy Rule (Vashishtha and Susan, 2019 a)	0.8115
	ANFIS	**0.45204**
SemEval 2015	Cavalcanti (Cavalcanti et al., 2011)	0.6839
	Ortega (Ortega et al., 2013)	0.6619
	Fuzzy Rule (Vashishtha and Susan, 2019 a)	0.7222
	ANFIS	**0.4844**
STS (Go et al., 2009)	Cavalcanti (Cavalcanti et al., 2011)	0.5226
	Ortega (Ortega et al., 2013)	0.6555
	Fuzzy Rule (Vashishtha and Susan, 2019 a)	0.7418
	ANFIS	**0.42304**

Bold values are the lowest RMSE values for each dataset

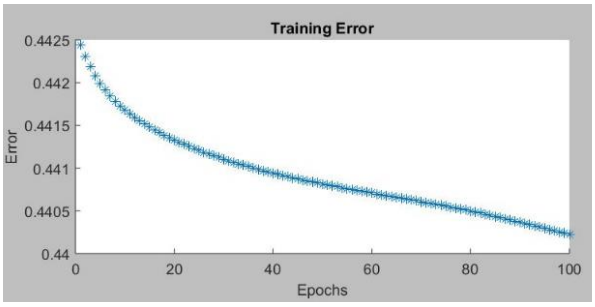


Fig. 5.4. Training Error for Apple Dataset with gaussmf in MultiLexANFIS

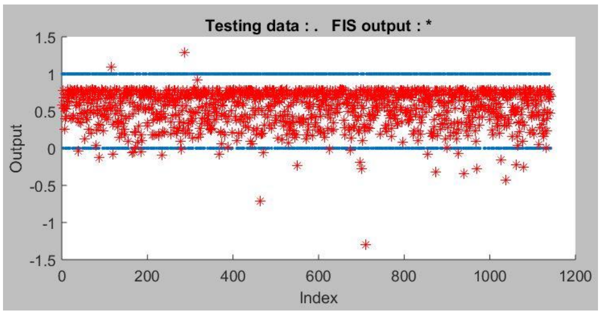


Fig. 5.5. Test Error for Apple Dataset with gaussmf in MultiLexANFIS*(red dots indicate the predicted values for testing data)*

5.4 CONCLUSION

In this chapter, a novel MultiLexANFIS system has been introduced for social sentiment analysis. This architecture is an Adaptive Neuro-Fuzzy Inference System (ANFIS) that incorporates inputs from multiple lexicons to perform SA of social media content. The key highlights of this work are: 1) a novel neuro-fuzzy system: MultiLexANFIS that combines sentiment scores from multiple lexicons to classify tweets, 2) a novel set of 64 rules for Sugeno-type Fuzzy Inference System, 3) single lexicon based ANFIS to classify tweets and 4) comparison of fuzzy approaches with non-fuzzy approaches affirms the supremacy of our proposed neuro-fuzzy system. ANFIS has deployed optimization with help of a hybrid learning process; that uses the least square approach and gradient descent to tune parameters.

We can deduce that all single lexicon based ANFIS systems and MultiLexANFIS have the lowest RMSE for all datasets compared to other approaches. Results reveal that among the single lexicon ANFIS systems VADER lexicon performs best. The best membership in MultiLexANFIS is triangular (trimf), and gaussian membership function (gaussmf) is at second position. The problem of confusion among the MFs for all datasets has been solved. The MultiLexANFIS is computationally faster than the total combination of three single lexicon based ANFIS for all lexicons. Our neuro-fuzzy network is targeted for applications that aim to search for neutral social media content posted by users. Our results thus indicate an effective and computationally feasible solution to social sentiment analysis that incorporates multiple lexicons onto the same platform. These proposed novel rules are domain-independent. We can apply these rules to any textual data that employs lexicons to extract positive and negative scores for each textual document or sentence or review. The proposed system can handle only those Twitter datasets which contain neutral tweets. At present, it cannot deal with mixed tweets, i.e. the tweets that include positive as well as the negative sentiment. Eventually, later in the future, we will compute the sentiment of mixed tweets by enhancing our model using deep neural networks.

CHAPTER 6

ANALYSIS OF SENTIMENT FROM SOCIAL MEDIA POSTS BASED ON FUZZY LOGIC

In this chapter[1], we evaluate the sentiment of social media data, posted by users, using a novel set of fuzzy rules associated with multiple lexicons and datasets. The proposed fuzzy system combines Natural Language Processing techniques and word sense disambiguation by applying a novel unsupervised nine fuzzy rule-based system to categorize the post into: positive, negative, or neutral sentiment classes. A comparative analysis of our method on nine public Twitter datasets, three sentiment lexicons, four state-of-the-art approaches for unsupervised Sentiment Analysis, and one state-of-the-art method for supervised machine learning has been performed. The results of this system can give a vision to researchers to select the best lexicon for handling social media. The proposed FIS is suitable for any lexicon and any two-class or three-class sentiment dataset. The blending of fuzzy logic concepts with sentiment lexicons for classifying sentiment has set up a new exemplar in the field of Sentiment Analysis.

:[1] The contents of this chapter are published in "Fuzzy rule based unsupervised sentiment analysis from social media posts." *Expert Systems with Applications* 138 (2019): 112834 and "Fuzzy logic based dynamic plotting of mood swings from tweets." In *International Conference on Innovations in Bio-Inspired Computing and Applications*, pp. 129-139. Springer, Cham, 2018.

The main contributions of this chapter are: i) formulation of nine fuzzy rules to compute sentiment of each tweet ii) the proposed unsupervised approach is suitable for any sentiment lexicon iii) also suitable for any two-class dataset or three-class dataset iv) comparison of our proposed rule-based approach for Sentiment Analysis with four state-of-the-art methods for unsupervised sentiment classification and one state-of-the-art method for supervised machine learning. The experimental results on various datasets produce a higher performance for our system in comparison to the state-of-the-art. The rest of the chapter is organized as follows. Our proposed fuzzy rule-based system is presented in Section 6.1. Section 6.2 is about the experimental setup & implementation. Results are discussed in Section 6.3. A fuzzy logic-based method for plotting mood swings is explained in Section 6.4. The overall conclusions are drawn in Section 6.5.

6.1 PROPOSED FUZZY RULE SYSTEM FOR SENTIMENT ANALYSIS

In this section, we introduce the details of the proposed fuzzy logic-based model. Fig. 6.1 describes the framework of a fuzzy logic-based model. Fuzzification is the process of transforming a crisp quantity into fuzzy. The crisp or real inputs are mapped to fuzzy sets whose elements have a degree of membership computed using fuzzy membership functions (MF). In this work, we select the triangular-fuzzy membership function because it is easy to understand and popular in use.

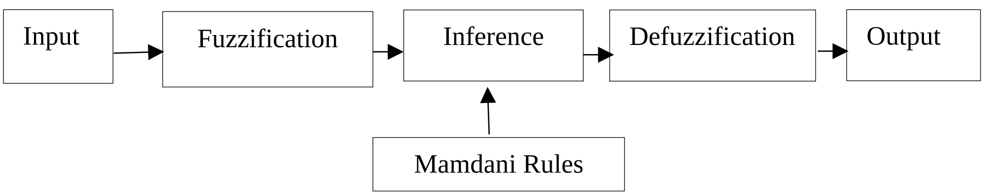


Fig. 6.1. *The framework of using a fuzzy logic-based model*

In the field of artificial intelligence, human knowledge can be easily represented by converting it into natural language expressions, further these expression can be depicted in the format of

IF-THEN rules (Zadeh, 1975). Complex systems can be represented in the form of linguistic variables with the help of fuzzy rule-based system. The representation of these linguistic variables is in the form of fuzzy sets and logical connectives of these sets (Zadeh, 1975). The three common methods of deductive inference for fuzzy systems based on linguistic rules are: (1) Mamdani systems, (2) Sugeno models, and (3) Tsukamoto models. In our work, we have used the commonly used Mamdani systems, developed by Mamdani and Assilian in 1975 (Mamdani *et al.*,1975). This is similar to a dual-input and single-output fuzzy system. A fuzzy system with two non-interactive inputs A and B (antecedents) and a single output C (consequent) is expressed by a number of r linguistic IF–THEN propositions in the Mamdani form:

$$\text{Rule } R_j\text{: } IF\ A\ is\ A_1^j\ and\ B\ is\ B_1^j\ THEN\ C\ is\ C_1^j,\ j = 1,2,\ldots r \qquad (6.1)$$

where A_1^j and B_1^j are the fuzzy sets representing the j^{th} antecedent or premise pairs and C_1^j is the fuzzy set representing the j^{th} consequent. We have used the max-min inference method. It is a popular inference method in fuzzy systems (Liu *et al.*, 2017), (Jefferson *et al.*, 2017), (Ishibuchi, *et al.*, 2001). The fuzzy output is obtained by applying the rules to fuzzy input. This output can be defuzzied using defuzzification methods. Defuzzification is the conversion process of a fuzzy quantity to an exact quantity. Some of the defuzzification methods are: mean of maximum (MOM), centroid, smallest of maximum (SOM), bisector, and largest of maximum (LOM) (Hellendoorn *et al.*, 1993). We have used centroid defuzzification method as it gives the best results. Our approach is based on an unsupervised strategy consisting of three major phases: text pre-processing, use of sentiment lexicon, and fuzzy rule system for sentiment polarity classification.

6.1.1 TEXT PRE-PROCESSING

The social media text is of limited size. On Twitter, the character limit for tweets is 280 characters. It was earlier limited to 140 characters. Users post additional information which depicts sentiment, using abbreviations, emoticons, hashtags, slang, or URLs. Thus, the text needs to be pre-processed to get relevant and useful information by removing the noisy data. First of all, we have eliminated URLs, and '@' symbol used to mention user names because they don't carry any sentiment. We have re-phrased commonly used phrases (like "can't") with their grammatical form (“can not”). Tokens consisting of “#” (hashtags), usually depict an emotion or opinion about the tweet’s topic, so we discard only the “#”.

6.1.2 USE OF SENTIMENT LEXICON

A sentiment lexicon contains a list of lexical features (e.g., words) which are usually labelled according to their semantic orientation as either positive or negative. Here, we have investigated three different sentiment lexicons: SentiWordNet (Baccianella *et al.*, 2010), AFINN (Nielsen *et al.*, 2011), and VADER (Hutto and Gilbert, 2014) in isolation from each other. The pre-processed text is used along with these lexicons to compute the positive and negative scores for each tweet. SentiWordNet is an extension of WordNet in which 147,306 synsets are annotated with three numerical scores relating to positivity, negativity, and objectivity (neutrality). It has high coverage of terms. Each score has range 0.0 - 1.0, and the sum of the positive, negative and objective scores is 1.0 for each synset (Baccianella *et al.*, 2010). It is a useful and popular lexicon for a wide range of tasks in text mining.

We interface with SentiWordNet via Python’s Natural Language Toolkit (Bird *et al.*, 2009). The method which uses SentiWordNet lexicon includes pre-processing of text: removal of stopwords, removal of punctuations, lemmatization, Part of Speech (POS) tagging by NLTK (Bird *et al.*, 2009), and Word Sense Disambiguation (WSD) by Lesk (Banerjee *et al.*, 2002). In a sentence, a word can have multiple contextual meanings, word-sense disambiguation is a process that recognizes the correct sense of a word among multiple senses. The aim of the WSD process is to detect the best <*word, POS-tag, sense*> match for each of the <*word, POS-tag*> pairs received as input. The SentiWordNet method obtains the scores of each word from this lexicon using *syn.pos_score ()* and *syn.neg_score ()*. Each word has a positive and negative score (6.2 and 6.3) computed using WSD, which can be interpreted as a fuzzy membership pertaining to the fuzzy sets *Pos* and *Neg* (6.4 and 6.5). The words which are having higher positive score than a negative score in a tweet are summed up to compute the positive score (*TweetPos*) of the tweet (6.6). Similarly, words which are having higher negative scores than positive scores in a tweet are summed up to compute the negative score (*TweetNeg*) of the tweet (6.7). These scores are computed for all tweets.

$$\mu_{Pos}(a) = syn.pos\_score() \tag{6.2}$$

$$\mu_{Neg}(a) = syn.neg\_score() \tag{6.3}$$

$$Pos = \{(a, \mu_{Pos}(a)\}, a \in X_i \tag{6.4}$$

$$Neg = \{(a, \mu_{Neg}(a)\}, a \in X_i \tag{6.5}$$

$$\begin{aligned} &if\,(\mu_{Pos}(a) > 0 \,\&\, \mu_{Pos}(a) > \mu_{Neg}(a)) \\ &then(TweetPos = \sum_{a=1}^{m} \mu_{Pos}(a)) \end{aligned} \tag{6.6}$$

$$\begin{aligned} &if\,(\mu_{Neg}(a) > 0 \,\&\, \mu_{Neg}(a) > \mu_{Pos}(a)) \\ &then(TweetNeg = \sum_{a=1}^{m} \mu_{Neg}(a)) \end{aligned} \tag{6.7}$$

where *a* is a word in a tweet, *m* is the number of selected words and X_i is the set of total words.

The AFINN lexicon is a list of English terms manually rated for valence with an integer between -5 (negative) and +5 (positive) by Finn Årup Nielsen in 2011. This lexicon is equipped to handle modern-day tweets due to its inclusion of Internet slang and obscene words. It has been created especially for Sentiment Analysis in microblogs, so we have included AFINN as one of the lexicons for our Twitter datasets. The AFINN method fetches the score of each word using AFINN lexicon (6.8), if it is greater than 0 it is a positive word, and if less than 0 it is a negative word. Each word has a positive and negative score and can be interpreted as a fuzzy membership pertaining to the fuzzy sets *Pos* and *Neg* (6.4 and 6.5). The positive words are summed up to compute the positive score of the tweet (6.11); similarly, the negative score is computed for each tweet (6.12).

$$\mu(a) = af.score(a) \tag{6.8}$$

$$if\,(\mu(a) > 0) then(\mu_{Pos}(a) = \mu(a)) \tag{6.9}$$

$$if\,(\mu(a) < 0) then(\mu_{Neg}(a) = -\mu(a) \tag{6.10}$$

$$TweetPos = \sum_{a=1}^{m} \mu_{Pos}(a) \tag{6.11}$$

$$TweetNeg = \sum_{a=1}^{m} \mu_{Neg}(a) \tag{6.12}$$

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based SA tool that is specifically accustomed to sentiment expressed in social media, it was created in 2014 (Hutto and Gilbert, 2014). It is quick and computationally economical without sacrificing accuracy. It works excellently well on social media text. It doesn't require any training data. It does not severely suffer from a speed-performance tradeoff. These factors inspired us to include this lexicon for our Twitter datasets. The VADER method computes the score of the overall tweet using VADER lexicon's *polarity_scores(a)* method and gives positive (*TweetPos*) and negative (*TweetNeg*) scores of a tweet as output.

6.1.3 Fuzzy Rule System

We have used one of the popularly used fuzzy inference techniques called the Mamdani fuzzy model. The Mamdani fuzzy model is performed in four stages, first stage is Fuzzification of input variables, the second stage is Rule evaluation, the next stage is Aggregation of the rule outputs and the last stage is Defuzzification.

6.1.3.1 Fuzzification

The positive and negative score of each tweet obtained from the second phase is fuzzified using the triangular membership function. When the triangular fuzzy membership is used, each linguistic term T involves three key points, d, *e, f* associated with the change of pattern of the fuzzy membership. A membership function (MF) for a fuzzy set *S* on the universe of discourse X is defined as $\mu_S : X \rightarrow [0,1]$, where each element of X is mapped to a value between 0 and 1. Following is the equation for the triangular function defined by a lower limit *d*, an upper limit *f*, and an intermediate value *e*, where $d < e < f$:

$$\mu_S(x) = \begin{cases} 0, x <= d \\ (x-d)/(e-d), d < x <= e \\ (f-x)/(f-e), e < x <= f \\ 0, x >= f \end{cases} \tag{6.13}$$

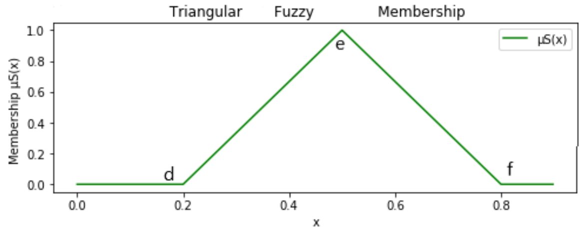


Fig. 6.2. *Triangular Fuzzy Membershi6*

The triangular fuzzy membership is graphically presented in Fig. 6.2, where the parameters are: d=0.2, e=0.5 and f=0.8 Three fuzzy sets: Low (L), Medium (M) and High (H) are created using triangular fuzzy membership for universe variables: positive (*x_p)*, negative (*x_n)* and output (*x_op)*. The range of *x_op* is (0-10) fixed for all lexicons. The range of *x_p* and x_n is calculated for each (dataset, lexicon) combination. We compute the global minimum (*min*), global maximum (*max*) values for all positive scores, *TweetPos* and all negative scores, *TweetNeg* of all tweets in a dataset. The range of *x_p* and *x_n* is (min, max). The *mid* value is calculated as:

$$mid = (\min + \max)/2 \tag{6.14}$$

The parameters required for building the triangular fuzzy membership for the fuzzy sets Low, Medium and High are: Low: {min, min, mid}; Medium: {min, mid, max}; High: {mid, max, max}. For the output variable, x_op, *min*=0 and *max*=10, thus range is 0-10 and the parameters for three fuzzy sets (Negative, Neutral and Positive) which depict the sentiment class are: Negative(*op_neg*): {0,0,5}; Neutral(*op_neu*): {0,5,10}; Positive(*op_pos*): {5,10,10}; *op_neu*, *op_neg* and *op_neu* are the MFs of consequent parts of proposed rules. These are graphically presented in Fig. 6.3.

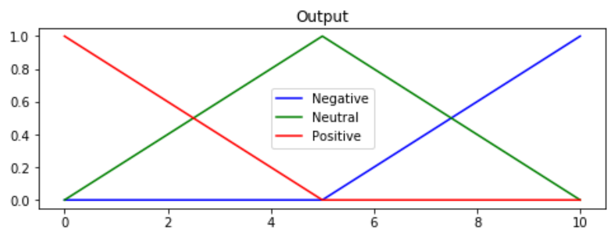


Fig. 6.3. *Triangular Fuzzy Membership sets for output variables*

6.1.3.2 Formulating the Rule Base

The novelty of this work is the proposal of nine rules, displayed in Table 6.1. Fig. 6.4 shows the visualization of our nine rules obtained by the intersection of two input variables (positive (*TweetPos)* and negative (*TweetNeg*) scores of a tweet), each with three fuzzy subsets. Every data point activates one and only one rule. The rules were devised based on the assumption that a higher score (positive or negative) indicates the sentiment. In the case of common scores, the sentiment is neutral. The rule evaluation is done on the basis of Table 6.1 and Fig. 6.4.

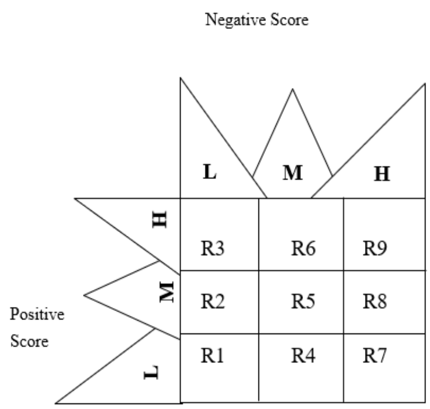


Fig. 6.4. *Visualization of nine rules*

Table 6.1. The proposed nine Mamdani rules.

Rule	Positive Score	Negative Score	Sentiment
R1	Low	Low	Neutral
R2	Medium	Low	Positive
R3	High	Low	Positive
R4	Low	Medium	Negative
R5	Medium	Medium	Neutral
R6	High	Medium	Positive
R7	Low	High	Negative
R8	Medium	High	Negative
R9	High	High	Neutral

$$w_{R1} = pos\_low \wedge neg\_low \tag{6.15}$$

$$w_{R2} = pos\_med \wedge neg\_low \tag{6.16}$$

$$w_{R3} = pos\_high \wedge neg\_low \tag{6.17}$$

$$w_{R4} = pos\_low \wedge neg\_med \tag{6.18}$$

$$w_{R5} = pos\_med \wedge neg\_med \tag{6.19}$$

$$w_{R6} = pos\_hi \wedge neg\_med \tag{6.20}$$

$$w_{R7} = pos\_low \wedge neg\_high \tag{6.21}$$

$$w_{R8} = pos\_med \wedge neg\_high \tag{6.22}$$

$$w_{R9} = pos\_high \wedge neg\_high \tag{6.23}$$

These equations (6.15-6.23) depict the nine rules $w_{R1} ... w_{R9}$ depict the firing strength of each rule and the symbol $\wedge$ represents fuzzy AND operator. The variables *pos_low, pos_med* and *pos_high* constitute the antecedent part of the fuzzy rules and they depict the low, medium and high fuzzy sets for the positive score *TweetPos*, respectively. Similarly, *neg_low, neg_med* and neg*_high* constitute the antecedent part of the fuzzy rules and they depict the low, medium, and high fuzzy sets for the negative score *TweetNeg*, respectively.

6.1.3.3 Aggregation of Rule outputs:

$$w_{neg} = w_{R4} \vee w_{R7} \vee w_{R8} \tag{6.24}$$

$$w_{neu} = w_{R1} \vee w_{R5} \vee w_{R9} \tag{6.25}$$

$$op\_activation\_low = w_{neg} \wedge op\_neg \tag{6.27}$$

$$op\_activation\_med = w_{neu} \wedge op\_neu \tag{6.28}$$

$$op\_activation\_high = w_{pos} \wedge op\_pos \tag{6.29}$$

$$aggregated = op\_activation\_low \cup op\_activation\_med \cup op\_activation\_high \tag{6.30}$$

In equations (6.24-6.26) w_{neg} depicts the overall firing strength or degree of fulfilment of the fuzzy rules pertaining to negative emotion, similarly w_{neu} and w_{pos} are for neutral and positive emotion respectively. These overall firing strengths represent the extent to which the antecedent part of the rule is contented (Jang *et al.*, 1997). In (6.27-6.29) o*p_neg, op_neu* and *op_pos* are the MFs of consequent parts of respective rules (6.15-6.23). The induced or resultant consequents MFs (*op_activation_low*, *op_activation_med* and *op_activation_high*) are computed by clipping the MFs of consequent parts with overall firing strength, given by (6.27-6.29). Overall output MF is obtained by aggregating the induced consequent MFs using union operator in (6.30).

6.1.3.4 Defuzzification

The final stage in fuzzy rule system is defuzzification. We have implemented the centroid defuzzification method as it delivers reliable results (Jang *et al.*, 1997). It returns the centre of area (COA) under the curve (Hellendoorn *et al.*, 1993). This method yields a discrete value based on the centre of gravity of the fuzzy set. The aggregated output (μ_A) computed in (6.30) is used to calculate the defuzzified output in (6.31), where z indicates the sample value in output variable, *x_op* described in section 6.1.3.1.

$$COA = \frac{\sum z\mu_A(z)}{\sum \mu_A(z)} \tag{6.31}$$

Finally, the defuzzified output is checked for different ranges to classify the tweet according to its polarity: Negative, Neutral or Positive class in (6.32). Since *min*=0 and *max*=10 for output range, we equally divide this range into three parts. Negative: 0-*max*/3, Neutral: (*max*)/3-2/3(*max*) and Positive: 2/3(*max*)-*max*. The overall process flow of unsupervised fuzzy rule system for Sentiment Analysis is presented in Fig 6.5.

$$Output = \begin{Bmatrix} Negative, 0 < COA < 3.3 \\ Neutral, 3.3 < COA < 6.7 \\ Positive, 6.7 < COA < 10 \end{Bmatrix} \tag{6.32}$$

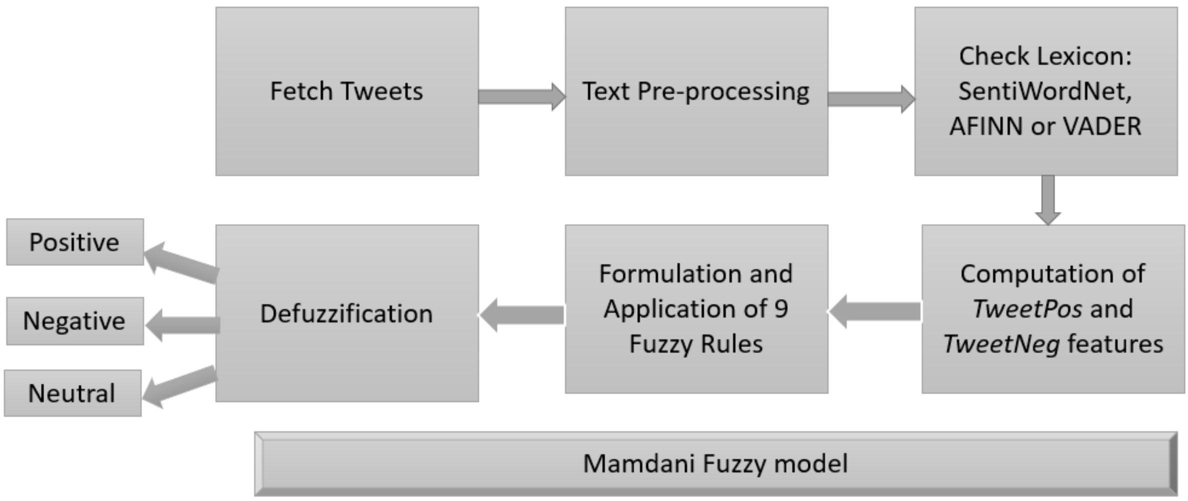


Fig. 6.5. Process Flow of Unsupervised Fuzzy Rule System for Sentiment Analysis

We have next explicitly compared two papers that use fuzzy inferencing: 1) (Srivastava *et al.*, 2013) and 2) (Haque *et al.*, 2014) with proposed unsupervised fuzzy rule system.

1. In (Srivastava *et al.*, 2013) they have constructed their own lexicon Opinion Words Lexicon (OWL) by performing some modifications on SentiWordNet data. This approach is SentiWordNet dependent. On the other hand, our fuzzy approach can be used with any lexicon: SentiWordNet, AFINN and VADER. POS (Part of Speech) Tagger is applied to extract only adverbs and adjectives, while our method focuses on noun, verb, adjective and adverbs. Using OWL, two fuzzy sets are created: positive opinion words and negative opinion words. The output is positive or negative polarity. In our fuzzy approach the input: positive and negative scores for each tweet are represented using Low, Medium, and High fuzzy sets; the output: negative, neutral, or positive sentiment. We detect the neutral sentiment while the previous work doesn't. In their approach, the fuzzy memberships of words are modified using a linguistic hedge. Overall aggregated output is achieved by taking the average sum of scores. We use a fuzzy rule based system to detect the final polarity of the tweet. The aggregation involves the union of output activation levels low, medium, and high.

2. In (Haque *et al.*, 2014), they have used SentiWordNet lexicon. The sentiment score for each term in the tweet is computed as the difference of positive and negative scores obtained from the lexicon. Our approach can be used with any lexicon: SentiWordNet, AFINN and VADER. Weights are assigned manually to the frequently used terms. The tweets are simply classified as positive or negative by calculating the sum of sentiment scores and checking its range. In our approach, the input: positive and negative scores for each tweet, are represented using Low, Medium and High fuzzy sets; the output: negative, neutral or positive sentiment. We use fuzzy rule system to detect the polarity. They have used 100 tweets for analysis while our approach has been applied to multiple datasets containing thousands of tweets. Their approach classifies tweets into positive or negative while our approach classifies tweets into positive, negative, or neutral.

Hence, we can observe our fuzzy approach is different and more scalable as it takes into account: three polarity classes, computes the level of positive and negative scores as Low, Medium and High. It can be used with any lexicon: SentiWordNet, AFINN and VADER and can be applied to both two class (polarity) or three class (polarity) dataset.

6.2 EXPERIMENTAL SETUP AND IMPLEMENTATION

This section reports the experimental setup and implementation of the proposed fuzzy rule-based classifier for Sentiment Analysis. We have implemented our fuzzy rule-based system in python version 3.6.5. The system has as Intel Core i5 processor, 64-bit operating system and 8GB RAM. The code containing the implementation of our work is available online[2]. Most of the papers use the Twitter API to extract tweets but we have used publicly available datasets. In this chapter, we have used a total of nine benchmark datasets: The Sanders Twitter Dataset, The Nuclear Twitter Dataset, The Apple Twitter Dataset, The Stanford Twitter Sentiment Test Set (STS-Test), The Sentiment140 Twitter Dataset, SemEval 2017, SemEval 2016, SemEval 2015 and Twitter data used by (Hutto and Gilbert, 2014). The Sanders Twitter Dataset consists of tweets on four different topics (Apple, Google, Microsoft, and Twitter) ("Sanders Twitter Dataset", 2019). Each tweet was manually labelled by one annotator as either positive, negative, neutral, or irrelevant with respect to the topic. We have not considered the irrelevant tweets. The Nuclear Twitter dataset is collection of tweets related to nuclear energy ("Nuclear

Twitter Dataset", 2019). The Apple Twitter dataset is a collection of tweets about Apple products and company ("Apple Twitter Dataset", 2019). The Stanford Twitter sentiment corpus (sentiment 140) was developed by (Go *et al.*, 2009). It consists of two different sets, training and test. The training set tweets are automatically annotated based on emoticons while the test set tweets are manually annotated. The tweets in the test set were collected by searching Twitter API with particular queries including names of products, people and companies. All the datasets are three-class (i.e. positive, negative and neutral) except for Sentiment140 training dataset which is two-class (i.e. positive and negative). The distribution of tweets in different datasets according to sentiment classes: positive, negative and neutral is specified in Table 6.2. Furthermore, we have used various SemEval twitter datasets: SemEval-2017 Task 4, subtaskA decides whether a given tweet conveys positive, negative or neutral sentiment (Rosenthal *et al.*, 2017); SemEval-2016 Task4 decides whether a given tweet and a topic, the sentiment conveyed towards that topic on a three-point scale: positive, negative or neutral (Nakov *et al.*,2016); SemEval-2015 Task 10 decides given a tweet, figure outs whether it expresses a positive, a negative, or a neutral/objective sentiment (Rosenthal *et al*.,2015). The last twitter dataset is obtained from (Hutto and Gilbert, 2014). Table 6.3 represents the distribution of tweets in these datasets. We have used three different sentiment lexicons: SentiWordNet (Baccianella *et al.*,2010), AFINN (Nielsen, 2011) and VADER (Hutto and Gilbert, 2014); these are described in section 6.1.2.

Table 6.2. Dataset Distribution of different datasets.

	Sanders Twitter	**Nuclear Twitter**	**Apple Twitter**	**STS-Test**	**Sentiment140**
Positive	519	10	423	182	248576
Negative	572	19	1219	177	799999
Neutral	2333	161	2162	139	Null
Total	3424	190	3804	498	1048575

[2]: https://github.com/SrishtiVashishtha/Fuzzy-Rule-based-Unsupervised-Sentiment-Analysis-from-Social-Media-Posts.

Table 6.3. Dataset Distribution of SemEval and Gilbert datasets.

	SemEval 2017	**SemEval 2016**	**SemEval 2015**	**Gilbert Tweets**
Positive	2375	5157	4377	2742

Negative	3972	1225	1745	1219
Neutral	5937	2667	5593	239
Total	12284	9049	11715	4200

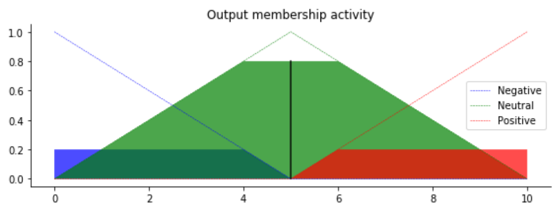
6.3 Results and Discussion

6.3.1 Processing of a single tweet

In this section, we present demonstrate a single tweet is being processed by our proposed fuzzy rule based unsupervised Sentiment Analysis model. Processing of a sample tweet of Nuclear twitter dataset ("Nuclear Twitter Dataset", 2019) using VADER lexicon (Hutto and Gilbert, 2014) is shown in Fig 6.6. Initially text preprocessing is done.

```
   rt mention the us nuclear industry is taking 7 steps to
reconfirm safety &amp; emergency preparedness at nuclear plants.
learn more: {link}


Positive Score for each  tweet :0.1
Negative Score for each  tweet :0.1

Firing Strength of Negative (wneg): 0.2
Firing Strength of Neutral (wneu): 0.8
Firing Strength of Positive (wpos): 0.2

Resultant consequents MFs:
op_activation_low: [0.2 0.2 0.2 0.2 0.2 0.  0.  0.  0.  0.  0. ]
op_activation_med: [0.  0.2 0.4 0.6 0.8 0.8 0.8 0.6 0.4 0.2 0. ]
op_activation_high: [0.  0.  0.  0.  0.  0.  0.2 0.2 0.2 0.2 0.2]

Aggregated Output: [0.2 0.2 0.4 0.6 0.8 0.8 0.8 0.6 0.4 0.2 0.2]

Defuzzified Output: 5.0

Output after Defuzzification: Neutral

Doc sentiment: neutral
```

Fig. 6.6. *Processing of a sample tweet using VADER lexicon*



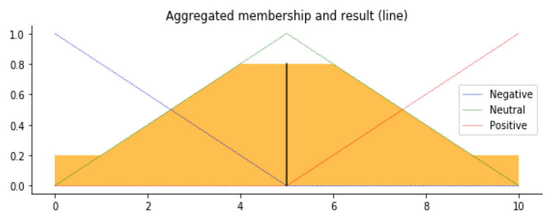


Fig. 6.7. *Output showing different emotions of the* tweet **Fig. 6.8.** *Aggregated Output of the tweet*

Then we apply VADER lexicons' *polarity_scores(a)* method which gives positive (*TweetPos*) score equal to 0.1 and negative (*TweetNeg*) score equal to 0.1 of the tweet as output. The fuzzy sets Low, Medium and High are created using triangular fuzzy membership for universe variables: positive (*x_p)* is (0-1), negative (x*_n)* is (0-1) and output (*x_op)* is (0-10). The fuzzy rules (6.15 - 6.23) are applied. The overall firing strength of tweet for different emotion classes are evaluated using (6.24 - 6.26). Fig. 6.7 is the visualization of membership values ($\mu_S(x)$) (firing strength) of different sentiment classes, blue color shows negative, green is for neutral and red is for positive class. The Resultant consequents MFs are computed using (6.27 – 6.29). Fig. 6.8 depicts the aggregated output membership ($\mu_A(x)$) computed in (6.30). The area under the aggregated output is used for centroid defuzzification in (6.31). The defuzzified output equal to 4.81 is shown as bold straight line. Finally, the sentiment of tweet is evaluated as ‘Neutral’ using (6.32). We can check the polarity of the tweet from dataset, and it turns out to be same.

6.3.2 COMPARISON AMONG LEXICONS

We can compare the performance of our fuzzy rule-based method in regard of the lexicon being used in the method. Fig. 6.9 shows a sample tweet (1008) of Sanders dataset. This tweet is being processed with different lexicons: SentiWordNet, AFINN and VADER. We can observe that all lexicons (Fig. 6.10 - 6.12) detect the correct sentiment (neutral) by the proposed scheme. Another sample tweet (3420) of Sanders dataset is depicted in Fig. 6.13. Here with the help of VADER lexicon (Fig. 6.16) correct sentiment class is detected (positive) while SentiWordNet (Fig. 6.14) and AFINN (Fig. 6.15) detect the wrong sentiment class (neutral). Further we have displayed the execution time of all methods with different lexicon-dataset combinations in Table 6.9 and Table 6.9. We can examine that VADER lexicon takes minimum time while SentiWordNet lexicon takes maximum time in executing the method. Our fuzzy rule-based method takes least time for execution, comparable with most of the methods in Table 6.9 and Table 6.9.

```
apple ios5 is all well and good and has nice new features, but i am still waiting
on an app that will go to work for me.
```

Fig. 6.9. *Sample tweet (1008) of Sanders Dataset*

```
Positive Score for each  tweet :
2.792

Negative Score for each  tweet :
0.375

Firing Strength of Negative (wneg): 0.0
Firing Strength of Neutral (wneu): 0.1875
Firing Strength of Positive (wpos): 0.736

Defuzzified Output: 6.62

Output after Defuzzification:
Neutral

Doc sentiment:
neutral
```

Fig. 6.10. *Tweet Processing by SentiWordNet*

```
Positive Score for each  tweet :
6.0

Negative Score for each  tweet :
0

Firing Strength of Negative (wneg): 0.0
Firing Strength of Neutral (wneu): 0.25
Firing Strength of Positive (wpos): 0.75

Defuzzified Output: 6.36

Output after Defuzzification:
Neutral

Doc sentiment:
neutral
```

Fig. 6.11. *Tweet Processing by AFINN*

```
Positive Score for each  tweet :
0.2

Negative Score for each  tweet :
0.0

Firing Strength of Negative (wneg): 0.0
Firing Strength of Neutral (wneu): 0.6
Firing Strength of Positive (wpos): 0.4

Defuzzified Output: 4.98

Output after Defuzzification:
Neutral

Doc sentiment:
neutral
```

Fig. 6.12. *Tweet Processing by VADER*

```
facebook, twitter , spongebob, nirvana. great way to
spend the night. :d
```

Fig. 6.13. *Sample tweet (3420) of Sanders Dataset*

```
Positive Score for each  tweet :
0.25

Negative Score for each  tweet :
0

Firing Strength of Negative (wneg): 0.0
Firing Strength of Neutral (wneu): 0.875
Firing Strength of Positive (wpos): 0.125

Defuzzified Output: 5.26

Output after Defuzzification:
Neutral

Doc sentiment:
positive
```

Fig. 6.14. *Tweet Processing by SentiWordNet*

```
Positive Score for each  tweet :
3.0

Negative Score for each  tweet :
0

Firing Strength of Negative (wneg): 0.0
Firing Strength of Neutral (wneu): 0.625
Firing Strength of Positive (wpos): 0.375

Defuzzified Output: 5.4

Output after Defuzzification:
Neutral

Doc sentiment:
positive
```

Fig. 6.15. *Tweet Processing by AFINN*

```
Positive Score for each  tweet :
0.5

Negative Score for each  tweet :
0.0

Firing Strength of Negative (wneg): 0.0
Firing Strength of Neutral (wneu): 0.0
Firing Strength of Positive (wpos): 1.0

Defuzzified Output: 7.67

Output after Defuzzification:
 Positive

Doc sentiment:
positive
```

Fig. 6.16. *Tweet Processing by VADER*

6.3.3 COMPARISON WITH STATE-OF-THE-ART

We have compared our proposed rule-based approach for Sentiment Analysis with four state-of-the-art methods for unsupervised sentiment classification: i) Cavalacanti.*et al.* ii) Ortega *et al.* iii) Hutto and Gilbert iv) Nielsen, 2011. The first two methods have used SentiWordNet (Baccianella *et al.*, 2010) lexicon, the third method has implemented simple Sentiment Analysis using VADER lexicon (Hutto and Gilbert, 2014) and the last method has used AFINN lexicon (Nielsen, 2011) to perform Sentiment Analysis. The implementation of fuzzy rule based-method using all the lexicons is performed in isolation with each other on nine publicly available twitter datasets. The F1-scores (Micro and Macro) of all the methods for different lexicon-dataset combinations has been presented in Table 6.4 and Table 6.5.

The highest F1-score for each dataset has been shown in bold, for both Micro and Macro. The proposed method yields consistently high scores for F1-Micro and acceptable results for F1-Macro for all datasets. We can observe that our unsupervised fuzzy rule-based method with VADER lexicon has performed the best among all methods with the highest F1-Micro score of 0.865 in Gilbert Tweets and 0.842 in Nuclear Twitter dataset. Our method with VADER lexicon has the highest F1-Micro scores among Sanders, Nuclear, Apple and Gilbert Twitter datasets. On the other hand, our fuzzy rule method with AFINN Lexicon (Nielsen, 2011) achieves highest F1-Mirco scores of 0.765 and 0.686 in two-class dataset-Sentiment 140 and SemEval 2017 respectively. For STS-test dataset, the Nielson *et al.* 's method performed the best whereas in SemEval 2016 and 2015 Hutto and Gilbert 's method has highest F1-scores.

VADER lexicon has performed the best because this lexicon is best suited for social media posts. It handles emojis, slang, emoticons, and acronyms very well and evaluates the emoticons contained in the text. Tremendous benefits can be obtained by using VADER in micro-blogging websites wherein the text data is of complex nature. SentiWordNet lexicon was developed in 2010, AFINN lexicon in 2011, and VADER lexicon was developed recently in 2014. Since our approach is unsupervised and we don't have any training data, the VADER lexicon is best suited for the task. The VADER lexicon doesn't require any training data. The next best lexicon is AFINN and then SentiWordNet lexicon.

Table 6.4 F1- Scores of Different methods, Lexicons and Twitter datasets.

Lexicons	Methods	Sanders		Nuclear		Apple		STS Test		Sentiment 140	
		F1-Micro	**F1-Macro**	**F1-Micro**	**F1-Macro**	**F1-Micro**	**F1-Macro**	**F1-Micro**	**F1-Macro**	**F1-Micro**	**F1-Macro**
SentiWordNet	Cavalcanti	0.255	0.266	0.074	0.110	0.307	0.287	0.502	0.423	0.600	0.38
	Ortega	0.568	0.424	0.196	0.184	0.524	0.43	0.456	0.448	0.339	0.265
	Fuzzy Rules	0.679	0.306	0.816	**0.384**	0.57	0.33	0.46	0.41	**0.763**	0.304
AFINN	Simple SA	0.558	**0.515**	0.484	0.338	0.557	**0.524**	**0.729**	0.**726**	0.527	0.348
	Fuzzy Rules	0.678	0.387	0.768	0.352	0.6	0.503	0.482	0.427	**0.765**	0.316
VADER	Simple SA	0.541	0.509	0.295	0.244	0.546	0.517	0.717	0.714	0.534	0.**528**
	Fuzzy Rules	**0.686**	0.425	**0.842**	0.338	**0.614**	0.416	0.642	0.642	0.528	0.333

Table 6.5 F1- Scores of Different methods, Lexicons and SemEval-Gilbert datasets.

Lexicons	Methods	SemEval 2017		SemEval 2016		SemEval 2015		Gilbert Tweets	
		F1-Micro	**F1-Macro**	**F1-Micro**	**F1-Macro**	**F1-Micro**	**F1-Macro**	**F1-Micro**	**F1-Macro**
SentiWordNet	Cavalcanti	0.358	0.334	0.436	0.314	0.372	0.309	0.549	0.406
	Ortega	0.473	0.419	0.255	0.253	0.467	0.428	0.363	0.332
	Fuzzy Rules	0.485	0.231	0.326	0.227	0.478	0.221	0.346	0.223
AFINN	Simple SA	0.558	0.515	0.308	0.185	0.618	0.594	0.079	0.073
	Fuzzy Rules	**0.686**	0.308	**0.457**	**0.419**	0.484	0.236	0.44	0.426
VADER	Simple SA	0.528	0.526	0.475	0.428	**0.604**	0.585	1	1
	Fuzzy Rules	0.525	0.381	0.34	0.232	0.524	0.319	**0.865**	**0.772**

Table 6.6 Precision Recall Macro Scores of Different methods, Lexicons and Twitter datasets.

Lexicons	Methods	Sanders		Nuclear		Apple		STS Test		Sentiment 140	
		Precision	**Recall**	**Precision**	**Recall**	**Precision**	**Recall**	**Precision**	**Recall**	**Precision**	**Recall**
SentiWordNet	Cavalcanti	0.410	0.430	0.079	0.372	0.409	0.457	0.524	0.469	0.4	0.425
	Ortega	0.419	0.424	0.374	0.373	0.439	0.436	0.506	0.480	0.302	0.297
	Fuzzy Rules	0.536	0.348	0.359	0.416	0.434	0.352	0.524	0.466	0.41	0.343
AFINN	Simple SA	0.527	0.581	0.389	0.455	0.527	0.589	**0.735**	**0.728**	0.417	0.437
	Fuzzy Rules	**0.597**	0.364	0.331	0.396	0.586	0.49	0.626	0.509	0.446	0.348
VADER	Simple SA	0.528	**0.589**	0.376	0.443	0.531	**0.594**	0.726	0.715	**0.628**	**0.661**
	Fuzzy Rules	0.583	0.421	**0.616**	0.347	**0.69**	0.426	0.705	0.655	0.365	0.375

Table 6.7 Precision Recall Macro Scores of Different methods, Lexicons and SemEval-Gilbert datasets.

Lexicons	Methods	SemEval 2017		SemEval 2016		SemEval 2015		Gilbert Tweets	
		Precision	**Recall**	**Precision**	**Recall**	**Precision**	**Recall**	**Precision**	**Recall**
SentiWordNet	Cavalcanti	0.458	0.454	0.345	0.392	0.457	0.458	0.467	0.447
	Ortega	0.43	0.418	0.303	0.296	0.444	0.446	0.436	0.493
	Fuzzy Rules	0.527	0.337	0.427	0.352	0.539	0.335	0.416	0.358
AFINN	Simple SA	0.526	**0.581**	**0.559**	0.346	0.595	**0.614**	0.672	0.352
	Fuzzy Rules	**0.637**	0.351	0.423	**0.438**	**0.799**	0.343	0.697	0.590
VADER	Simple SA	0.539	0.561	0.426	0.448	0.598	0.613	1	1
	Fuzzy Rules	0.617	0.414	0.493	0.362	0.691	0.383	**0.748**	**0.881**

Table 6.6 and Table 6.7 presents the Precision and Recall Macro scores of all methods for different datasets-lexicon combinations. The highest precision and recall scores are shown in bold. Our fuzzy rules method with VADER lexicon has the highest precision in Nuclear, Apple dataset and best precision- recall in Gilbert Tweets. On the other hand, our fuzzy rule method with AFINN lexicon achieves the highest precision in Sanders, SemEval 2017, and SemEval 2015 datasets; and the highest recall in SemEval 2016. Nielson *et al.* 's method performed the best in the STS test dataset in both recall and precision; highest recall in Nuclear, SemEval 2017, and SemEval 2015 datasets; highest precision in SemEval 2016. In the sentiment 140 dataset, the highest scores were scored by Hutto and Gilbert's method. This method gained the highest recall in Sanders and Apple datasets as well. We can conclude that AFINN and VADER lexicon performed better compared to the SentiWordNet lexicon.

Table 6.8 Execution Time (in sec) of Different methods, Lexicons and Twitter datasets.

Lexicons	Methods	Sanders	Nuclear	Apple	STS Test	Sentiment 140
SentiWordNet	Cavalcanti	7.86	0.55	19.79	1.077	2042.85
	Ortega	222.87	2.19	393.5	6.32	15000
	Fuzzy Rules	15.81	1.02	18.46	2.87	8924.22
AFINN	Simple SA	4.91	0.51	5.106	0.61	2511.15
	Fuzzy Rules	9.95	0.68	11.42	1.32	2834.90
VADER	Simple SA	1.75	0.27	2.36	0.328	936.93
	Fuzzy Rules	7.65	0.46	5.6	1.18	2111.85

Table 6.9 Execution Time (in sec) of Different methods, Lexicons & SemEval-Gilbert datasets.

Lexicons	Methods	SemEval 2017	SemEval 2016	SemEval 2015	Gilbert Tweets
SentiWordNet	Cavalcanti	31.87	23.22	44.89	10.04
	Ortega	4599	1584.10	2781	277.5
	Fuzzy Rules	54.9	40.75	57.61	18.43

Lexicons	Methods	SemEval 2017	SemEval 2016	SemEval 2015	Gilbert Tweets
AFINN	Simple SA	4.80	27.8	31.96	14.3
	Fuzzy Rules	11.7	12.8	40.2	9.62
VADER	Simple SA	7.76	5.64	7.45	2.24
	Fuzzy Rules	25.65	26.96	27.62	6.89

Execution time of each method for all datasets-lexicon combinations are presented in Table 6.8 and Table 6.9. Execution time depends upon various factors: the size of the dataset, lexicon and type of calculations in a method. Small size datasets take very less time compared to bigger size datasets, for example, STS test and Nuclear datasets take less than 1 sec while Sentiment140 takes hours to execute. Among the lexicons, VADER is the fastest and SentiWordNet is the slowest lexicon. Ortega *et al.* method takes maximum time compared to other methods. The fuzzy rule method with AFINN or VADER lexicon performs faster compared to methods that implement the SentiWordNet lexicon. The fuzzy rule method is unsupervised and requires no training time and is not dependent on the size of the dataset. This is the advantage of our approach. VADER is quick and computationally economical without comprising F1 scores. It works excellently well on social media text. It doesn't require any training data.

6.4 FUZZY LOGIC BASED DYNAMIC PLOTTING OF MOOD SWINGS FROM TWEETS

Twitter is one the most of popular social media platforms. Users express their feelings easily on social media regarding any trending event. We have proposed a fuzzy logic-based approach for dynamic plotting of mood swings from tweets. The novelty of this approach is the use of linguistic hedges with fuzzy logic to compute the sentiment of a tweet. A comparison of our approach with existing methods, on real-time tweets extracted from an online website, confirms the superiority and efficiency of our method. The tweets used in our experiments are extracted from the timeline of the India Vs Pakistan final ICC world-cup match in June 2017. They reflect the moods of the Twitter users as the match progresses. Using our fuzzy logic-based approach we successfully plot the dynamic mood vs time and compute the polarity of the sentiment at each time instant.

6.4.1 PROPOSED APPROACH

Our proposed approach for dynamic analysis of the sentiment of tweets is based on fuzzy logic. It uses SentiWordNet (SWN), a lexicon resource for sentiment analysis. Following is the description of the approach followed:

6.4.1 Tokenization and Lemmatization of Tweets

Tweets are stored in a document/paragraph. These tweets are stored with their corresponding timestamp. We need to first extract tweets from the document. Splitting up of paragraphs into sentences is termed as sentence tokenization (Bird *et al.*, 2009). Tokenizing a sentence is a process of splitting a sentence into a list of words. In other words, a tokenizer parses a sentence into a list of tokens (words). The output of sentence tokenization process will be stored in a dynamic list. The tokenized words are lemmatized to get the root word.

6.4.2 Bag of Words

In sentiment analysis Bag of words mean those important words which are essential for mining reviews, opinions, etc. It is a model that transforms documents into numerical form, where each word in the document is assigned some value in the range of 0 to 1. In our approach, the Universe of discourse, for fuzzy set, is the set of all the words in each tweet. Each tweet contains a bag of word list. Each word has a positive and negative score (6.35 and 6.36) obtained from SWN. This bag of words can be interpreted as a fuzzy membership pertaining to the fuzzy sets Pos and Neg. In our proposed model, we have used NLTK POS Tagger (Bird *et al.*, 2009) to extract words which are nouns, adjectives, verbs or adverbs. The fuzzy sets *Pos* and *Neg* are represented as:

$$Pos = \{(a, \mu_{Pos}(a)\}, a \in X_i \tag{6.33}$$

$$Neg = \{(a, \mu_{Neg}(a)\}, a \in X_i \tag{6.34}$$

where a is the word and it belongs to the i^{th} Bag of words X_i, where i=1,2,…,n. If we have n tweets then there is n bag of words for each tweet. Each tweet has its own unique bag of words. The membership functions $\mu_{Pos}(a)$ and $\mu_{Neg}(a)$ associated with the tokenized word a are defined below.

$$\mu_{Pos}(a) = \frac{\sum_{synsets} [syn.pos\_score()]}{length(synsets)} \tag{6.35}$$

$$\mu_{Neg}(a) = \frac{\sum_{synsets} [syn.neg\_score()]}{length(synsets)} \tag{6.36}$$

where syn.pos_score () and syn.neg_score () are the scores obtained from SWN; synsets is the set of synonyms of each word present in SWN.

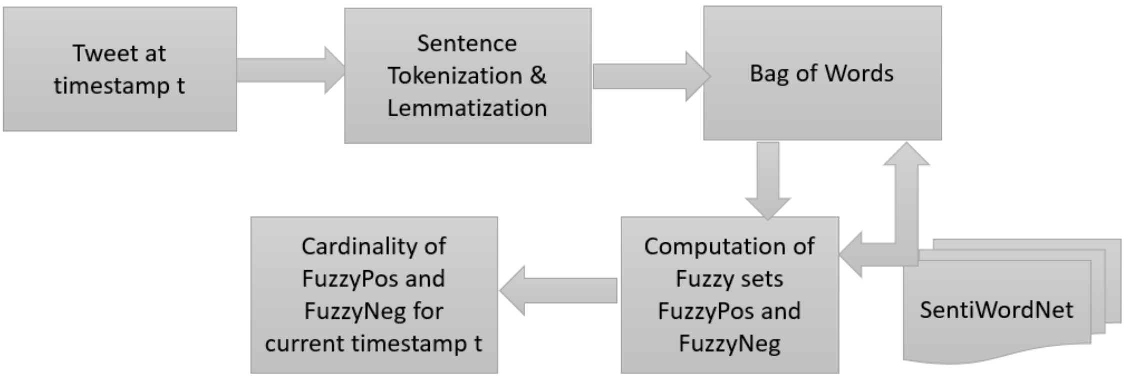


Fig. 6.17. *Fuzzy Processing of a single tweet by our proposed Fuzzy Approach*

6.4.3 Formulation of Fuzzy Sets for plotting mood swings

First of all, Union of fuzzy sets is carried out to check the presence of emotion flowing through the tweets. Union fuzzy set is Pos $\cup$ Neg and the membership function is:

$$\mu_{Union}(a) = \mu_{Pos}(a) \vee \mu_{Neg}(a), a \in X_i \tag{6.37}$$

Linguistic hedges are modifiers, they are adverb or adjective like very, slightly, more or less, fairly, etc. The membership function for a basic atomic term can be modified using the concept of linguistic hedges (Ross, 2004). In our approach we have used the linguistic hedges *Very*, *More or less* and *Not* in (6.38 and 6.39).

$$\mu_{Verypos}(a) = [\mu_{Pos}(a)]^2,\ \mu_{NotVerypos}(a) = 1 - [\mu_{Pos}(a)]^2,\ \mu_{MoreLesspos}(a) = [\mu_{Pos}(a)]^{0.5} \tag{6.38}$$

$$\mu_{MoreLessneg}(a) = [\mu_{Neg}(a)]^{0.5},\ \mu_{Veryneg}(a) = [\mu_{Neg}(a)]^2,\ \mu_{NotVeryneg}(a) = 1 - [\mu_{Neg}(a)]^2 \tag{6.39}$$

We have proposed *FuzzyPos* and *FuzzyNeg* sets for all tweets:

$$FuzzyPos = \{(a, \mu_{FuzzyPos}(a))\}, a \in X_i \tag{6.40}$$

$$FuzzyNeg = \{(a, \mu_{FuzzyNeg}(a))\}, a \in X_i \tag{6.41}$$

where X_i is the *i*th BOW and a is the word in BOW. The membership functions $\mu_{FuzzyPos}$ and $\mu_{FuzzyNeg}$ are defined as:

$$\mu_{FuzzyPos}(a) = \mu_{Morelesspos}(a) \wedge \mu_{NotVeryneg}(a) \tag{6.42}$$

$$\mu_{FuzzyNeg}(a) = \mu_{Morelessneg}(a) \wedge \mu_{NotVerypos}(a) \tag{6.43}$$

Now let W be the set of words $\{w\}$ in tweet at timestamp t, m is the number of words in set W and p = 1,2…*total*, where *total* is the total number of tweets. Next, we calculate cardinality of *FuzzyPos* and *FuzzyNeg* for each tweet p, as shown below:

$$Card\_FuzzyPos(p) = \sum_{w=1}^{m} \mu_{FuzzyPos}(w), w \in W \tag{6.44}$$

$$Card\_FuzzyNeg(p) = \sum_{w=1}^{m} \mu_{FuzzyNeg}(w), w \in W \tag{6.45}$$

The fuzzy processing for single tweet using our proposed fuzzy approach is shown in Fig. 6.17. Next for dynamic plotting of the mood swings, we fix time windows containing 3 tweets each. We sum up the cardinality for all 3 tweets in each time window as shown below:

$$FuzzyPos\_value(n) = \sum_{p=3n+1}^{3n+3} Card\_FuzzyPos(p) \tag{6.46}$$

$$FuzzyNeg\_value(n) = \sum_{p=3n+1}^{3n+3} Card\_FuzzyNeg(p) \tag{6.47}$$

where n is the number of time windows n=0,1…(*total*/3)-1

Finally, *Sentiment* value is calculated for each time window by de-fuzzifying of *FuzzyPos* and *FuzzyNeg* values to Positive (*P*) and Negative(*N*) respectively is done to get crisp result, i.e. the sentiment of tweets.

$$Sentiment(n) = \begin{cases} P, FuzzyPos\_value > FuzzyNeg\_value \\ Neu, FuzzyPos\_value = FuzzyNeg\_value \\ N, otherwise \end{cases} \quad (6.48)$$

where *Sentiment* is the sentiment at particular time window *n*.

Cavalcanti *et.al*, 2011 proposed an approach to compute the sentiment score of each word using positive and negative score in (6.35) and (6.36) as shown below:

$$score(w) = \frac{\sum_{w \in W} [\mu_{Pos}(w) - \mu_{Neg}(w)]}{m}, w \in W \quad (6.49)$$

The overall sentiment of each tweet is calculated by summing up the score(w) for each tweet. If the value of overall sentiment is higher than zero then tweet is positive, if the value is less than zero then tweet is negative and if it is equal to zero then tweet is neutral. Ortega *et.al*, 2013 developed an unsupervised SA approach on Twitter and SMS messages based on word sense disambiguation (WSD). The polarity of each word is calculated using positive score: *syn.pos_score ()* and negative score: *syn.neg_score ()*, obtained from SWN.

$$polarity(w) = \begin{cases} 4, syn.pos\_score(w) \geq 0.75 \\ -4, syn.neg\_score(w) \geq 0.75 \\ 2, syn.pos\_score(w) < 0.75 and\ syn.pos\_score(w) \geq 0.5 \\ -2, syn.neg\_score(w) < 0.75 and\ syn.neg\_score(w) \geq 0.5 \\ 0, otherwise \end{cases} \quad (6.50)$$

The overall positive value (*PosS*(*t*)) and overall negative value (*NegS*(*t*)) is computed by summing the *syn.pos_score ()* and *syn.neg_score ()* respectively for each tweet t. If *PosS*(*t*) is greater than *NegS*(*t*) then the tweet is considered as positive. On the other hand, if *PosS*(*t*) is less than *NegS*(*t*) the tweet is negative. At last, if *PosS*(*t*) is equal to *NegS*(*t*) the tweet is assigned neutral label. For dynamic plotting of both methods, we add up scores of all words for every 3 tweets in each time window in the same procedure as mentioned above.

6.4.4 Overall Methodology

Following are the steps of overall methodology applied:

1) Procure single tweet with timestamp *t*.
2) Do the fuzzy processing of tweet as per Fig. 6.17.
3) Repeat 1) and 2) for two consecutive tweets.
4) Sum up the respective cardinality for all 3 tweets as per (6.46) and (6.47).
5) Decode the final sentiment for current time window using (6.48).
6) Repeat steps1-5 for all consecutive time windows.
7) Plot the *FuzzyPos_value* and *FuzzyNeg_value* with time window to show the dynamic mood swings.

6.4.2 EXPERIMENTAL SETUP AND IMPLEMENTATION

Our proposed approach is based on social media: Twitter. We have collected 30 plaintext tweets from online websites[3]. These tweets have been posted by cricket fans who were watching India Vs Pakistan ICC Champions Trophy Final World Cup match on 18th June,2017. The identity of fans is kept anonymous. They are expressing their feelings, mood. These 30 tweets are used as a test case while experimenting the proposed fuzzy approach. We have used these tweets as input to fuzzy processing method. We have compared our method with two methods, first given by Cavalcanti *et.al*, 2011 and second given by Ortega *et.al*, 2013. All methods are unsupervised techniques and their dynamic graphs can be plotted in real-time. Same set of tweets are processed by these methods. The code containing the implementation of our work is available online[4].

6.4.3 RESULTS AND DISCUSSION

In this section we discuss the results of our proposed fuzzy-logic based approach. We implemented the methodology, given in section 3, for processing tweets posted by cricket fans.

---

[3]: https://www.scoopwhoop.com/48-Tweets-Which-Define-How-Embarrassing- Indias-Champions-Trophy-Defeat-To-Pakistan-Was/#.l1uoc3uu5
[4]: https://github.com/SrishtiVashishtha/Fuzzy-logic-based-Dynamic-Plotting-of-Mood-Swings-from-Tweets

Following are the results:

6.4.3.1 Union

The first operation we have applied is the union of positive and negative scores of each tweet. It shows that some kind of emotion (either positive or negative) and sentiment is running. Overall reading of graph is high. Fig. 6.18(a) shows the union values for all 30 tweets and Fig. 6.18(b) shows for 10-time windows.

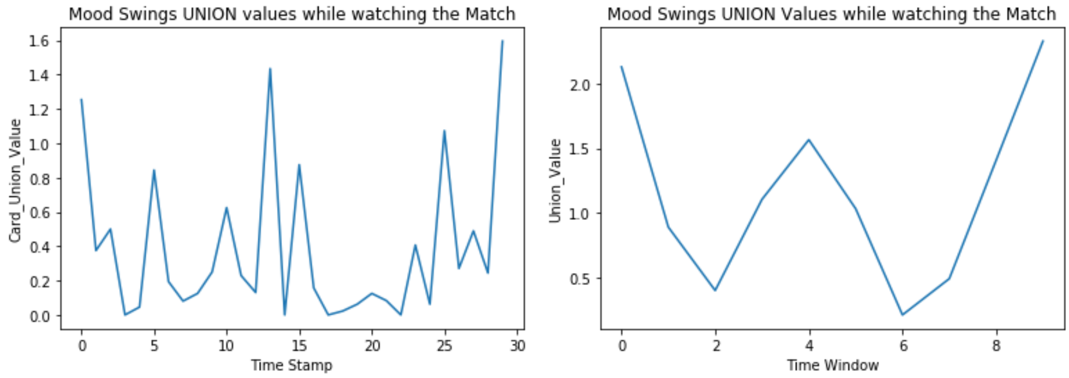


Fig. 6.18 *Union Values for a) 30 tweets b) 10-time windows*

6.4.3.2 FuzzyPos

The next operation is our proposed *FuzzyPos* set (6.40) which states that there is more or less positivity but not very negativity in tweets. Fig. 6.19(a) shows the values for all 30 tweets and Fig. 6.19(b) shows for 10-time windows. We can observe initially the tweets are positive about the match but as match proceeds there is no positivity in tweet 10,11,12 hence the values are 0 and we can see dip, later on due to the fall of wickets of Indian team the fans get angry and post negative tweets like tweet no 5: “SharmaJi ka beta fail ho Gaya!! #INDvPak#CT17Final”, tweet 23: “Hurts #IndvsPak #PakistanMurdabad”, tweet 25: “TV sets broken on the streets of Ahmedabad after India's defeat by Pakistan in the #CT2017Final #IndVsPak”. But at the end some kind of positivity came up to support Indian team like tweet 30: “Do not become sad ur team (india) is one of the best team i think they loss becoz of overconfidence.”

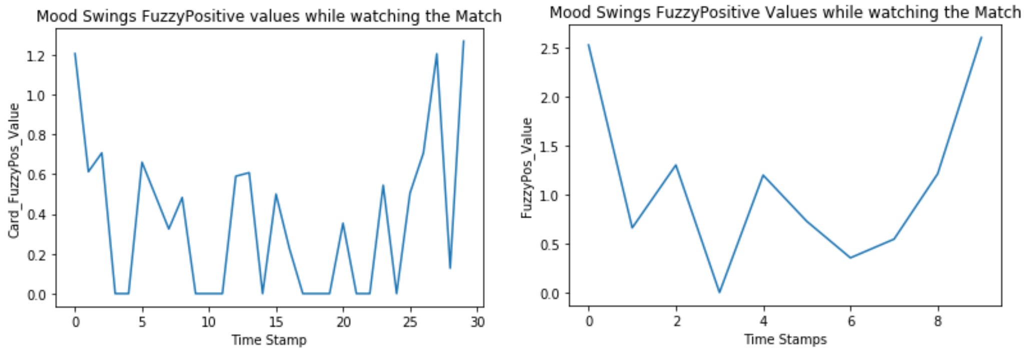


Fig. 6.19 *Fuzzy Positive Values for a) 30 tweets b) 10-time windows*

6.4.3.3 FuzzyNeg

Another operation proposed is *FuzzyNeg* set (6.41) which states that there is more or less negativity but not very positivity in tweets. The graph is shown in Fig. 6.20. Fig. 6.20(a) shows the values for all 30 tweets and Fig. 6.20(b) shows for 10-time windows. We can observe initially the tweets are less negative about the match but as match proceeds India starts losing the match the negativity starts rising and reaches its peak, this implies the fans are sad and angry like in tweet 17: "We lost the match bulleya, They played so well bulleya, We need a slap bulleya, Cup unka, cup unka #INDvPAK". The peak is achieved due to words "massacre" in tweet 10, "not" in tweet 11, "down" in tweet 12 and highest at tweet 14 due to the double occurrence of "not". Some fans are still in hope of winning the match and supporting the team hence the negativity falls in between but finally when India loses the match negativity rises again.

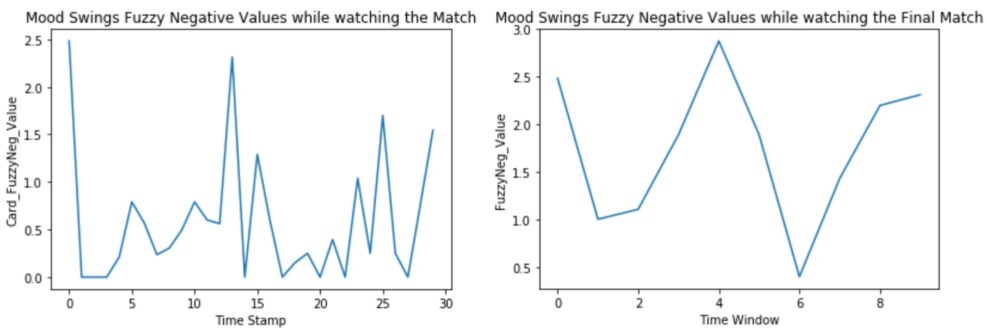


Fig. 6.20 *Fuzzy Negative Values for a) 30 tweets b) 10-time windows*

6.4.3.4 Difference by Cavalcanti et.al

We have plotted graph for Cavalcanti et.al method (6.49) for the same set of tweets. The graph is shown in Fig. 6.21. Fig. 6.21(a) shows the values for all 30 tweets and Fig. 6.21(b) shows for 10-time windows.

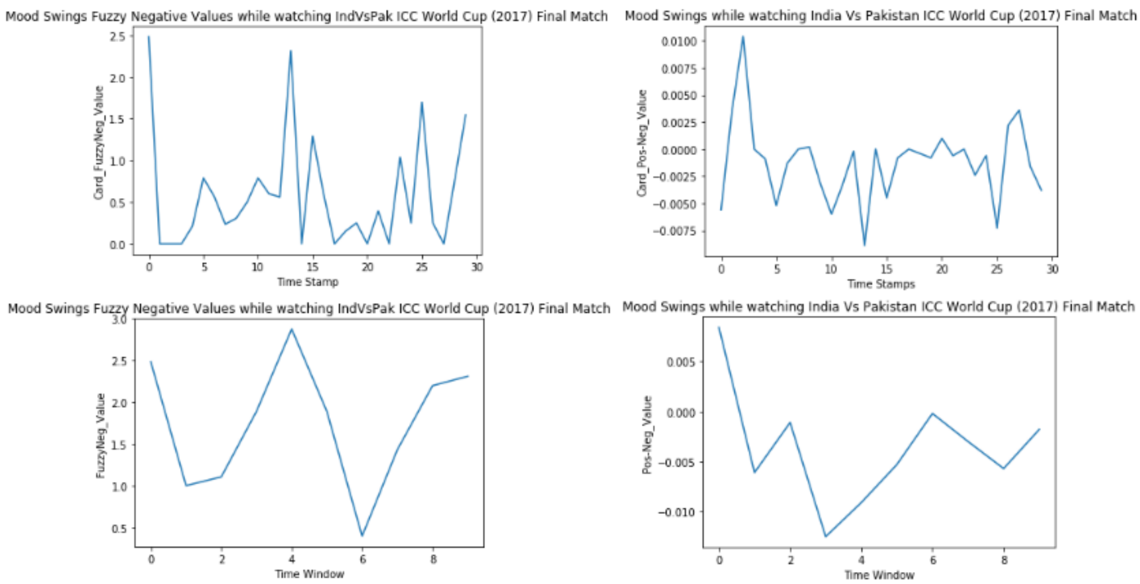


Fig. 6.21 *Difference Values for a) 30 tweets b) 10-time windows*

We can observe that this method is not able to catch optimism in the end. As we can see initially the first value is positive in Fig. 6.20(b) but later on all values are negative. It is not able to show positivity of tweet 30: “Do not become sad ur team (india) is one of the best team i think they loss becoz of overconfidence.”. We have calculated the sentiment of Cavalcanti *et.al* method, the positive values denote the positive sentiment and negative values denote negative sentiment as shown in Table 6.10. The sentiment of tweets at each and particular time window, using our method, is also displayed in Table 6.10.

Table 6.10. Sentiment of Tweets.

Time Window	*FuzzyPos*	*FuzzyNeg*	*Fuzzy Sentiment*		*Cavalcanti et.al Sentiment*		*Ortega et.al Sentiment*
TW1	**2.523**	2.4826	2.523	**P**	0.0084	**P**	**P**
TW2	0.6599	**1.0039**	1.0039	**N**	-0.0061	**N**	**N**
TW3	**1.3005**	1.1038	1.3005	**P**	-0.0011	**N**	**N**
TW4	0	**1.8912**	1.8912	**N**	-0.0125	**N**	**N**
TW5	1.1975	**2.8758**	2.8758	**N**	-0.0091	**N**	**N**
TW6	0.7265	**1.8899**	1.8899	**N**	-0.0053	**N**	**Neu**
TW7	0.3536	**0.4007**	0.4007	**N**	-0.0002	**N**	**Neu**
TW8	0.5445	**1.4324**	1.4324	**N**	-0.003	**N**	**Neu**
TW9	1.2115	**2.1987**	2.1987	**N**	-0.0057	**N**	**P**
TW10	**2.6014**	2.3113	2.6014	**P**	-0.0018	**N**	**P**

Table 6.10 is equivalent to the time graph for dynamic plotting of mood swings. It shows the computed sentiment values for both methods. We can see initially the tweets are positive implying fans are positive for the match, as the match progresses India starts losing the match people become sad, and angry and thus the tweets are negative. At the last time window sentiment becomes positive because people are supporting team India and portraying their love for team India. Our fuzzy logic-based approach catches the optimism at the end and computes the sentiment at each time window correctly. Comparison with Cavalcanti *et.al* method's sentiment values shows that it is not able to detect positive sentiment at the 3rd and the 10th time window and Ortega *et.al* method's sentiment values show that it is not able to detect negative emotion from the 6th-9th time window. Our approach is sensitive to mood swings. Analysis and interpretation of our proposed fuzzy logic-based method demonstrate that processing of plaintext tweets for decoding the mood swings of cricket fans while watching the India Vs Pakistan ICC Champions final world cup match is superior than the Cavalcanti *et.al* method.

6.5 CONCLUSION

In this chapter, we have proposed a fuzzy rule-based methodology for Sentiment Analysis of social media posts specifically for twitter datasets. The novelty of this chapter is i) the formulation of nine fuzzy rules to evaluate the sentiment class of tweets, ii) proposed unsupervised approach is suitable to any lexicon and iii) and any two- class or three-class sentiment dataset. Two-class datasets have positive and negative sentiment classes while three-class datasets have an additional neutral sentiment class. We learn that fuzzy rules are able to incorporate the fuzziness of positive and negative scores. Fuzzy logic-based systems can deal with vagueness and ambiguity. Advantages of using the fuzzy approach are summarized as i) An important contribution of fuzzy logic is that it provides a way for computing with words, i.e. words can be transformed into numerical values for further computation, ii) Fuzzy logic provides us a desirable way to deal with linguistic problems and iii) it is equipped with logical reasoning that aids in SA process by providing closer views to the exact sentiment values.

We have implemented our proposed method using three various lexicons: SentiWordNet, AFINN, and VADER in isolation from each other on nine publicly available Twitter datasets. Comparison with four state-of-the-art methods for unsupervised sentiment classification reveals that our fuzzy rule-based method performs consistently the best with respect to F1-Micro scores. Our fuzzy rule based method scores higher F1 Micro scores, Precision, and Recall in the majority of datasets (7 out of 9). The F1- Macro scores are acceptable in all cases if not always the best. The highest F1-Micro scores of 0.865 and 0.842 are achieved by the VADER lexicon in Gilbert Tweets and Nu-clear Twitter datasets respectively. Moreover, the methods which implement VADER lexicon execute in the least time while the methods which implement the SentiWordNet lexicon take maximum time in execution. The metrics precision and recall scores for unsupervised methods AFINN and VADER lexicon performed better compared to SentiWordNet lexicon. VADER is quick and computationally economical without comprising F1 scores. It works excellently well on social media text. It doesn’t require any training data. It has performed the best because this lexicon is best suited for social media posts. It handles emojis, slang, emoticons, and acronyms very well and evaluates the emoticons contained in the text. Tremendous benefits can be obtained by using VADER in micro-blogging websites wherein the text data is of complex nature.

We have also proposed a fuzzy logic- based approach for dynamic plotting of mood swings of tweets. These tweets are posted by cricket fans while watching India Vs Pakistan's final ICC world-cup match in June 2017. The novelty of our approach is the use of linguistic hedges: very, more or less and not, with fuzzy logic. These hedges describe the mood of the user in the tweet. The fuzzy membership values of these hedges are used to evaluate the fuzzy positive and fuzzy negative sentiment of tweets. Finally, we apply defuzzification to get the sentiment of each tweet. We have compared our method with existing methods, the analysis demonstrates that our approach is more sensitive to mood swings and decodes the correct sentiment at each time window.

DESIGN OF MULTI-MODAL SPEECH EMOTION RECOGINITION BY FUZZY INFERENCING

Sentiment Analysis mainly focuses on the classification of data into three sentiment classes: positive, negative, and neutral. These sentiment features can be applied to extract emotion classes of the same data. Previous chapters concentrate on text-based sentiment analysis models with fuzzy inferencing but in this chapter audio and text features, are analyzed for emotion recognition. The textual sentiment features, along with speech cues, aid in recognizing the emotions. A novel text and speech based fuzzy rule-based system has been proposed for multimodal sentiment analysis of reviews posted on social media.

In last half decade an increasing number of works published has manifested the tremendous progress in multimodal sentiment analysis. In real-life communication, people are spontaneously modulating their tone to accentuate specific points or to express their sentiments. In this chapter[1] we have proposed two fuzzy rule-based algorithms for speech emotion recognition. The first research work introduces a supervised fuzzy rule-based system for multimodal sentiment classification that can identify the sentiment expressed in reviews on social media platform. It has been demonstrated that multimodal sentiment analysis can be effectively performed by the joint use of linguistic and acoustic modalities. In this chapter computation of the sentiment using an ingenious set of fuzzy rules has been deployed to label

[1:] The contents of this chapter are published in "Inferring Sentiment from Supervised Classification of Text and Speech cues using Fuzzy Rules." *Procedia Computer Science* 167 (2020): 1370-1379 and "Unsupervised Fuzzy Inference System for Speech Emotion Recognition using audio and text cues (Workshop Paper)." *In 2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM), pp. 394-403. IEEE, 2020*.

the review into positive or negative sentiment. The confidence score from supervised Support Vector Machine (SVM) classification of text and speech cues is considered as input variable for the fuzzy rules. The fusion of fuzzy logic with acoustic and linguistic features for classifying sentiment contributes to a new exemplar in multimodal sentiment analysis. Our fuzzy approach has been compared with eight state-of-the-art techniques for supervised machine learning. The experiments on benchmark datasets yield 82.5 % accuracy for our approach which is higher in contrast to the state-of-the-art.

With the growth of communication technology, the high availability of computers and smartphones, and the rapid rise of social media have greatly boosted content sharing in form of reviews or comments. There is a shift toward a multimodal social web, where users post their opinion in the form of audio clips, or in the form of text on Twitter and Instagram (Soleymani *et al.,* 2017). Multimodal sentiment analysis is a new dimension that combines different modalities such as text, audio, and video to integrate the features for better evaluation of opinions (Morency *et al.*, 2011). These vocal modulations in the visual data, together with text data, provide significant hints to determine the correct emotional states of the opinion holder. Thus, a combo of text and audio data helps to build an improved sentiment analysis model (Soleymani *et al.,* 2017). We have attempted to build a fuzzy rule-based sentiment analysis system for multimodal text and speech, using a decision-level fusion.

The key highlights of this work are: i) the creation of four novel fuzzy rules based on text and speech cues to evaluate the sentiment of each review ii) the proposed decision-level fusion approach performs better than unimodal and basic feature-level text-speech fusion using supervised machine learning classifier: Support Vector Machines (SVM) and iii) comparison of our proposed rule-based system for sentiment analysis with eight state-of-the-art techniques for supervised machine learning. Section 7.1 covers the related studies on multimodal sentiment analysis based on online social platforms. Section 7.2 describes our proposed fuzzy rule-based system. The experimental setting & implementation of our work is given in Section 7.3. Results are presented in Section 7.4. The second algorithm- Unsupervised Fuzzy Inference System for Speech Emotion Recognition using audio and text cues is described in Section 7.5. Section 7.6 offers a summary of our work.

7.1 PROPOSED FUZZY RULE-BASED SYSTEM FOR FIRST ALGORITHM

A supervised fuzzy rule-based system for multimodal sentiment classification has been introduced in this chapter, which can identify the sentiment expressed in reviews on social media platform. The experiments demonstrate that multimodal sentiment analysis can be effectively performed by the joint use of linguistic and acoustic modalities. Our approach consists of three major phases: Unimodal Feature Extraction, Confidence Score Computation from supervised classification and Fuzzy-Rule System incorporating decision-level fusion.

7.1.1 UNIMODAL FEATURE EXTRACTION

The first task of our methodology is to extract the unimodal text and speech features of product reviews. To take into account linguistic information, Term Frequency–Inverse Document Frequency (TF-IDF) (Joachims, 1996) cues are considered the measurable features from raw texts. TF-IDF is a mathematical statistic that is used to indicate the significance of a word in a document. TF means term frequency while IDF means inverse document frequency. The TF-IDF value is directly proportional to the number of times a word appears in the document and is counteracted by the frequency of the word in the set of documents.

$$TF-IDF(t,r) = TF(t,r) * IDF(t) \tag{7.1}$$

$$IDF(t) = \log\frac{1+n_r}{1+df(r,t)} + 1 \tag{7.2}$$

where TF (t, r) shows the number of times the word t appears in the review r, n_r is the total count of reviews and df (r, t) is the count of reviews that contain the word t. A review is considered as equivalent to a document. A total of 3079 text features are extracted.

For speech feature extraction, openSMILE is used to extract the pitch and voice intensity automatically. Using openSMILE, several Low-Level Descriptors (LLD) (e.g., voice intensity, pitch) and their statistical functionals are extracted (e.g., amplitude mean, arithmetic mean, standard deviation, flatness, skewness, kurtosis, and quartiles). Our work utilizes the configuration file of openSMILE (IS13-ComParE). A total of 6373 features are extracted from each input audio segment. The text and speech features are in normalized form. This is followed by Support Vector Machines (SVM)(Vapnik, 1995) classification of the text and speech cues, in isolation of each other.

7.1.2 CONFIDENCE SCORE COMPUTATION FROM SUPERVISED CLASSIFICATION

The SVM classifier is used to evaluate text and speech cues, its class membership probability estimate gives confidence scores for each sample (Wu *et al.*, 2004; Platt, 1999). These scores are titled confidence scores because these scores represent to what level the predicted value is correct and matches to the true value. The confidence scores are computed for text and speech features in isolation from each other. The normalization process is applied to these scores as shown in (7.3) so that the values are comparable.

$$norm(c) = \frac{c - \min(data)}{\max(data) - \min(data)} \tag{7.3}$$

where *norm(c)* represents the normalized value of a score *c*, *data* is a list of scores, *min* and *max* function calculates the minimum and maximum score from the list respectively. The corresponding normalized confidence score for text and speech are *TextConf* and *SpeechConf* respectively. Text and speech confidence scores are evaluated (*TextConf* and *SpeechConf)* for the test data.

7.1.3 FUZZY-RULE SYSTEM INCORPORATING DECISION-LEVEL FUSION.

This work used a Takagi-Sugeno or Sugeno fuzzy model, developed by Tagaki and Sugeno, 1993 and (Sugeno and Kang, 1988). The Sugeno fuzzy model uses an efficient method to generate fuzzy rules for a provided input-output data. In this model, a fuzzy rule has the form:

$$\text{If } x \text{ is } A \text{ and } y \text{ is } B \text{ then } z = f(x,y) \tag{7.4}$$

where *A* and *B* are the fuzzy sets in the antecedent, *x* and *y* are input variables, and *z* is the crisp function in the consequent. This process is achieved in three steps: Fuzzification of input variables, Rule evaluation (inference), and Defuzzification (Jang *et al.*, 1997).

7.1.3.1 Fuzzification

The confidence score of each review for text and speech (*TextConf* and *SpeechConf)* acquired from second phase is fuzzified using triangular membership function. In triangular fuzzy membership, each linguistic term *T* depends on three scalar parameters, *d*, *e*, *f*, where *e* locates the peak and, *d* and *f* locate base of the triangle. A membership function (MF) for a fuzzy set *S* on the universe of discourse X is defined as $\mu_S : X \rightarrow [0,1]$, where each element of X is transformed to a value between 0 and 1. The equation for triangular function defined by the three parameters, *d*, *e* and *f*, where $d < e < f$ is given in (7.5).

$$\mu_S(x) = \begin{cases} 0, x \le d \\ (x-d)/(e-d), d < x \le e \\ (f-x)/(f-e), e < x \le f \\ 0, x \ge f \end{cases} \tag{7.5}$$

The graphical presentation of this MF is in Fig. 7.1 (a) having parameter values: *d*=0.2, *e*=0.5 and *f*=0.8. This MF is applied to build two input fuzzy sets: Low (L) and High (H) for the normalized input variables: text (*x_t)* and speech (*x_s)*. The range of these variables *x_t* and *x_s* is [0,1]. The normalized values have a minimum value, *min*=0 and maximum value, *max*=1. The *mid* value is calculated in (7.6). Following parameters are involved in creating the triangular fuzzy MF for the fuzzy sets Low and High as: - Low: {*min*, *min*, *mid*}and High: {*mid*, *max*, *max*}. These are graphically presented in Fig. 7.1 (b).

$$mid = \frac{(\min + \max)}{2} \tag{7.6}$$

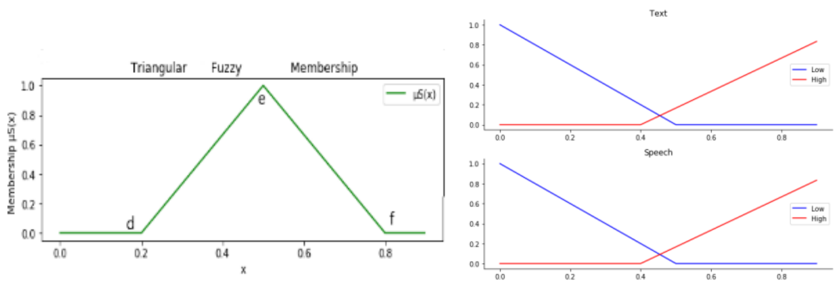


Fig. 7.1. *(a) Triangular Fuzzy Membership; (b) Triangular Fuzzy Membership sets for text and speech input variables.*

7.1.3.2 Formulation of Rules

The novelty of our work is the proposal of the four fuzzy rules illustrated in Table 7.1. The rules were designed based on the hypothesis that if both text and speech confidence scores are low it indicates that the predicted sentiment of review by unimodal SVM should be flipped, and if both scores of text and speech are high then it indicates that the predicted sentiment of review by unimodal SVM should be retained. In cases when one of the text or speech confidence scores is low and the other is high, then the lower text confidence score implies a flip and a higher text confidence score implies retaining the predicted sentiment. 'Flip' means a change of the polarity of sentiment from positive to negative or vice versa and 'Retain' means to not change the polarity of the predicted sentiment. This is inspired by the fact that in a unimodal SVM classifier, the linguistic features have proved to perform better than acoustic features (Sun *et al.*, 2019; Zhao *et al.*, 2019; Cai *et al.*, 2019; Vashishtha and Susan, 2020 a). The formulation of the suggested fuzzy rule-based system is explained next.

$$w_{R1} = text\_low \wedge speech\_low \tag{7.7}$$
$$w_{R2} = text\_high \wedge speech\_low \tag{7.8}$$
$$w_{R3} = text\_low \wedge speech\_high \tag{7.9}$$
$$w_{R4} = text\_high \wedge speech\_high \tag{7.10}$$

The firing strengths of each rule, $w_{R1} \ldots w_{R4}$, are formulated in (7.7-7.10). The fuzzy AND operator are represented using the symbol "$\wedge$". The variables *text_low* and *text_high* includes the antecedent part of fuzzy rules which characterize the low and high fuzzy sets for the confidence score of text, *TextConf*. Likewise, speech_low and speech_high include the antecedent part of fuzzy rules which characterize the low and high fuzzy sets for the confidence score of speech, *SpeechConf*.

Table 7.1 The proposed (Sugeno) Fuzzy Rules.

Rule	Text	Speech	Sentiment Output
R1	Low	Low	**Flip** Predicted Sentiment
R2	High	Low	**Retain** Predicted Sentiment
R3	Low	High	**Flip** Predicted Sentiment
R4	High	High	**Retain** Predicted Sentiment

7.1.3.3 Defuzzification

Defuzzification is opposite to fuzzification. It is a conversion process of fuzzy membership functions to discrete formats. In Sugeno fuzzy model, each rule has a discrete output (indicated by the last column in Table 1), thus weighted average is used to get the final output (Rahman *et al.*, 2016). The defuzzified output, z, is computed in (7.12), where w_i indicates the firing strength of each rule i evaluated in (7.7-7.11) and z_i indicates the corresponding sentiment output for each rule computed using Table 7.1.

$$z = \frac{\sum w_i z_i}{\sum w_i} \tag{7.12}$$

The last step is to check the range of the defuzzified output to categorize the review into: Negative or Positive class as shown in (7.13). For the output range *min*=0 and *max*=1. This range is equally divided into two parts: - Negative: 0-*mid* and Positive: *mid*-*max*.

$$Output = \begin{cases} Negative, 0 \le z < 0.5 \\ Positive, 0.5 \le z \le 1 \end{cases} \tag{7.13}$$

To summarize, the main contribution of our work is a decision-level fusion of text and speech classifiers, involving a novel set of fuzzy rules. The probabilistic confidence scores of the classifiers are interpreted as fuzzy memberships. The fuzzy rules imply that in case of a conflicting decision, the text classifier is given more precedence for determining the sentiment. The overall process flow of Text- Speech fuzzy rule model is depicted in Fig. 7.2.

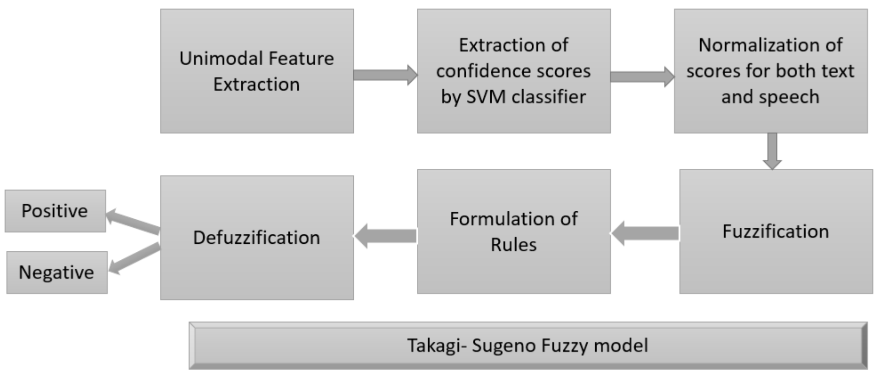


Fig. 7.2. *Overall process flow of Text- Speech Fuzzy Rule Model.*

7.2 EXPERIMENTAL SETUP AND IMPLEMENTATION

Our supervised fuzzy rule-based system for multimodal sentiment analysis has been carried out as per the procedure explained in section 7.1. There are numerous multimodal datasets that consists of sentiment annotations, including the CMU-MOSI dataset (Zadeh *et al.*, 2016), , YouTube (Morency *et al.*, 2011), ICT-MMMO (Wöllmer *et al.*, 2013) and MOUD (Pérez-Rosas *et al.*, 2013), yet only CMU-MOSI provides English dataset with utterance level sentiment labels. Thus, this dataset is used in our work. CMU-MOSI dataset is marked for video opinions of YouTube movie reviews. This dataset has abundant sentimental expressions, with product reviews in English given by 93 people. The videos are subdivided into clips, and a sentiment score between -3 to +3 has been allotted to each clip by five different annotators. The labels' average is measured as the sentiment polarity and categorized into: positive and negative classes.

This dataset also contains the audio and transcripts of these videos. There is total of 2199 opinion utterances, out of them 1176 are positive and 1023 are negative. Each audio has an average of 23.2 opinion segments and average length 4.2 seconds. The opinion utterance contains a total of 26,295 words. The authors have implemented supervised classification on this dataset for audio clips and its transcripts. For the supervised learning method- SVM, the train/test split of the dataset is 70-30% for different runs. Particularly 1539 and 660 utterances are randomly selected for training and test respectively. For comparison purposes, SVM classifier with five-fold classification is applied for only text features, only speech features and feature-level fusion of text-speech (concatenation) features. Linear kernel is used for only text-based classification, while Gaussian (rbf) kernel with C=10 and auto-mode for gamma is used for only speech and fusion based classification. The proposed supervised classification-based method on fuzzy rules uses 70% training and 30 % testing split for the dataset of 2199 utterances, and is averaged for five-fold classification. The source code in python is available online[2].

[2]: https://github.com/SrishtiVashishtha/Inferring-Sentiments-from-Supervised-Classification-of-Text-and-Speech-cues-using-Fuzzy-Rules

A comparative analysis of our fuzzy rule-based system with the other state-of-the-art methodologies has been performed. These techniques are implemented on the same CMU-MOSI dataset. Their train/test split of the dataset is almost 70-30%. First 62 individuals are incorporated in the train set. Remaining 31 speaker videos comprise the test set. The training and test sets consists of 1447 and 752 utterances respectively. These state-of-the art techniques have incorporated different supervised classifiers- some of them have implemented SVM classifier (Lian *et al.*, 2018) (Poria *et al.*, 2018) while other have used bidirectional LSTM (Poria *et al.*, 2017), LSTM (Zadeh *et al.*, 2017), Neural network (Majumder *et al.*, 2019), Gated Recurring Units (GRU) (Majumder *et al.*, 2019) and Convolutional Neural Network (CNN) (Poria *et al.*, 2016).

7.3 RESULTS

Firstly, the unimodal features of text and speech are extracted and supervised SVM is applied on these feature vectors, in isolation, for sentiment classification as described in section 7.1.1 The accuracy for text and speech obtained in unimodal sentiment analysis is 71.21% and 58.0% respectively. Then in order to implement multimodal sentiment analysis, the basic feature-level fusion of text and speech features is performed by simply concatenating the features to produce a multimodal feature vector. For sentiment analysis, SVM is applied on to this feature vector which yields an accuracy of 57.5%. These results highlight the supremacy of text cues over speech cues for sentiment analysis. Finally, the fuzzy rule-based system using decision-level fusion (section 3.3) is applied for multimodal sentiment analysis. The proposed approach yields the highest accuracy of 82.5%, precision is 0.826, recall is 0.823 and F-score is 0.844 compared to other SVM methods. Table 7.2 presents the comparison of our methods.

Table 7.2 Comparison of Text- Speech Fuzzy Rule sysetm with baseline methods.

Methods	**Accuracy**	**Precision**	**Recall**	**F-Score**
Supervised SVM **Text**	71.21	0.712	0.71	0.71
Supervised SVM **Speech**	58.0	0.585	0.559	0.534
Supervised SVM **Text+ Speech**	57.5	0.567	0.553	0.537
Text- Speech Fuzzy Rule Model	**82.5**	**0.826**	**0.823**	**0.844**

```
Confidence score of Text: 0.263
Confidence score of Audio: 0.516
Firing Strength of Rules:
wr1: 0.0
wr2: 0.0
wr3: 0.19333333333333333
wr4: 0.0

Predicted Label by unimodal SVM: 0
Fuzzy Output(z): 1.0

Output after Defuzzification: Positive
Review actual sentiment: 1
(0: Negative, 1: Positive)
```

```
Confidence score of Text: 0.263
Confidence score of Audio: 0.685
Firing Strength of Rules:
wr1: 0.0
wr2: 0.0
wr3: 0.474
wr4: 0.0

Predicted Label by unimodal SVM: 1
Fuzzy Output(z): 0.0

Output after Defuzzification: Negative
Review actual sentiment: 0
(0: Negative, 1: Positive)
```

Fig. 7.3. *(a) Sample Positive Review of CMU-MOSI dataset; (b) Sample Negative Review of CMU-MOSI dataset.*

Processing of a single positive and a negative review of CMU-MOSI dataset is demonstrated in Fig. 7.3 (a) and Fig 7.3 (b) respectively (label for SVM positive: 1 and negative :0). One can observe the sentiment label predicted by unimodal SVM is incorrect. By the application of Sugeno fuzzy model the fuzzy output detects the correct sentiment.

The proposed fuzzy rule-based system is compared with eight state-of-the-art techniques for supervised sentiment analysis: i) SVM: SOTA[1]= Poria *et al.*, 2018, ii) bc-LSTM: SOTA[2]= Poria *et al.*, 2017, iii) Hierarchical fusion: SOTA[3]= Majumder *et al.*, 2018, iv) Context-aware hierarchical fusion: SOTA[3]= Majumder *et al.*, 2018, v) Early fusion: SOTA[3]= Majumder *et al.*, 2018 vi) Tensor Hierarchical Fusion: SOTA[4] = Zadeh *et al.*, 2017 vii) Tensor Context-aware Hierarchical Fusion: SOTA[4]= Zadeh *et al.*, 2017 and viii) Convolutional MKL SOTA[5]= Poria *et al.*, 2016 .

Table 7.3 Comparison of Text- Speech Fuzzy Rule system with state-of-the-art techniques.

S. No	Methods	Classifiers	Training Parameter	Accuracy
1.	SOTA[1] **T+S**	SVM	≈Train-70%, Test-30%	75.8
2.	SOTA[2] **T+S**	bc-LSTM	≈Train-70%, Test-30%	80.2
3.	SOTA[3] HFusion **T+S**	Neural Network	≈Train-70%, Test-30%	77.3
4.	SOTA[3] CHFusion **T+S**	GRU	≈Train-70%, Test-30%	79.1
5.	SOTA[3] Early Fusion **T+S**	SVM	≈Train-70%, Test-30%	77.1
6.	SOTA[4]Tensor Hierarchical Fusion **T+S**	Deep Neural Network	≈Train-70%, Test-30%	76.3

7.	SOTA[4]Tensor Context-aware Hierarchical Fusion **T+S**	LSTM	≈Train-70%, Test-30%	77
8.	SOTA[5] Convolutional MKL **T+S**	CNN	≈Train-70%, Test-30%	77.3
9.	Fuzzy Rule Based **T+S**	Fuzzy Rule based	Train-70%, Test-30%	**82.5**

Table 7.3 compares our fuzzy rule-based approach with other fusion methods for CMU-MOSI dataset in terms of accuracy; bold font depicts highest accuracy, where T stands for text and S for Speech. Simple concatenation of Speech and text features produces bimodal features (Poria *et al.*, 2018; Majumder *et al.*, 2018). These bimodal features are fed into SVM for sentiment classification. The accuracy yielded by these methods are 75.8% (Poria *et al.*, 2018) and 77.1% (Majumder *et al.*, 2018). Majumder *et al.* performed three different methods for multimodal sentiment analysis: Hierarchical fusion, Context-aware hierarchical fusion and Early fusion (Majumder *et al.*, 2018). The first one implements a simple neural network with softmax output gives 77.3% accuracy, second one is the enhancement of first method by introducing GRU (Gated Recurring Units) to model semantic dependency among the utterances yields 79.1% accuracy and the third one is a simple concatenation of text and speech features fed into SVM. This method gives 77.1% accuracy.

Zadeh *et al.*, 2017 proposed a tensor fusion network that learns intra-modality and inter-modality dynamics end-to-end in multimodal sentiment analysis. This tensor network is a fully connected deep neural network, it achieves 76.3% accuracy for Hierarchical fusion and achieves 77% by adding LSTM for Context-aware Hierarchical fusion of text and speech features. A multiple kernel learning classifier based on Convolutional Neural Network (CNN) developed by Poria *et al.*, 2016 yields 77.3% accuracy. Observations from the results report that our fuzzy rule-based system achieves an outstanding accuracy of 82.5% compared to supervised machine learning methods. Fig. 7.4. shows the dataset distribution of CMU MOSI dataset and summarizes the comparison of accuracies for different classifiers outlined in the second column of Table 7.3 in the form of a visual graph.

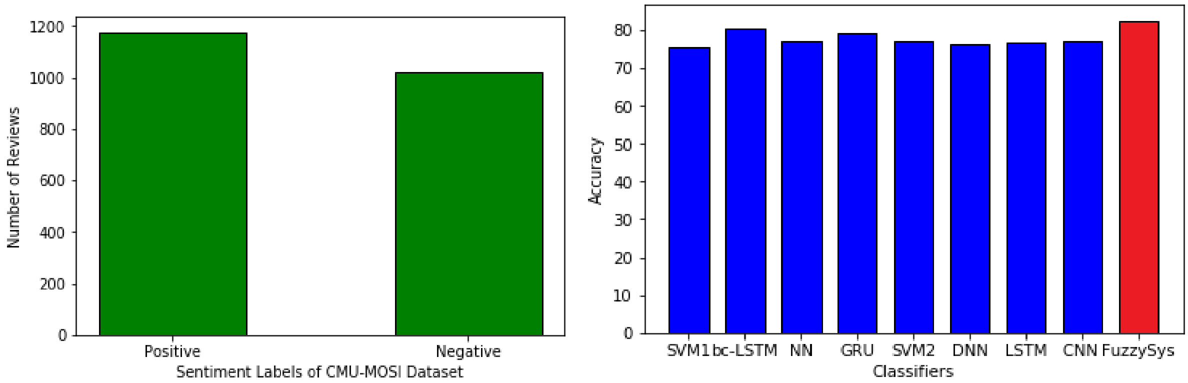


Fig. 7.4 *(a) Dataset Distribution of CMU MOSI dataset; (b) Comparison of accuracies for different classifiers*

7.4 UNSUPERVISED FUZZY INFERENCE SYSTEM FOR SPEECH EMOTION RECOGNITION USING AUDIO AND TEXT CUES

Speech Emotion Recognition (SER) is the technique for determining underlying emotions from speech samples. Usually, text transcripts supplement vocal cues and contain additional information that boosts the SER process. An unsupervised Fuzzy Inference System (FIS) is developed for SER, in this chapter, that incorporates audio and text features. The extracted features are pitch, energy, and textual sentiment score. The proposed system is based on the Mamdani Fuzzy Inference model and is capable of determining four emotions: happy, sad, angry, and neutral. Our FIS has three variants based on the sentiment lexicon- AFINN, SentiWordNet, and VADER selected for computing textual sentiment score as the text feature. The main highlights of this work are: i) formulation of eleven novel fuzzy rules based on audio and text cues for SER ii) comparative analysis of all variants of our proposed unsupervised FIS with six state-of-the-art supervised machine learning approaches for SER iii) both speaker-independent SER and speaker-dependent SER are executed iv) investigations reveal that for few speakers in speaker-dependent SER have received higher accuracies than the others and v) the proposed unsupervised FIS can handle multiple datasets without any training while the supervised machine learning algorithms fail for cross-dataset evaluation. The experiments conducted on speech datasets: SAVEE and RAVDESS, indicate that our FIS has achieved higher accuracy and f1-scores in comparison to the other state-of-the-art methods.

Emotion recognition is usually conducted with the help of a single modality: audio features. But with increasing number of audio posts on social media, textual information is also required to aid the SER process. Models for detecting emotions related to social stigmatizing issues like suicides (Sinha *et al.*, 2019) and sexual harassment (Chowdhury *et al.*, 2019), on social media have been developed by applying sentiment lexicon embeddings for text classification. In text-based Sentiment Analysis (SA), the system can determine sentiment expressed by handling the words (Vashishtha *et al.*, 2020 b), phrases, sentences, and dependencies among them. While in our daily communications, the emotion in the text is generally characterized by the tone of the voice. Hence the combination of both type of features- audio and text, enhances the emotion identification process. In this work, we have adopted both audio features: pitch and energy; along with textual sentiment score for determining emotions of audio clips. We have developed a Mamdani Fuzzy Inference System (FIS) with the help of eleven novel rules based on these features. Our system is able to recognize four emotions: angry, sad, neutral, and happy. The advantage of this unsupervised FIS is that it easily handles the uncertainties, ambiguities, or vagueness that exist while interpreting the emotions; and also mimics the logic of human thought. Moreover, it doesn't require any training thus it saves computational complexity and time.

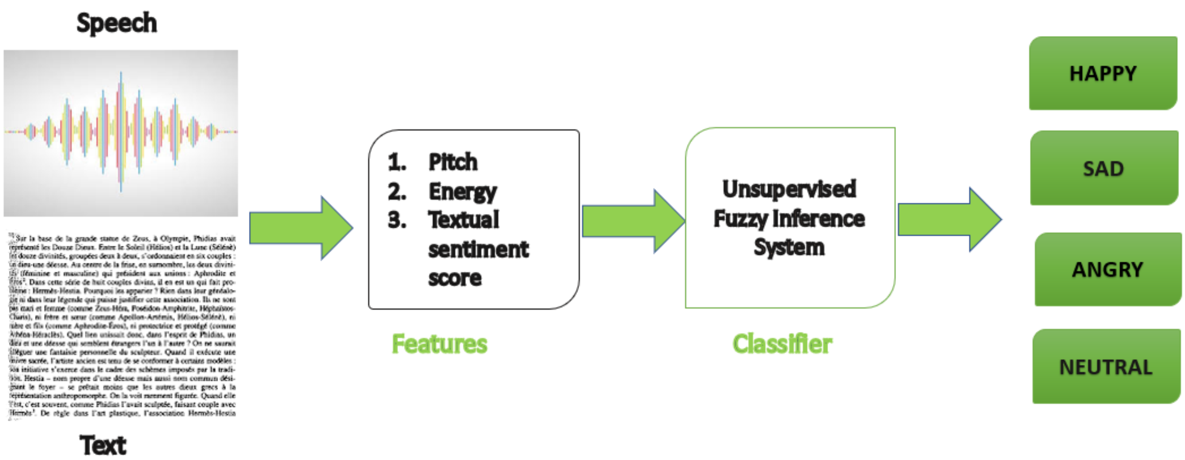


Fig. 7.5 *Overall process flow for Unsupervised Fuzzy Inference System for Speech Emotion Recognition*

7.4.1 PROPOSED FUZZY INFERENCE SYSTEM

7.4.1.1 Audio and Text Feature Extraction

An unsupervised fuzzy inference system for speech emotion recognition has been introduced. This FIS is based on multimodal features: audio and text which are extracted from the audio clips. We have demonstrated how the fusion of acoustic and linguistic features in a fuzzy rule-based system can effectively identify emotion classes: happy, sad, angry, and neutral. The overall process flow for unsupervised FIS for Speech Emotion Recognition is shown in Fig. 7.5.

The audio features extracted from the audio samples are average pitch, p, and average energy, e computed for all frames *n*, each frame *i* having pitch P_i, and Energy E_i, as shown in (7.14 - 7.15). The energy *E* of the frame is computed in (7.16), where the *K* is the length of the frame and the samples {x} in a frame of the audio signal are given by $k= 0$ to $k= K-1$.

$$\text{avg\_ pitch:} \quad p = \frac{\sum_{i=1}^{n} P_i}{n} \tag{7.14}$$

$$\text{avg\_ energy:} \quad e = \frac{\sum_{i=1}^{n} E_i}{n} \tag{7.15}$$

$$\text{energy of an audio signal:} \quad E = \sum_{k=0}^{K-1} x^2(k) \tag{7.16}$$

The text feature is obtained from different sentiment lexicons: AFINN, SentiWordNet, and VADER. The text score t_a of a sentence *s* as computed using AFINN as shown in (7.17).

$$t_a = afn.score(s) \tag{7.17}$$

For the SentiWordNet lexicon, the Universe of Discourse is the set of all the words in each sentence. Each word has one or more synonyms (*synsets* in the SentiWordNet lexicon). Each synonym has positive and negative score derived from the lexicon. The positive score, $\mu_{Pos}(w)$ and negative score, $\mu_{Neg}(w)$ of each word *w*, is computed in (7.18) and (7.19) and can be interpreted as a fuzzy membership pertaining to the fuzzy sets *Pos* and *Neg* (7.20 and 7.21); and X_i is the set of total words in a sentence. The text score t_s of a sentence *s*, computed using

SentiWordNet from (7.19 - 7.20) is shown in (7.23), where *w* is a word in a sentence, *m* is the count of words in the sentence.

$$\mu_{Pos}(w) = \frac{\sum_{synsets} [syn.pos\_score()]}{length(synsets)} \tag{7.18}$$

$$\mu_{Pos}(w) = \frac{\sum_{synsets} [syn.pos\_score()]}{length(synsets)} \tag{7.19}$$

$$\mu_{Neg}(w) = \frac{\sum_{synsets} [syn.neg\_score()]}{length(synsets)} \tag{7.20}$$

$$Pos = \{(w, \mu_{Pos}(w)\}, w \in X_i \tag{7.21}$$

$$Neg = \{(w, \mu_{Neg}(w)\}, w \in X_i \tag{7.22}$$

$$t_s = \sum_{w=1}^{m} \mu_{Pos}(w) - \mu_{Neg}(w) \tag{7.23}$$

The textual score of a sentence is also computed using VADER lexicon's *polarity_scores(s)* method. This method gives compound score of the sentence as an output named as t_v. We consider textual score, denoted in generality as *t*, from each lexicon, one at a time, in our proposed Mamdani FIS. The next subsection discusses the proposed unsupervised Mamdani FIS.

7.4.1.2 Mamdani Fuzzy Inference System based on Audio and Text Features

A fuzzy system with three non-interactive inputs A, B and C (antecedents) and a single output D (consequent) is expressed by a number of r IF–THEN propositions in the Mamdani form:

$$Rule: IF\ A\ is\ A_1^j, B\ is\ B_1^j\ and\ C\ is\ C_1^j, THEN\ D\ is\ D_1^j, j = 1 \ldots r \tag{7.24}$$

where A_{1j}, B_{1j} and C_{1j} are the fuzzy sets representing the *j*th antecedent or premise terms and D_{1j} is the fuzzy set representing the *j*th consequent (Jang *et al.*, 1997). There are three features: two audio and one text, these are the antecedents; and one emotion label as output, this is the consequent. Max-min inference method has been used in our system. The fuzzy output, *D* is obtained by applying the rules to fuzzy input. This output can be defuzzied to a crisp output using defuzzification methods. We have created eleven novel rules using audio and text cues, these are presented in Table 7.4.

1) Justification of formulation of fuzzy rules

Psychological studies prove that prosody features: pitch and energy are two of the important factors that distinguish between emotions, according to human perception. Generally, emotional features are related to arousal. This leads to confusion between anger and happiness, and between sadness and neutrality. Apart from audio features, written text also carries information about the underlying affective state. The semantic part of the speech contains linguistic information which reveals the characteristics of the pronunciation of the utterances based on the rules of the language. Speeches associated with emotions such as angry and happy are loud, fast, and uttered with strong high frequency energy. On the other hand, audios of the emotion: sad and neutral are slow, low in volume, and possess lower high frequency energy (Revathy *et al.*, 2015). These factors influenced us to formulate the rules for audio cues *p* and *e* in Table 7.4. Further, text cues are used for the classification of emotion classes. A low text score implies there is negativity hence it can be either be angry or sad; while a high text score implies there is positivity: a happy state or neutral, as observed from the t column in Table 7.4. The number of rules has been decided based on the analytical experiments performed on the multimodal datasets.

Table 7.4 The proposed eleven Mamdani rules.

Rule	**Pitch (*p*)**	**Energy (*e*)**	**Text (*t*)**	**Emotion**
R1	Low	Low	Med	Neutral
R2	Low	Med	Low	Sad
R3	Low	Med	High	Neutral
R4	Med	Med	Med	Happy
R5	Med	Med	High	Happy
R6	High	High	Low	Angry
R7	High	High	Med	Angry
R8	None	Low	None	Sad
R9	Low	Med	None	Neutral
R10	High	Med	None	Happy
R11	High	High	None	Angry

2) Computation of fuzzy rules

$$w_{R1} = p\_low \wedge e\_low \wedge t\_med \qquad (7.25)$$

$$w_{R2} = p\_low \wedge e\_med \wedge t\_low \qquad (7.26)$$

$$w_{R3} = p\_low \wedge e\_med \wedge t\_high \qquad (7.27)$$

$$w_{R4} = p\_med \wedge e\_med \wedge t\_med \qquad (7.28)$$

$$w_{R5} = p\_med \wedge e\_med \wedge t\_high \qquad (7.29)$$

$$w_{R6} = p\_high \wedge e\_high \wedge t\_low \qquad (7.30)$$

$$w_{R7} = p\_high \wedge e\_high \wedge t\_med \qquad (7.31)$$

$$w_{R8} = e\_low \qquad (7.32)$$

$$w_{R9} = p\_low \wedge e\_med \qquad (7.33)$$

$$w_{R10} = p\_high \wedge e\_med \qquad (7.34)$$

$$w_{R11} = p\_high \wedge e\_high \qquad (7.35)$$

$$w_{angry} = w_{R6} \vee w_{R7} \vee w_{R11} \qquad (7.36)$$

$$w_{sad} = w_{R2} \vee w_{R8} \qquad (7.37)$$

$$w_{neutral} = w_{R1} \vee w_{R3} \vee w_{R9} \qquad (7.38)$$

$$w_{happy} = w_{R4} \vee w_{R5} \vee w_{R10} \qquad (7.39)$$

The eqs. (7.24 - 7.35) represent the firing strength of eleven rules $w_{R1} ... w_{R11}$ shown in Table 7.4; the symbol $\wedge$ represents fuzzy AND operator and the symbol $\vee$ represents the fuzzy OR operator. The variables *p_low*, *p_med* and *p_high* constitute the antecedent part of the fuzzy rules and they depict the *Low*, *Medium* and *High* fuzzy sets for the pitch score *p*, respectively. Similarly, *e_low*, *e_med*, *e_high*, *t_low*, *t_med* and *t_high* constitute the antecedent part of the fuzzy rules and they depict the *Low*, *Medium* and *High* fuzzy sets for the energy score *e* and text score *t*, respectively. The fuzzy sets: *Low*, *Medium* and *High* are created using triangular fuzzy membership for input variables: pitch (*p*), energy (*e*), text (*t*). The output variable (*op*) has *Angry*, *Sad*, *Neutral* and *Happy* fuzzy sets.

The aggregated output (μ_{AG}) in (7.40) utilizes the four firing strengths computed in (7.36 - 7.39), and is used to calculate the defuzzified output in (7.41), where *z* indicates a sample value in the output variable *C*. Defuzzification is the process of converting a fuzzy quantity to a definite quantity (Hellendoorn and Thomas, 1993). We have applied the centroid defuzzification method. This method provides a crisp value based on the center of gravity of the fuzzy set. Finally, the defuzzified output is classified into different emotion classes in (7.42). The overall process of speech emotion recognition using our Mamdani FIS based on audio and text cues is presented in Fig.7.6.

$$\mu_{AG} = (w_{angry} \wedge op\_a) \cup (w_{sad} \wedge op\_s) \cup (w_{neutral} \wedge op\_n) \cup (w_{happy} \wedge op\_h) \tag{7.40}$$

$$C = \frac{\sum z \mu_{AG}(z)}{\sum \mu_{AG}(z)} \tag{7.41}$$

$$Output = \begin{Bmatrix} Angry, 0 < C \leq 0.2 \\ Sad, 0.2 < C \leq 0.4 \\ Neutral, 0.4 < C \leq 0.6 \\ Happy, 0.6 < C \leq 0.8 \end{Bmatrix} \tag{7.42}$$

Let's process a transcript "*Withdraw all phony accusations at once.*" The average pitch and average energy computed using (1-3) is *3.99* and *35.6*. We can compute text score by choosing any lexicon, AFINN computes t_a: -2 using (4), SentiWordNet evaluates t_s: -0.125 using (5-9) and VADER calculates t_V: *-0.318*. The speech features with text feature computed from one lexicon at a time are fed as input to our proposed FIS with eleven rules to compute the aggregated fuzzy output (11-26). The defuzzified output is evaluated using (27), in case of AFINN it is *0.5* while for other two it is *0.104*; and the emotion label is detected by (28). The actual label for this audio is “angry”. The AFINN predicts neutral emotion while SentiWordNet and VADER predicts angry emotion correctly.

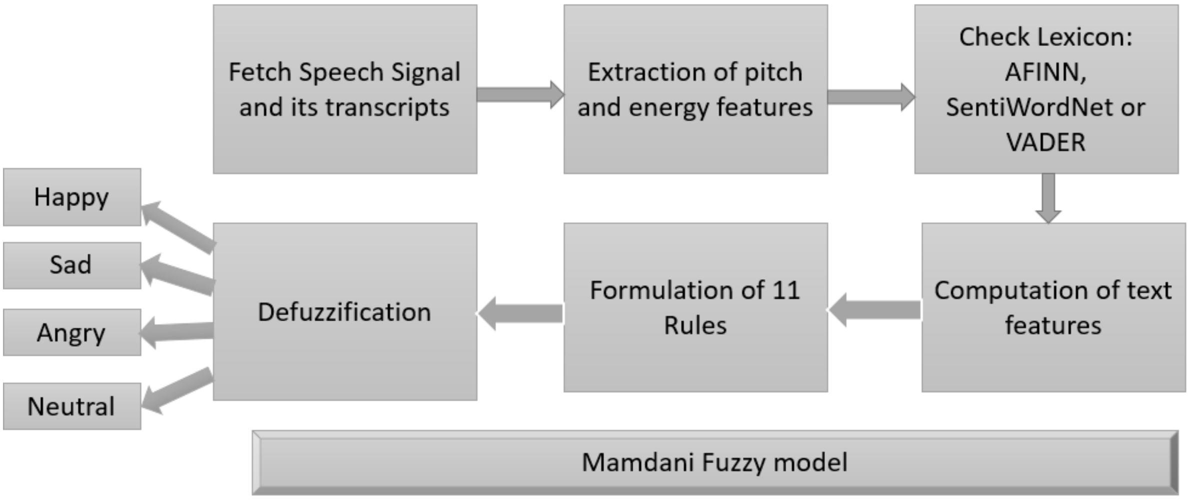


Fig. 7.6. *Speech Emotion Recognition using Mamdani FIS based on audio and text cues.*

In Fig. 7.7, a single sample audio file by VADER variant FIS is processed. The transcript of the file is “*Withdraw all phony accusations at once*”; the audio features: average pitch: *3.99* and average energy: *35.6* is computed using (1-3) and text score: *-0.318* is computed by applying VADER lexicon’s *polarity scores(w)* method. These features are fed as input to our proposed unsupervised Mamdani FIS with eleven rules (11-26). The defuzzified output: *0.104* is evaluated using (27) and finally the emotion label is detected by (28). The actual and predicted label for this audio is “angry”.

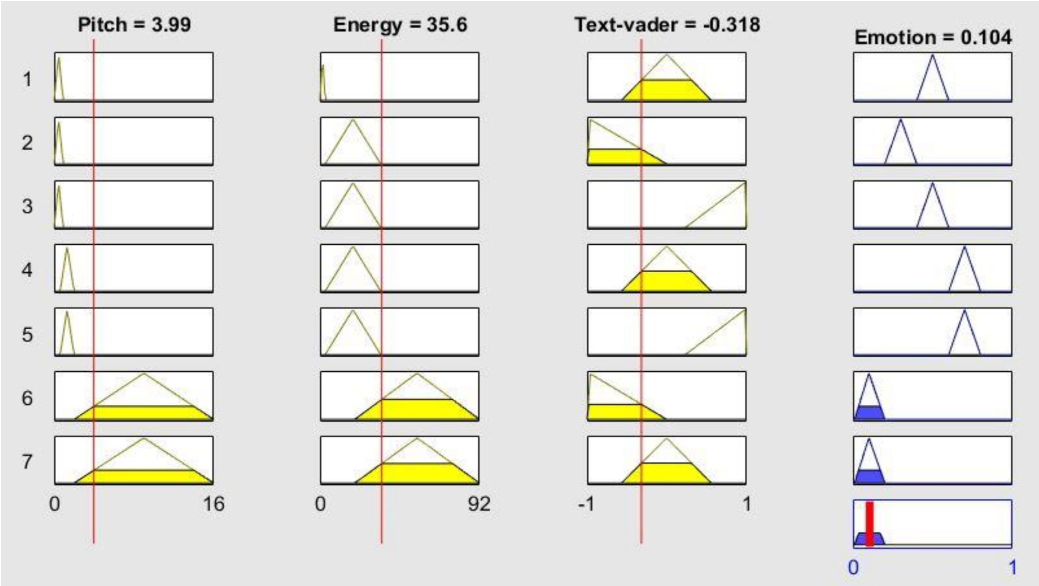


Fig. 7.7. *Processing of a sample audio file by our proposed FIS.*

7.4.2 EXPERIMENTAL SETUP

Our unsupervised fuzzy rule-based FIS for speech emotion recognition has been implemented as per the methodology explained in the previous section. Our code just took about 2-3 seconds to run for each lexicon type. We have extracted the audio features using librosa, which is a popular python library for speech processing. The text features are extracted using three different lexicons: AFINN (Nielsen, 2011), SentiWordNet (Baccianella *et al.*, 2010) and VADER (Hutto and Gilbert, 2014) in python. Our Mamdani FIS is implemented in MATLAB. In our Mamdani FIS we have used max-min inference method, triangular fuzzy membership for all the features and centroid defuzzification method.

In our experiments, we have used two datasets: SAVEE (Jackson and Haq, 2014) and RAVDESS (Livingstone and Russo, 2018). Surrey Audio-Visual Expressed Emotion (SAVEE) dataset was created by collecting the recordings of four native English male speakers all of whom are university students (Jackson and Haq, 2014). It contains 15 sentences for each of the seven emotions: surprise, anger, happiness, sadness, disgust, fear, and neutral. The number of utterances in this database is 480. We have considered four emotions: anger, happiness, sadness, and neutral, thus there is a total of 300 files. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) is a dataset of facial and vocal expressions in English developed by recordings of 24 professional actors (12 female and 12 male) (Livingstone and Russo, 2018). Only the speech audio files are used in our work. There are only two statements spoken in eight emotions: calm, happy, sad, angry, fearful, surprise, and disgust. There is a total of 1440 files, but since we have considered four emotions: happy, sad, angry and neutral, hence there exists 672 audio files. The dataset distribution of both datasets for each emotion class is shown in Table 7.5. We have examined these datasets because of their popularity in the field of emotion recognition.

Table 7.5 Dataset Distribution for each emotion class.

Dataset	Angry	Sad	Neutral	Happy
SAVEE (Jackson and Haq, 2014)	60	60	120	60
RAVDESS (Livingstone and Russo, 2018)	192	192	96	192

Our unsupervised FIS is compared with six state-of-the-art learning methodologies for SER: i) Linear SVM (Sinith *et al.*, 2015), ii) Gaussian SVM (Ramakrishnan and Emary, 2013), iii) SVM (Aouani and Ayed, 2018), iv) Random Forest (Iliou and Anagnostopoulos, 2009), v) Multi- Layer Perceptron (MLP) (Iliou and Anagnostopoulos, 2009) and vi) LSTM (Rahman *et al*., 2016). The details about these methodologies is discussed in next section. The training-testing split of the datasets is 70-30% and we report five-fold cross validation results.

7.4.3 Results And Discussions

We have proposed a fuzzy-rule based system that detects emotions from audio clips using both audio and text features. We have conducted three types of experiments on both datasets: speaker-independent SER, speaker-dependent SER and cross-dataset experiment on Multi-Layer Perceptron (MLP) (Iliou and Anagnostopoulos, 2009). There are six state-of-the art methods that have been compared with our proposed Mamdani FIS. Sinith *et al*. used MFCC, pitch and energy speech features to determine the emotion classes from speech samples using Linear SVM classifier (Sinith *et al.*, 2015). Ramakrishnan *et al*. employed Gaussian SVM classifier by utilizing the following speech features: pitch, centroid, flatness, MFCC, chroma, mel filter, contrast, zero crossing rate for SER (Ramakrishnan and Emary, 2013). Another SVM classifier (Aouani and Ayed, 2018) was implemented for SER using MFCC and mel filter: speech cues to detect emotions of SAVEE dataset (Jackson and Haq, 2014). Different classifiers- Random Forest classifier and MLP for emotion recognition were compared by extracting MFCC, pitch, and energy speech features (Iliou and Anagnostopoulos, 2009). A Long short-term memory (LSTM) model (Rahman *et al*., 2016) has been implemented by extracting textual features. We have not compared to our previous work (Vashishtha and Susan, 2020 a), because there, the proposed fuzzy rules for Sugeno FIS system were designed for the sentiment classification while the aim of the current research work is SER. We modified these methods by determining emotions for four classes: happy, sad, angry, and neutral on both datasets; and the train/test split for these supervised machine learning classifiers is set to 70-30% with five-fold cross-validation. Comparitive analysis of our unsupervised FIS with state-of-the-art is done by considering the 30% test data only.

7.4.3.1 Experiments on SAVEE dataset

The speaker-independent speech emotion recognition using our Mamdani Fuzzy Inference System based on audio and text cues is conducted. We have three variants- AFINN, SentiWordNet, and VADER, based on the sentiment lexicon applied for extracting text cues. The system's performance is measured using accuracy, macro f1-score ,and micro f1- score. The results are depicted in Table 7.6. We can observe VADER lexicon variant achieves the highest scores: 52.67% accuracy, 0.395 macro f1-score ,and 0.527 micro f1- score. The details of the comparative analysis of our system with state-of-the-art methods are shown in Table 7.7. We can identify that our proposed system has gained the highest scores for all the measures. The speaker-dependent SER using our proposed system is executed. There are four speakers in SAVEE dataset, the accuracy is computed for each speaker-lexicon combination. The comparison chart is presented using a bar plot in Fig. 7.8. We can discover that speaker 2 and speaker 4 have performed the worst by achieving 37.33%, while speaker 3 has accomplished the highest accuracy for all the lexicons. The comparison between the lexicons in speaker 3 reveals that SentiWordNet has the highest accuracy of 64% while AFINN has 62.67% and VADER has 60%. The accuracies for all the lexicon variants in speaker 1 and speaker 3 are greater than the accuracies obtained in speaker-independent SER.

Table 7.6 Our Mamdani Fuzzy Inference System based on Speech & Text features on SAVEE dataset.

Lexicon	Accuracy (%)	Macro F1	Micro F1
AFINN	48.44	0.371	0.484
SentiWordNet	50.67	0.381	0.507
VADER	**52.67**	**0.395**	**0.527**

Table 7.7 Comparison of state-of-the-art vs Speech Text Mamdani FIS on SAVEE dataset.

Method	Classifier	Accuracy (%)	Macro F1	Micro F1
(Sinith *et al.*, 2015)	Linear SVM	22.22	0.208	0.222
(Ramakrishnan and Emary, 2013)	Gaussian SVM	45.6	0.156	0.456
(Aouani and Ayed, 2018)	SVM	37.78	0.137	0.378
(Iliou and Anagnostopoulos, 2009)	Random Forest	46.17	0.394	0.462
(Iliou and Anagnostopoulos, 2009)	MLP	37.03	0.313	0.37
(Rahman *et al.*, 2016)	LSTM	23.33	0.079	0.189
Speech Text Mamdani FIS	**Mamdani FIS**	**52.67**	**0.395**	**0.527**

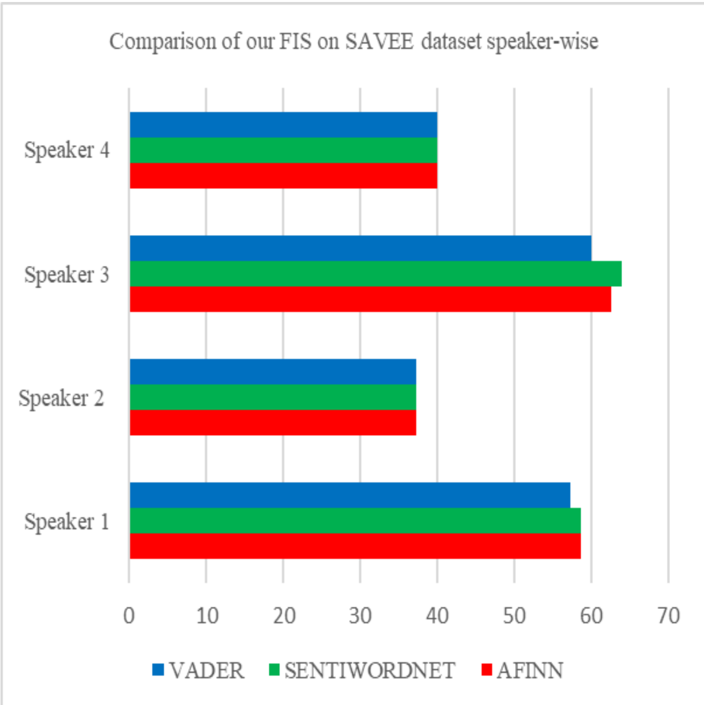


Fig. 7.8. *Comparison of accuracies of our FIS for SER on SAVEE dataset speaker-wise.*

7.4.3.2 Experiments on RAVDESS dataset

Similar experiments have been conducted on RAVDESS dataset. This dataset contains only two statements: “Kids are talking by the door" and "Dogs are sitting by the door”. These statements have neutral content with respect to linguistic information. The results of speaker independent SER using our proposed Mamdani FIS for all the lexicon variants is displayed in Table 7.8. The SentiWordNet variant has accomplished the highest score values of 52.52% accuracy, 0.411 macro f1-score and 0.525 micro f1-score. This is because the textual sentiment scores obtained by this lexicon are better than other lexicons for these two statements. Table 7.9 depicts the comparison of our proposed system with state-of-the-art methods. Our system has attained the highest scores with 52.52 % accuracy, 0.403 macro f1-score, and 0.525 micro f1-score. Random Forest classifier has acquired second-highest scores in both datasets. The speaker-dependent SER by adopting our proposed technique is implemented. RAVDESS dataset contains 24 speakers; a comparison of accuracy achieved by each speaker for each lexicon is presented in Fig. 7.9. From the observations, we report that except for some singular cases, like speaker 15 has 35.7 % accuracy in SentiWordNet and 42.86% in other lexicons, the three lexicons perform at par with each other. Speakers 9, 10, and 14 have gained the highest accuracy of 53.6% in all lexicon variants which is higher than the accuracies achieved in speaker-independent SER. Four speakers 6, 7, 21, and 23 have acquired 50% accuracy, while speaker 4 depicts the worst performance by obtaining 28.57% accuracy.

Table 7.8 Our Mamdani Fuzzy Inference System based on Speech & Text features on RAVDESS dataset.

Lexicon	Accuracy (%)	Macro F1	Micro F1
AFINN	47.57	0.38	0.475
SentiWordNet	**52.52**	**0.411**	**0.525**
VADER	49.55	0.403	0.495

Table 7.9 Comparison of state-of-the-art vs our Mamdani FIS on RAVDESS dataset.

Method	Classifier	Accuracy (%)	Macro F1	Micro F1
(Sinith *et al.*, 2015)	Linear SVM	29.4	0.275	0.294
(Ramakrishnan and Emary, 2013)	Gaussian SVM	26.24	0.104	0.262
(Aouani and Ayed, 2018)	SVM	24.8	0.131	0.248

(Iliou and Anagnostopoulos, 2009)	Random Forest	38.8	0.349	0.388
(Iliou and Anagnostopoulos, 2009)	MLP	31.8	0.159	0.318
(Rahman *et al*., 2016)	LSTM	28.22	0.11	0.282
Speech Text Mamdani FIS	**Mamdani FIS**	**52.52**	**0.403**	**0.525**

The class-wise metrics report for the best variant- VADER in SAVEE, and for the best variant- SentiWordNet in RAVDESS, is depicted in Table 7.10. The Ablation analysis of our Mamdani FIS on SAVEE and RAVDESS datasets for their best variants is shown in Table 7.11. It can be clearly observed that "only speech" and "only text" features underperform; the combination of both features achieves higher accuracy and f1-scores.

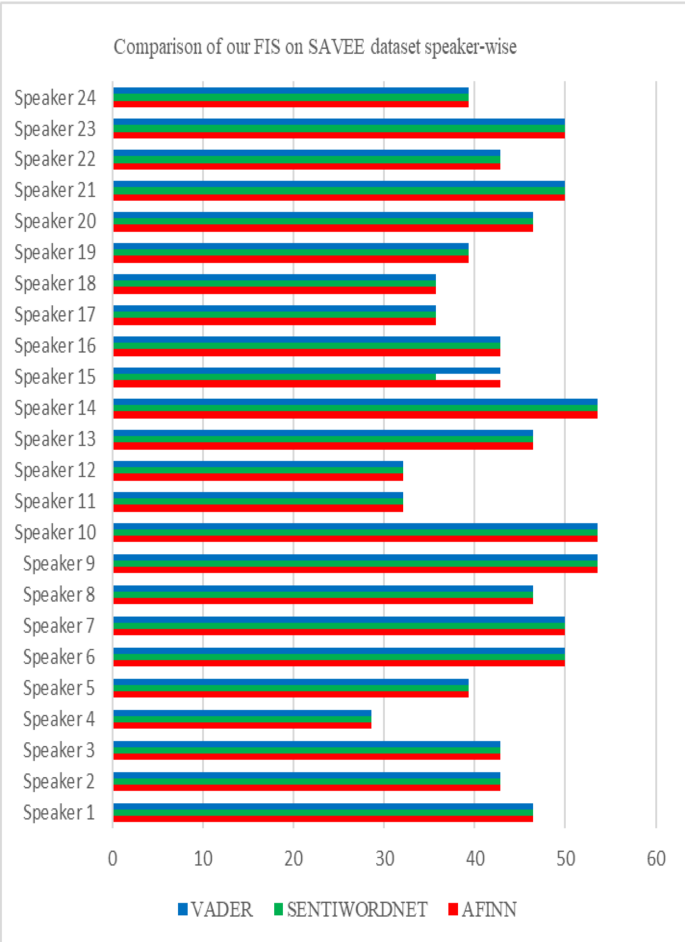


Fig. 7.9. *Comparison of accuracies of our FIS for SER on RAVDEES dataset speaker-wise.*

Table 7.10 Emotion class – wise F1 score report for both datasets.

Dataset	Angry	Sad	Neutral	Happy
SAVEE	0.52	0.11	0.70	0.33
RAVDESS	0.36	0.65	0.18	0.49

Table 7.11 Ablation Analysis of our Mamdani FIS for both datasets.

Dataset	Features	Accuracy (%)	Macro F1	Micro F1
SAVEE	Only Speech	45.56	0.359	0.456
SAVEE	Only Text	20.894	0.1344	0.209
SAVEE	Speech + Text	**52.67**	**0.395**	**0.527**
RAVDESS	Only Speech	39.6	0.299	0.396
RAVDESS	Only Text	26.73	0.105	0.267
RAVDESS	Speech + Text	**52.52**	**0.403**	**0.525**

The third type of experiment is cross-dataset experiment on Multi-Layer Perceptron (MLP) for SER (Iliou and Anagnostopoulos, 2009). We have trained a machine learning algorithm with SAVEE dataset and tested it on RAVDESS dataset and noted the accuracy and f1-scores. Similar experiments are conducted by training the supervised machine learning algorithm on RAVDESS dataset and testing it on SAVEE dataset. We selected MLP classifier for this purpose and conducted the experiment with five-fold cross validation. The results of this experiment are depicted in Table 7.12. We can observe an accuracy of less than 1% in first case and around 8% in second case. We can deduce that the supervised machine learning algorithms are not able to classify data correctly when cross-dataset testing is applied; while for our unsupervised algorithm there is no need for training, the same system can work for both datasets. There is no overhead of training, and computational time is saved.

Table 7.12 Cross- dataset testing using MLP for both datasets.

Cross Dataset	Accuracy (%)	Macro F1	Micro F1
Train SAVEE Test RAVDESS	0.89	0.009	0.009
Train RAVDESS Test SAVEE	7.9	0.042	0.079

7.5 CONCLUSION

In this chapter, a novel text and speech based fuzzy rule-based system has been proposed for multimodal sentiment analysis of reviews posted on social media. The key highlights of this work are: i) creation of four novel fuzzy rules based on text and speech cues to evaluate each review's sentiment, ii) the proposed decision-level fusion method performs better than unimodal and basic feature-level text-speech fusion using supervised machine learning: Support Vector Machine (SVM) and iii) comparison of our proposed rule-based system for sentiment analysis with eight state-of-the-art techniques for supervised machine learning. Experiments unveil that the proposed system achieves sufficiently well accuracy of 82.5%. This due to the fact that by incorporating fuzzy inferencing with linguistic and acoustic behavior, the system deals well with vagueness and ambiguity related to text and speech.

In this chapter, a novel fuzzy rule-based Mamdani FIS has been proposed for SER using both audio and text features. The system is implemented on two speech datasets: SAVEE and RAVDESS. The main highlights of this work are: i) formulation of eleven novel fuzzy rules based on audio and text cues for SER ii) comparative analysis of all variants of our proposed unsupervised rule-based technique for SER with six state-of-the-art supervised machine learning approaches iii) both speaker-independent SER and speaker-dependent SER are executed iv) investigations unveil that few speakers in speaker-dependent SER have gained higher accuracies than the others and v) proposed unsupervised FIS can handle multiple datasets without any training while the supervised machine learning algorithms fail for cross-dataset evaluation. The state-of-the-art methods use only audio features but, in our FIS, we have included textual sentiment score as a text feature along with audio features. This is because apart from audio features, the written text also carries information about the hidden emotional state, and SA is closely related to the detection of emotions. The semantic part of the speech contains linguistic information which reveals the characteristics of the pronunciation of the utterances based on the rules of the language. Our results indicate an effective and computationally feasible solution for detecting emotions with the help of audio and text features. In future, we would take into account language disparities, like slang and mixed-language content, in usage across different social media platforms.

CHAPTER 8

CONCLUSION

The fuzzy logic-based sentiment analysis system has been developed in this thesis for evaluating the sentiment expressed in online reviews and social media posts. This concept is used to solve several sentiment analysis problems such as

- Extraction of sentiment cognition words
- Extraction of sentiment cognition keyphrases
- Implementation of neuro-fuzzy networks to detect sentiment
- Computational analysis of social media posts based on fuzzy logic
- Multimodal SA and Speech Emotion Recognition by fuzzy inferencing

All the problems except multimodal SA and speech emotion recognition, relies on the application of fuzzy logic-based algorithms, using text features, for identifying the sentiment contained in the text. The last problem is solved by applying supervised and unsupervised classification algorithms with fuzzy rules based on both text and speech features. A novel fuzzy model based on fuzzy entropy, clustering, and sentiment lexicon has been developed to extract sentiment cognition words. It has been extended to extract the significant phrases by incorporating fuzzy linguistic hedges. A new collection of 9 fuzzy rules is proposed to compute the sentiment of text using any sentiment lexicon. A novel neuro-fuzzy system, containing a novel set of 64 fuzzy rules, that combines sentiment scores from multiple lexicons to classify tweets has been deployed.

A supervised fuzzy rule-based system for multimodal sentiment classification has been built that can identify the sentiment expressed in reviews on social media platform. This fuzzy rule-based system, consists of 4 novel fuzzy rules, deployed by the joint use of linguistic and acoustic features. An unsupervised Fuzzy Inference System (FIS) is developed for Speech Emotion Recognition that incorporates audio and text features. A novel set of 11 novel fuzzy rules based on audio and text cues has been formulated, that is capable of determining emotions from speech signals. It can be deduced that for analysing sentiment from movie review dataset the fuzzy sentiment cognition models that shortlists words and keyphrases, described in Chapter 3 and Chapter 4, can be used for sentiment classification. The sentiment analysis of social media data- Twitter datasets can be executed by the fuzzy rule-based model described in Chapter 6.

In Chapter 3, a novel model has been developed that extracts out the high sentiment cognition words from text in an automated way using fuzzy entropy measure and clustering for fuzzy reasoning. The fuzzy scores are extracted from the SentiWordNet lexicon that has been compiled with the help of human annotations. A clustering technique is applied to these fuzzy entropy values to divide them into two clusters of significant and nonsignificant words. We are interested in the cluster with low fuzzy entropy values since they correspond to the important words which contribute in the evaluation of the sentiment of text. Our approach emulates the way human mind works by glancing at only a few key important words in text, to determine the polarity of a review. These shortlisted words are trained using an LSTM network. Our approach is implemented on two movie review datasets: 1) the IMDB dataset and 2) the polarity dataset by Pang and Lee. We have compared both data sets with other state-of-the-art approaches of SA. It is observed that our model has achieved the highest accuracies of 89.8% in IMDB and 78.5% in Pang and Lee data sets, as compared to other methods. A fuzzy logic-based technique is applied to online reviews to compute the fuzzy sentiment score. Two sentiment lexicons- SentiWordNet and AFINN are used to compute the sentiment score of words. The key highlights are: i) proposed an unsupervised approach based on fuzzy logic for sentiment analysis of textual reviews, ii) the proposed model uses fuzzy cardinality as the measure for the evaluation of word polarity scores, iii) our model has two versions based on the sentiment lexicon deployed in the model, iv) our fuzzy cardinality approach is compared to non-fuzzy state-of-the-art methods. Our proposed fuzzy methodology is better than nonfuzzy methods. This is because fuzzy deals with ambiguity in real-world problems. Our approach calculates the strength of average positive and negative scores of each word in each review and these scores are fuzzy. Thus, the strength of fuzzy sets gives better results than simple average scores. The application of fuzzy logic with NLP provides us with results that match human interpretation for sentiment classification. Our approach can be applied to any textual data that is based on online or social media content, like Twitter datasets, product reviews datasets, any other customer review datasets, etc. The limitation of our work is that the scores of words are dependent on lexicons; some words which do not exist in lexicons cannot be processed further.

An unsupervised sentiment classification system that comprehensively formulates phrases, coand mputes their senti-scores (sentiment scores) and polarity using fuzzy linguistic hedges is proposed in Chapter 4. Further, it extracts the keyphrases, significant for SA, using fa uzzy entropy filter with k-means clustering. The key contributions of the work are: i) An unsupervised phrase-level SA approach has been proposed to perform sentiment analysis on online reviews using n-gram techniques, specifically a combination of unigram, bigram, and trigram, ii) Phrases are constructed comprehensively using part-of-speech (POS) Tagger, list of concentrators, dilators, and negators. Their senti-scores and polarity are computed using SentiWordNet lexicon and fuzzy linguistic hedges, iii) Document-level SA on online reviews is executed by extracting high sentiment bearing keyphrases filtered out by fuzzy entropy and k-means clustering, and finally computing the sentiment of the review and iv) The performance of our fuzzy technique is evaluated using the parameters of accuracy and f-score. The results indicate higher classification scores as compared to the state of the art. The advantage of this unsupervised algorithm is that it easily handles the uncertainties, ambiguities or vagueness that exist while interpreting the sentiment; and also mimics the logic of human thought. Moreover, it doesn’t require any training thus it saves computational complexity and time. To the best of our knowledge, till now, keyphrase extraction in Sentiment Analysis and computation of sentiment scores of keyphrases using fuzzy measures, has not been investigated by any researcher.

In Chapter 5, a novel MultiLexANFIS system has been proposed for social sentiment analysis. This architecture is an Adaptive Neuro-Fuzzy Inference System (ANFIS) that incorporates inputs from multiple lexicons to perform sentiment classification of social media content. The key highlights of this work are: 1) a novel neuro-fuzzy system: MultiLexANFIS that combines sentiment scores from multiple lexicons to classify tweets, 2) a novel set of 64 rules for Sugeno-type Fuzzy Inference System, 3) single lexicon based ANFIS to classify tweets and 4) comparison of fuzzy approaches with non-fuzzy approaches affirms the supremacy of our proposed neuro-fuzzy system. ANFIS has deployed optimization with help of a hybrid learning process; that uses least square technique and gradient descent to tune parameters. We can deduce that all single lexicon based ANFIS systems and MultiLexANFIS have lowest RMSE for all datasets compared to other approaches. Results reveal that among the single lexicon ANFIS systems VADER lexicon performs best. The best membership in MultiLexANFIS is triangular (trimf), and gaussian membership function (gaussmf) is at second position. The problem of confusion among the MFs for all datasets has been solved. The MultiLexANFIS is computationally faster than the total combination of three single lexicon based ANFIS for all lexicons. Our neuro-fuzzy network is targeted for applications that aim to search neutral social media content posted by the users. Our results thus indicate an effective and computationally feasible solution to social sentiment analysis that incorporates multiple lexicons onto the same platform. These proposed novel rules are domain independent. We can apply these rules for any textual data that employs lexicons to extract positive and negative scores for each textual document or sentence or review. The proposed system can handle only those Twitter datasets which contain neutral tweets. At present it cannot deal with mixed tweets, i.e., the tweets that include both positive and negative opinions. In future, we will compute the sentiment of mixed tweets by enhancing our model using deep learning.

We have proposed a fuzzy rule-based approach for SA of social media data specifically for twitter datasets in Chapter 6. The novelty of this work is i) the formulation of nine fuzzy rules to evaluate the sentiment class of tweets, ii) proposed unsupervised approach is suitable to any lexicon and iii) and any two- class or three-class sentiment dataset. Two-class datasets have positive and negative sentiment classes while three- class datasets have an additional neutral sentiment class. We learn that fuzzy rules are able to incorporate the fuzziness of positive and negative scores. Fuzzy logic-based systems can deal with vagueness and ambiguity. Advantages of using the fuzzy approach are summarized as i) An important contribution of fuzzy logic is that it provides a way for computing with words, i.e., words can be transformed into numerical values for further computation, ii) Fuzzy logic provides us a desirable way to deal with linguistic problems and iii) it is equipped with logical reasoning that aids in SA process by providing closer views to the exact sentiment values. We have implemented our proposed method by applying three various lexicons: SentiWordNet (Baccianella *et al.*, 2010), AFINN (Nielsen, 2011), and VADER (Gilbert & Hutto, 2014) in isolation with each other on nine publicly available Twitter datasets. Comparison with four state-of-the-art methods for unsupervised sentiment classification and one state-of-the-art supervised machine learning involving SVM classifier reveals that our fuzzy rule-based method performs consistently the best with respect to F1- Micro scores. Our fuzzy rule-based method scores higher F1 Micro scores, Precision, and Recall in the majority of datasets (7 out of 9). The F1- Macro scores are acceptable in all cases if not always the best. The highest F1-Micro scores of 0.865 and 0.842 are achieved by VADER lexicon in Gilbert Tweets and Nuclear Twitter datasets respectively. Moreover, the methods which implement VADER lexicon execute in the least time while the methods which implement SentiWordNet lexicon take maximum time in execution. The metrics precision and recall scores of unsupervised methods AFINN and VADER lexicon performed better compared to SentiWordNet lexicon.

Our unsupervised fuzzy rule-based method with VADER lexicon has performed much better than supervised machine learning involving SVM in terms of all metrics. VADER is quick and computationally economical without comprising F1-scores. It works excellently well on social media text. It doesn't require any training data. It has performed the best because this lexicon is best suited for social media posts. It handles emojis, slang, emoticons, and acronyms very well and evaluates the emoticons contained in the text. Tremendous benefits can be obtained by using VADER in micro-blogging websites wherein the text data is of complex nature. In future, we can implement our fuzzy rule-based approach on other domains like movie reviews, product reviews, etc. for Sentiment Analysis and opinion mining. A fuzzy logic-based approach for dynamic plotting of mood swings of tweets is developed. These tweets were posted by cricket fans while watching India Vs Pakistan final ICC world-cup match in June 2017. The novelty of our approach is use of linguistic hedges: *Very*, *More or Less* and *Not*, with fuzzy logic. These hedges describe the mood of user in tweet and captures the sentiment even when it is present in moderate amount. The fuzzy membership values of these hedges are used to plot the fuzzy positive and fuzzy negative sentiment of tweet with respect to time. Finally, we apply defuzzification to get the sentiment of each tweet. We have compared our method with existing methods. The analysis demonstrates that our approach is more sensitive to mood swings and decodes the correct sentiment at each time window.

In Chapter 7, a novel text and speech based fuzzy rule-based system has been proposed for multimodal sentiment analysis of review posted on social media. The key highlights of this work are: i) the creation of four novel fuzzy rules based on text and speech cues to evaluate

each review’s sentiment ii) the proposed decision-level fusion method performs better than unimodal and basic feature-level text-speech fusion using supervised machine learning: Support Vector Machine (SVM) and iii) comparison of our proposed rule-based system for sentiment analysis with eight state-of-the-art techniques for supervised machine learning. Experiments unveil that the proposed system achieves sufficiently well accuracy of 82.5%. This due to the fact that by incorporating fuzzy inferencing with linguistic and acoustic behaviour, the system deals well with vagueness and ambiguity related to text and speech.
A novel fuzzy rule-based Mamdani FIS has been proposed for Speech Emotion Recognition (SER) using both audio and text features. The system is implemented on two speech datasets: SAVEE and RAVDESS. The main highlights of this work are: i) formulation of eleven novel fuzzy rules based on audio and text cues for SER ii) comparative analysis of all variants of our proposed unsupervised rule-based technique for SER with six state-of-the-art supervised machine learning approaches iii) both speaker-independent SER and speaker dependent SER are executed iv) investigations unveil that few speakers in speaker-dependent SER have gained higher accuracies than the others and v) proposed unsupervised FIS can handle multiple datasets without any training while the supervised machine learning algorithms fail for cross-dataset evaluation. The state-of-the-art methods use only audio features but, in our FIS, we have included textual sentiment score as a text feature along with audio features. This is because apart from audio features, the written text also carries information about the hidden emotional state, and SA is closely related to the detection of emotions. The semantic part of the speech contains linguistic information which reveals the characteristics of the pronunciation of the utterances based on the rules of the language. Our results indicate an effective and computationally feasible solution for detecting emotions with the help of audio and text features. In future, we would take into account language disparities, like slangs and mixed language content, in usage across different social media platforms.

As an extension to the work done in this thesis, we suggest to expand our research work on speech emotion recognition and computer vision. We would like to extend our fuzzy rule system proposed in Chapter 7 for classifying more emotion labels including happy, surprise, angry, disgust, fear, sad, calm or neutral in speech emotion recognition. Our fuzzy rule-based system in Chapter 6 can be widened for sentiment classification of image datasets. The utilities of Neuro-fuzzy networks proposed in Chapter 5 can be further explored not only in the field of machine learning and natural language processing, but also in computer vision and pattern recognition problems. We propose to extend the sentiment cognition algorithm for shortlisting words in Chapter 3 by incorporating eye-movement data in our future work. We can also employ a supervised sentiment classification system with keyphrases and its senti-scores as features. We would compute senti-scores for emoticons used in the reviews that were not considered for automatic score calculation executed in Chapter 4. Some jargon, slang, misspellings, and oddly spelled words such as "*ROFL*", "*LOL*", "*verrrrryyyy*", "*funnyyyyy*", "*omggggg*" that probably emphasizes certain feeling is also ignored in this work. We will tackle such acronyms in our future work. We can also take into account language disparities, like slang and mixed-language content, in usage across different social media platforms. We can enhance our algorithms by incorporating fuzzy inferencing into deep neural network models in comparison to the state-of-the-art in deep learning.

LIST OF PUBLICATIONS

Journals

1. Vashishtha, Srishti, and Seba Susan. "Sentiment Cognition from Words Shortlisted by Fuzzy Entropy." *IEEE Transactions on Cognitive and Developmental Systems* (2019). (Impact Factor: 3.379, Citations: 11)
2. Vashishtha, Srishti, and Seba Susan. "Fuzzy rule based unsupervised sentiment analysis from social media posts." *Expert Systems with Applications* 138 (2019): 112834. (Impact Factor: 6.954, Citations: 11)
3. Vashishtha, Srishti, and Seba Susan. "Highlighting Keyphrases using Senti-Scoring and Fuzzy Entropy for Unsupervised Sentiment Analysis." *Expert Systems with Applications* 169 (2021): 114323. (Impact Factor: 6.954, Citations: 90)
4. Vashishtha, Srishti, and Seba Susan. "Neuro-Fuzzy Network incorporating Multiple Lexicons for Social Sentiment Analysis". *Soft Computing* (2021): 1-21. (Impact Factor: 3.643, Citations:0)

Conferences

1. Vashishtha, Srishti, and Seba Susan. "Fuzzy logic based dynamic plotting of mood swings from tweets." In *International Conference on Innovations in Bio-Inspired Computing and Applications*, pp. 129-139. Springer, Cham, 2018. (Citations: 12)
2. Vashishtha, Srishti, and Seba Susan. "Inferring Sentiments from Supervised Classification of Text and Speech cues using Fuzzy Rules." Procedia Computer Science 167 (2020): 1370-1379. (Citations: 12)
3. Vashishtha, Srishti, and Seba Susan. "Fuzzy Interpretation of Word Polarity Scores for Unsupervised Sentiment Analysis." In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT*), pp. 1-6. IEEE, 2020. (Citations: 5)
4. Vashishtha, Srishti, and Seba Susan. "Unsupervised Fuzzy Inference System for Speech Emotion Recognition using audio and text cues (Workshop Paper)." In *2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM)*, pp. 394-403. IEEE, 2020. (Citations: 5)

REFERENCES

[1] Acampora G, Cosma G (2014). A hybrid computational intelligence approach for efficiently evaluating customer sentiments in E-commerce reviews. In 2014 IEEE Symposium on Intelligent Agents (IA), pp. 73-80. IEEE, 2014.

[2] Adamov, A. Z., & Adali, E. (2016, October). Opinion mining and sentiment analysis for contextual online-advertisement. In *2016 IEEE 10th International Conference on Application of Information and Communication Technologies (AICT)* (pp. 1-3). IEEE.

[3] Agarwal, B., Sharma, V. K., & Mittal, N. (2013 a, August). Sentiment classification of review documents using phrase patterns. In 2013 *International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 1577-1580). IEEE.

[4] Agüero-Torales, M. M., Salas, J. I. A., & López-Herrera, A. G. (2021). Deep learning and multilingual sentiment analysis on social media data: An overview. *Applied Soft Computing*, *107*, 107373.

[5] Alharbi, J. R., & Alhalabi, W. S. Sentimental Analysis Using Fuzzy Logic for Cloud Service Feedback Evaluation.

[6] Ali, F., Shaker, E. S., Khan, P., & Kwak, K. S. (2018, October). Feature-based Transportation Sentiment Analysis Using Fuzzy Ontology and SentiWordNet. In *2018 International Conference on Information and Communication Technology Convergence (ICTC)* (pp. 1350-1355). IEEE.

[7] Andreevskaia, A., & Bergler, S. (2006, April). Mining wordnet for a fuzzy sentiment: Sentiment tag extraction from wordnet glosses. In *11th conference of the European chapter of the Association for Computational Linguistics.IEEE.*

[8] Aouani, H., & Ayed, Y. B. (2018, March). Emotion recognition in speech using MFCC with SVM, DSVM and auto-encoder. In *2018 4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)* (pp. 1-5). IEEE.

[9] Appel, O., Chiclana, F., Carter, J., & Fujita, H. (2016). A hybrid approach to the sentiment analysis problem at the sentence level. *Knowledge-Based Systems*, 108, 110-124.

[10] Arslan, Y., Birturk, A., Djumabaev, B., & Küçük, D. (2017, December). Real-time Lexicon-based sentiment analysis experiments on Twitter with a mild (more information, less data) approach. In *2017 IEEE International Conference on Big Data (Big Data)* (pp. 1892-1897). IEEE.

[11] Austermann, A., Esau, N., Kleinjohann, L., & Kleinjohann, B. (2005 a, August). Prosody based emotion recognition for MEXI. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 1138-1144). IEEE.

[12] Austermann, A., Esau, N., Kleinjohann, L., & Kleinjohann, B. (2005 b, August). Fuzzy emotion recognition in natural speech dialogue. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.* (pp. 317-322). IEEE.

[13] Baccianella, S., Esuli, A., & Sebastiani, F. (2010, May). Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In *Lrec* (Vol. 10, No. 2010, pp. 2200-2204).

[14] Bakhtiyari, K., & Husain, H. (2014). Fuzzy model on human emotions recognition. *arXiv preprint arXiv:1407.1474*.

[15] Barnaghi, P. , Ghaffari, P. , & Breslin, J. G. (2016). Opinion mining and sentiment polar- ity on twitter and correlation between events and sentiment. In *Big Data Com- puting Service and Applications (BigDataService), 2016 IEEE Second International Conference on* (pp. 52–57). IEEE. 2016 .

[16] Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc.".

[17] Bedi, P., & Khurana, P. (2020). Sentiment Analysis Using Fuzzy-Deep Learning. In *Proceedings of ICETIT 2019* (pp. 246-257). *Springer, Cham.*

[18] Beineke, P., Hastie, T., & Vaithyanathan, S. (2004, July). The sentimental factor: Improving review classification via human-provided information. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04)* (pp. 263-270).

[19] Bedi, P., & Khurana, P. (2020). Sentiment Analysis Using Fuzzy-Deep Learning. In *Proceedings of ICETIT 2019* (pp. 246-257). *Springer, Cham.*

[20] Bernabé-Moreno, J., Tejeda-Lorente, A., Herce-Zelaya, J., Porcel, C., & Herrera-Viedma, E. (2020). A context-aware embeddings supported method to extract a fuzzy sentiment polarity dictionary. *Knowledge-Based Systems*, 190, 105236.

[21] Bhaskar, J., Sruthi, K., & Nedungadi, P. (2015). Hybrid approach for emotion classification of audio conversation based on text and speech mining. *Procedia Computer Science*, *46*, 635-643.

[22] Bhoir, P., & Kolte, S. (2015, December). Sentiment analysis of movie reviews using lexicon approach. In *2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)* (pp. 1-6). IEEE.

[23] Bidulya, Y., & Brunova, E. (2016, October). Sentiment analysis for bank service quality: A rule-based classifier. In 2016 *IEEE 10th International Conference on Application of Information and Communication Technologies (AICT)* (pp. 1-4). IEEE.

[24] Bing, L., & Chan, K. C. (2014, December). A fuzzy logic approach for opinion mining on large scale twitter data. In *2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing* (pp. 652-657). IEEE.

[25] Bravo-Marquez, F., Mendoza, M., & Poblete, B. (2014). Meta-level sentiment models for big social data analysis. *Knowledge-based systems*, 69, 86-99.

[26] Bruno, G. (2016, December). Text mining and sentiment extraction in central bank documents. In *2016 IEEE International Conference on Big Data (Big Data)* (pp. 1700-1708). IEEE.

[27] Cai, L., Hu, Y., Dong, J., & Zhou, S. (2019). Audio-textual emotion recognition based on improved neural networks. *Mathematical Problems in Engineering*, *2019*.

[28] Çakıt E, Karwowski W, Servi L (2019). Application of soft computing techniques for estimating emotional states expressed in Twitter® time series data. Neural Computing and Applications: 1-14, 2019.

[29] Cavalcanti, D. C., Prudêncio, R. B., Pradhan, S. S., Shah, J. Y., & Pietrobon, R. S. (2011, November). Good to be bad? Distinguishing between positive and negative citations in scientific impact. In *2011 IEEE 23rd International Conference on Tools with Artificial Intelligence* (pp. 156-162). IEEE.

[30] Chang, Y. C. , Chen, S. M. , & Churn-Jung, L. (2008). "Fuzzy interpolative reasoning for sparse fuzzy-rule-based systems based on the areas of fuzzy sets. *IEEE Transactions on Fuzzy Systems, 16* (5 (2008)), 1285.

[31] Chen, C. , Wang, Z. , & Wenjie, L. (2018). Tracking dynamics of opinion behaviors with a content-based sequential opinion influence model. *IEEE Transactions on Affec- tive Computing, 1* , 1 .

[32] Choudhary, N., Singh, R., Bindlish, I., & Shrivastava, M. (2018). Sentiment analysis of code-mixed languages leveraging resource rich languages. *arXiv preprint arXiv:1804.00806*.

[33] Chowdhury, A. G., Sawhney, R., Shah, R., & Mahata, D. (2019, July). # YouToo? detection of personal recollections of sexual harassment on social media. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 2527-2537).

[34] Cosma G, Acampora G (2016). Neuro-fuzzy sentiment analysis for customer review rating prediction. In Sentiment Analysis and Ontology Engineering, pp. 379-397. Springer, Cham, 2016.

[35] Dalal, M. K., & Zaveri, M. A. (2014). Opinion mining from online user reviews using fuzzy linguistic hedges. *Applied computational intelligence and soft computing*, *2014*.

[36] de Sousa, R. F., Rabêlo, R. A., & Moura, R. S. (2015, August). A fuzzy system-based approach to estimate the importance of online customer reviews. In *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (pp. 1-8). IEEE.

[37] Deng, Z. H., Luo, K. H., & Yu, H. L. (2014). A study of supervised term weighting scheme for sentiment analysis. *Expert Systems with Applications*, 41(7), 3506-3513.

[38] Deshmukh, K. C., & Khot, P. G. (2011). Generalized measures of fuzzy entropy and their properties.

[39] De Luca, A., & Termini, S. (1972). A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory. *Information and control*, *20*(4), 301-312.

[40] Dey, A., Jenamani, M., & Thakkar, J. J. (2018). Senti-N-Gram: An n-gram lexicon for sentiment analysis. *Expert Systems with Applications*, *103*, 92-105.

[41] Dong, X., & De Melo, G. (2018, April). Cross-lingual propagation for deep sentiment analysis. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

[42] Dundar, B., Ozdemir, S., & Akay, D. (2016, May). Opinion mining and fuzzy quantification in hotel reviews. In *2016 International Symposium on Networks, Computers and Communications (ISNCC)* (pp. 1-4). IEEE.

[43] Dutu, L. C. , Mauris, G. , & Philippe, B. (2018). A fast and accurate rule-base generation method for Mamdani fuzzy systems. *IEEE Transactions on Fuzzy Systems, 26* (2 (2018)), 715–733.

[44] Elbarougy, R., & Akagi, M. (2012, December). Speech emotion recognition system based on a dimensional approach using a three-layered model. In *Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference* (pp. 1-9). IEEE.

[45] Esau, N., Kleinjohann, L., & Kleinjohann, B. (2005, September). An Adaptable Fuzzy Emotion Model for Emotion Recognition. In *EUSFLAT Conf.* (pp. 73-78).

[46] Ghiassi, M., Skinner, J., & Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with applications*, 40(16), 6266-6282.

[47] Giachanou, A. , & Crestani, F. (2016). Like it or not: A survey of twitter sentiment analysis methods. *ACM Computing Surveys (CSUR), 49* (2), 28 .

[48] Gosai, D. D., Gohil, H. J., & Jayswal, H. S. (2018). A review on a emotion detection and recognization from text using natural language processing. *International Journal of Applied Engineering Research*, *13*(9), 6745-6750.

[49] Gupta, N., & Abhinav, K. R. (2013, October). Fuzzy sentiment analysis on microblogs for movie revenue prediction. In *2013 International Conference on Emerging Trends in Communication, Control, Signal Processing and Computing Applications (C2SPCA)* (pp. 1-4). IEEE.

[50] Hamdan, H. , Béchet, F. , & Bellot, P. (2013). Experiments with DBpedia, WordNet and SentiWordNet as resources for sentiment analysis in micro-blogging. In *Second Joint Conference on Lexical and Computational Semantics (∗SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013): 2* (pp. 455–459). 2013 .

[51] Han, Y., & Kim, K. K. (2017, June). Sentiment analysis on social media using morphological sentence pattern model. In *2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA)* (pp. 79-84). IEEE.

[52] Haque, M. (2014). Sentiment analysis by using fuzzy logic. *arXiv preprint arXiv:1403.3185*.

[53] Hellendoorn, H., & Thomas, C. (1993). Defuzzification in fuzzy controllers. *Journal of Intelligent & Fuzzy Systems*, *1*(2), 109-123.

[54] Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 8, No. 1).

[55] Hossain, M. S., & Muhammad, G. (2019). Emotion recognition using deep learning approach from audio–visual emotional big data. *Information Fusion*, *49*, 69-78.

[56] Houjeij, A., Hamieh, L., Mehdi, N., & Hajj, H. (2012, April). A novel approach for emotion classification based on fusion of text and speech. In *2012 19th International Conference on Telecommunications (ICT)* (pp. 1-6). IEEE.

[57] Iliou, T., & Anagnostopoulos, C. N. (2009, September). Comparison of different classifiers for emotion recognition. In *2009 13th Panhellenic Conference on Informatics* (pp. 102-106). IEEE.

[58] Ishibuchi, H. , & Tomoharu, N. (2001). Effect of rule weights in fuzzy rule-based classification systems. *IEEE Transactions on Fuzzy Systems, 9* (4 (2001)), 506–515 . Ishibuchi, H. , & Yamamoto, T. (2005). Rule weight specification in fuzzy rule-based classification systems. *IEEE Transactions on Fuzzy Systems, 13* (4 (2005)), 428–435.

[59] Jain, A. P. , & Katkar, V. D. (2015). Sentiments analysis of Twitter data using data mining. In *Information Processing (ICIP), 2015 International Conference on* (pp. 807–810). IEEE. 2015.

[60] Jain, R., Sawhney, R., & Mathur, P. (2018, March). Feature selection for cryotherapy and immunotherapy treatment methods based on gravitational search algorithm. In *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)* (pp. 1-7). IEEE.

[61] Jackson, P., & Haq, S. (2014). Surrey audio-visual expressed emotion (savee) database. *University of Surrey: Guildford, UK*.

[62] Jamatia, A., Swamy, S. D., Gambäck, B., Das, A., & Debbarma, S. (2020). Deep learning based sentiment analysis in a code-mixed English-Hindi and English-Bengali social media corpus. *International journal on artificial intelligence tools*, *29*(05), 2050014.

[63] Jang, J. S. R., Sun, C. T., & Mizutani, E. (1997). Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence [Book Review]. *IEEE Transactions on automatic control*, *42*(10), 1482-1484.

[64] Jefferson, C., Liu, H., & Cocea, M. (2017, July). Fuzzy approach for sentiment analysis. In *2017 IEEE international conference on fuzzy systems (FUZZ-IEEE)* (pp. 1-6). IEEE.

[65] Joachims, T. (1996). *A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization*. Carnegie-mellon univ pittsburgh pa dept of computer science.

[66] Johnson, R., & Zhang, T. (2016, June). Supervised and semi-supervised text categorization using LSTM for region embeddings. In *International Conference on Machine Learning* (pp. 526-534). PMLR.

[67] Joshi, S., Mehta, S., Mestry, P., & Save, A. (2016, March). A new approach to target dependent sentiment analysis with onto-fuzzy logic. In *2016 IEEE International Conference on Engineering and Technology (ICETECH)* (pp. 730-735). IEEE.

[68] Jusoh, S., & Alfawareh, H. M. (2013, January). Applying fuzzy sets for opinion mining. In *2013 International Conference on Computer Applications Technology (ICCAT)* (pp. 1-5). IEEE.

[69] Kamil A, Rustamov S, Clements MA, Mustafayev E (2018). Adaptive neuro-fuzzy inference system for classification of texts. In Recent Developments and the New Direction in Soft-Computing Foundations and Applications, pp. 63-70. Springer, Cham, 2018.

[70] Kanclerz, K., Miłkowski, P., & Kocoń, J. (2020). Cross-lingual deep neural transfer learning in sentiment analysis. Procedia Computer Science, 176, 128-137.

[71] Katta P, Hegde NP (2019). A Hybrid Adaptive Neuro-Fuzzy Interface and Support Vector Machine Based Sentiment Analysis on Political Twitter Data. International Journal of Intelligent Engineering and Systems 12, no. 1 (2019): 165-173.

[72] Kennedy, A., & Inkpen, D. (2006). Sentiment classification of movie reviews using contextual valence shifters. *Computational intelligence*, *22*(2), 110-125.

[73] Khan, F. H., Bashir, S., & Qamar, U. (2014). TOM: Twitter opinion mining framework using hybrid classification scheme. *Decision support systems*, 57, 245-257.

[74] Krishna, B. V., Pandey, A. K., & Kumar, A. S. (2018). Feature based opinion mining and sentiment analysis using fuzzy logic. *In Cognitive science and artificial intelligence* (pp. 79-89). Springer, Singapore.

[75] Kulkarni S, Nikumbh PJ, Anuradha G, Nikam S (2012). S-ANFIS: Sentiment aware adaptive network-based fuzzy inference system for Predicting Sales Performance using Blogs/Reviews. International Journal of Multidisciplinary in Cryptology and Information Security 1, no. 2 (2012).

[76] Kulkarni S, Vidhate A, Anuradha G. Neuro Fuzzy Approach for Predicting Sales Performance of Movies Considering Sentiments in Online Reviews.

[77] Kumar, A., & Singh, J. P. (2019). Demonetization in India: Good or Bad in Context of Social Media. *Available at SSRN 3349021*.

[78] Kumar, H. M., Harish, B. S., & Darshan, H. K. (2019). Sentiment Analysis on IMDb Movie Reviews Using Hybrid Feature Extraction Method. *International Journal of Interactive Multimedia & Artificial Intelligence*, *5*(5).

[79] L.A. Zadeh (1975). "Calculus of Fuzzy Restrictions." , Fuzzy Sets and their Applications to cognitive and decision processes, pages 1-40,1975.

[80] Lal, Y. K., Kumar, V., Dhar, M., Shrivastava, M., & Koehn, P. (2019, July). De-mixing sentiment from code-mixed text. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop* (pp. 371-377).

[81] Li, J., Fong, S., Zhuang, Y., & Khoury, R. (2014, September). Hierarchical classification in Text mining for Sentiment Analysis. *In 2014 International Conference on Soft Computing and Machine Intelligence* (pp. 46-51). IEEE.

[82] Li, X., Xie, H., Chen, L., Wang, J., & Deng, X. (2014). News impact on stock price return via sentiment analysis. *Knowledge-Based Systems*, 69, 14-23.

[83] Li, Xiaodong, Haoran Xie, Li Chen, Jianping Wang, and Xiaotie Deng (2014). "News impact on stock price return via sentiment analysis." *Knowledge-Based Systems* 69 (2014): 14-23.

[84] Lian, Z., Li, Y., Tao, J., & Huang, J. (2018). Investigation of multimodal features, classifiers and fusion methods for emotion recognition. *arXiv preprint arXiv:1809.06225*.

[85] Likitha, M. S., Gupta, S. R. R., Hasitha, K., & Raju, A. U. (2017, March). Speech based human emotion recognition using MFCC. In *2017 international conference on wireless communications, signal processing and networking (WiSPNET)* (pp. 2257-2260). IEEE.

[86] Liu, G., Huang, X., Liu, X., & Yang, A. (2020). A novel aspect-based sentiment analysis network model based on multilingual hierarchy in online social network. The Computer Journal, 63(3), 410-424.

[87] Livingstone, S. R., & Russo, F. A. (2018). The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PloS one*, *13*(5), e0196391.

[88] Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.

[89] Liu, H., & Cocea, M. (2017, February). Fuzzy rule based systems for interpretable sentiment analysis. In *2017 Ninth International Conference on Advanced Computational Intelligence (ICACI)* (pp. 129-136). IEEE.

[90] Liu, K. L. , Li, W. J. , & Guo, M. (2012, July). Emoticon smoothed language models for twitter sentiment analysis. *Twenty-sixth aAAI conference on artificial intelligence* .

[91] Liu, S. , Cheng, X. , Li, F. , & Li, F. (2015). TASC: Topic-adaptive sentiment classification on dynamic tweets. *IEEE Transactions on Knowledge and Data Engineering, 27* (6 (2015)), 1696–1709 .

[92] Liu, Y., Bi, J. W., & Fan, Z. P. (2017). Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory. *Information Fusion*, 36, 149-161.

[93] López, V. , Río, S. D. , Benítez, J. M. , & Herrera, F. (2015). Cost-sensitive linguistic fuzzy rule based classification systems under the MapReduce framework for imbalanced big data. *Fuzzy Sets and Systems, 258* (2015), 5–38.

[94] Lucas, G. M. , Gratch, J. , Malandrakis, N. , Szablowski, E. , Fessler, E. , & Nichols, J. (2017). GOAALLL!: Using sentiment in the world cup to explore theories of emotion. *Image and Vision Computing, 65* (2017), 58–65 .

[95] MacQueen, J. (1967, June). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, No. 14, pp. 281-297).

[96] Majumder, N., Hazarika, D., Gelbukh, A., Cambria, E., & Poria, S. (2018). Multimodal sentiment analysis using hierarchical fusion with context modeling. *Knowledge-based systems*, *161*, 124-133.

[97] Mamdani, E. H. , & Assilian, S. (1975). "An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies, 7* (1), 1–13 1975.

[98] Mamdani, E. H., & Assilian, S. (1993). An experiment in linguistic synthesis with a fuzzy logic. *Readings in Fuzzy Sets for Intelligent Systems Dubois, Morgan Kaufmann Publishers, Inc., Los Altos, CA*, 283-289.

[99] Mäntylä, M. V., Graziotin, D., & Kuutila, M. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review*, 27, 16-32.

[100] Márquez, F. A., Peregrín, A., & Herrera, F. (2007). Cooperative evolutionary learning of linguistic fuzzy rules and parametric aggregation connectors for Mamdani fuzzy systems. *IEEE Transactions on Fuzzy Systems*, *15*(6), 1162-1178.

[101] Martínez-Cámara, E. , Martín-Valdivia, M. T. , Urena-López, L. A. , & Monte- jo-Ráez, A. R. (2014). Sentiment analysis in Twitter. *Natural Language Engineer- ing, 20* (1), 1–28 .
[102] Maas, A., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011, June). Learning word vectors for sentiment analysis. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies* (pp. 142-150).
[103] Matsumoto, S., Takamura, H., & Okumura, M. (2005, May). Sentiment classification using word sub-sequences and dependency sub-trees. In *Pacific-Asia conference on knowledge discovery and data mining* (pp. 301-311). Springer, Berlin, Heidelberg.
[104] Mazumder P, Chowdhury NA, Bhuiya MA, Akash SH, Rahman RM (2018). A Fuzzy Logic Approach to Predict the Popularity of a Presidential Candidate. In Modern Approaches for Intelligent Information and Database Systems, pp. 63-74. Springer, Cham, 2018.
[105] Mejova, Y. (2009)., "Sentiment analysis: An overview,"Comprehensive exam pa- per, available on http://www.cs.uiowa.edu/∼ymejova/publications/CompsYelena Mejova.pdf [2010-02-03
[106] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
[107] Minaee, S., Azimi, E., & Abdolrashidi, A. (2019). Deep-sentiment: Sentiment analysis using ensemble of cnn and bi-lstm models. *arXiv preprint arXiv:1904.04206*.
[108] Ming, Z., Feng, Z., & Zhengbiao, J. (2015). A study on speech emotion recognition based on fuzzy K nearest neighbor. *International Journal of Multimedia and Ubiquitous Engineering*, *10*(10), 57-66.
[109] Montoro, A., Olivas, J. A., Peralta, A., Romero, F. P., & Serrano-Guerrero, J. (2018, July). An ANEW based fuzzy sentiment analysis model. In *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (pp. 1-7). IEEE.
[110] Morency, L. P., Mihalcea, R., & Doshi, P. (2011, November). Towards multimodal sentiment analysis: Harvesting opinions from the web. In *Proceedings of the 13th international conference on multimodal interfaces* (pp. 169-176).
[111] Mostafa, Mohamed M. "Global halal food discourse on social media: a text mining approach." *The Journal of International Communication* 26, no. 2 (2020): 211-237.
[112] Mouthami, K., Devi, K. N., & Bhaskaran, V. M. (2013, February). Sentiment analysis and classification based on textual reviews. In *2013 international conference on Information communication and embedded systems (ICICES)* (pp. 271-276). IEEE.
[113] Mullen, T., & Collier, N. (2004, July). Sentiment analysis using support vector machines with diverse information sources. In *Proceedings of the 2004 conference on empirical methods in natural language processing* (pp. 412-418).
[114] Nadali, Samaneh, Masrah Azrifah Azmi Murad, and Rabiah Abdul Kadir (2010). "Sentiment classification of customer reviews based on fuzzy logic." In *2010 International Symposium on Information Technology*, vol. 2, pp. 1037-1044. IEEE, 2010.
[115] Neethu, M. S. , & Rajasree, R. (2013). "Sentiment analysis in twitter using machine learning techniques. In *Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on* (pp. 1–5). IEEE. 2013.
[116] Neri, F., Aliprandi, C., Capeci, F., Cuadros, M., & By, T. (2012, August). Sentiment analysis on social media. *In 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 919-926). IEEE.
[117] Nielsen, F. Å. (2011). Afinn. *Richard Petersens Plads, Building, 321*.
[118] Ortega R., Fonseca A., Gutierrez Y., Montoyo A.: "Ssa-uo: unsupervised twitter sentiment analysis," In: *Second joint conference on lexical and computational semantics (* SEM)*, vol. 2, 2013, pp. 501–507.
[119] Padmaja K, Hegde NP (2019). Twitter sentiment analysis using adaptive neuro-fuzzy inference system with genetic algorithm. In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), pp. 498-503. IEEE, 2019.
[120] Pang, B., Lee, L., & Vaithyanathan, S. (2002, July). Thumbs up?: sentiment classification using machine learning techniques. *In Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume* 10 (pp. 79-86). Association for Computational Linguistics.
[121] Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *arXiv preprint cs/0409058*.
[122] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2), 1-135.
[123] Park, C. W. , & Seo, D. R. (2018). Sentiment analysis of Twitter corpus related to artificial intelligence assistants. In *2018 5th International Conference on Industrial Engineering and Applications (ICIEA)* (pp. 4 95–4 98). IEEE. 2018.
[124] Park, E., Kang, J., Choi, D., & Han, J. (2020). Understanding customers' hotel revisiting behaviour: a sentiment analysis of online feedback reviews. *Current Issues in Tourism*, *23*(5), 605-611.
[125] Parveen H, Pandey S (2016). Sentiment analysis on Twitter Data-set using Naive Bayes algorithm. In 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), pp. 416-419. IEEE, 2016.
[126] Pedrycz, W., & Gomide, F. (2007). *Fuzzy systems engineering: toward human-centric computing*. John Wiley & Sons.
[127] Peng, T. C., & Shih, C. C. (2010, August). An unsupervised snippet-Based sentiment classification method for chinese unknown phrases without using reference word pairs. In *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology* (Vol. 3, pp. 243-248). IEEE.
[128] Pérez-Rosas, V., Mihalcea, R., & Morency, L. P. (2013, August). Utterance-level multimodal sentiment analysis. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 973-982).
[129] Phan, H. T., Nguyen, N. T., Van Cuong, T., & Hwang, D. (2019, July). A Method for Detecting and Analyzing the Sentiment of Tweets Containing Fuzzy Sentiment Phrases. In *2019 IEEE International Symposium on INnovations in Intelligent SysTems and Applications (INISTA)* (pp. 1-6). IEEE.
[130] Phan, H. T., Tran, V. C., Nguyen, N. T., & Hwang, D. (2020). Improving the Performance of Sentiment Analysis of Tweets Containing Fuzzy Sentiment Using the Feature Ensemble Model. *IEEE Access*, 8, 14630-14641.
[131] Platt, J. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, *10*(3), 61-74.
[132] Polanyi, L., & Zaenen, A. (2006). Contextual valence shifters. In *Computing attitude and affect in text: Theory and applications* (pp. 1-10). Springer, Dordrecht.
[133] Poria, S., Cambria, E., Hazarika, D., Majumder, N., Zadeh, A., & Morency, L. P. (2017, July). Context-dependent sentiment analysis in user-generated videos. In *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 873-883).
[134] Poria, S., Majumder, N., Hazarika, D., Cambria, E., Gelbukh, A., & Hussain, A. (2018). Multimodal sentiment analysis: Addressing key issues and setting up the baselines. *IEEE Intelligent Systems*, *33*(6), 17-25.

[135] Qing, M., & Li, T. R. (2004, March). Some properties and new formulae of fuzzy entropy. In *IEEE International Conference on Networking, Sensing and Control, 2004* (Vol. 1, pp. 401-406). IEEE.
[136] Rahman, L., Mohammed, N., & Al Azad, A. K. (2016, September). A new LSTM model by introducing biological cell state. In *2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)* (pp. 1-6). IEEE.
[137] Ramakrishnan, S., & El Emary, I. M. (2013). Speech emotion recognition approaches in human computer interaction. *Telecommunication Systems*, *52*(3), 1467-1478.
[138] Rao, Y., Li, Q., Mao, X., & Wenyin, L. (2014). Sentiment topic models for social emotion mining. Information Sciences, 266, 90-100.
[139] Rattrout, A., & Ateeq, A. (2019, July). Sentiment Analysis on Arabic Content in Social Media: Hybrid Model of Dictionary Based and Fuzzy Logic. In *Proceedings of the 3rd International Conference on Future Networks and Distributed Systems* (pp. 1-11).
[140] Ravi, K., Ravi, V., & Prasad, P. S. R. K. (2017). Fuzzy formal concept analysis based opinion mining for CRM in financial services. *Applied Soft Computing*, 60, 786-807.
[141] Revathy, A., Shanmugapriya, P., & Mohan, V. (2015, March). Performance comparison of speaker and emotion recognition. In *2015 3rd International Conference on Signal Processing, Communication and Networking (ICSCN)* (pp. 1-6). IEEE.
[142] Ross, T. J. (2004). *Fuzzy logic with engineering applications* (Vol. 2). New York: Wiley.
[143] Roustakiani, A., Abdolvand, N., & Harandi, S. R. (2018). An Improved Sentiment Analysis Algorithm Based on Appraisal Theory and Fuzzy Logic. *Information Systems & Telecommunication*, 88.
[144] Rustamov S (2018). A hybrid system for subjectivity analysis. Advances in Fuzzy Systems 2018 (2018).
[145] Rustamov S, Mustafayev E, Clements M (2013). Sentence-level subjectivity detection using neuro-fuzzy models. In Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pp. 108-114. 2013.
[146] Rustamov S, Mustafayev E, Clements MA (2013). Sentiment analysis using Neuro-Fuzzy and Hidden Markov models of text. In 2013 Proceedings of IEEE Southeastcon, pp. 1-6. IEEE, 2013.
[147] Sun, S., Luo, C., & Chen, J. (2017). A review of natural language processing techniques for opinion mining systems. *Information fusion*, *36*, 10-25.
[148] Sahu, G. (2019). Multimodal speech emotion recognition and ambiguity resolution. *arXiv preprint arXiv:1904.06022*.
[149] Saleena, N. (2018). An ensemble classification system for Twitter sentiment analysis. *Procedia Computer Science, 132* (2018), 937–946 .
[150] Salehan, M., & Kim, D. J. (2016). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems*, 81, 30-40.
[151] Sanz, J. A . , Fernandez, A . , Bustince, H. , & Herrera, F. (2013). IVTURS: A linguistic fuzzy rule-based classification system based on a new interval-valued fuzzy rea- soning method with tuning and rule selection. *IEEE Transactions on Fuzzy Systems, 21* (3 (2013)), 399–411 .
[152] Sasikala, P., & Sheela, L. M. I. (2020). Sentiment analysis of online product reviews using DLMNN and future prediction of online product using IANFIS. *Journal of Big Data*, 7(1), 1-20.
[153] Sawhney, R., Mathur, P., & Shankar, R. (2018, May). A firefly algorithm based wrapper-penalty feature selection method for cancer diagnosis. In *International Conference on Computational Science and Its Applications* (pp. 438-449). Springer, Cham.
[154] Sawhney, R., Shah, R. R., Bhatia, V., Lin, C. T., Aggarwal, S., & Prasad, M. (2019, June). Exploring the impact of evolutionary computing based feature selection in suicidal ideation detection. In *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (pp. 1-6). IEEE.
[155] Scherer, S., Kane, J., Gobl, C., & Schwenker, F. (2013). Investigating fuzzy-input fuzzy-output support vector machines for robust voice quality classification. *Computer Speech & Language*, *27*(1), 263-287.
[156] Shah, A. F., & Anto, P. B. (2017, March). Hybrid spectral features for speech emotion recognition. In *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)* (pp. 1-4). IEEE.
[157] Shakeel, M., & Karwal, V. (2016, December). Lexicon-based sentiment analysis of Indian Union Budget 2016–17. In *2016 International Conference on Signal Processing and Communication (ICSC)* (pp. 299-302). IEEE.
[158] Sharmista, A., & Ramaswami, M. SVM and Fuzzy SVM Based Opinion Mining In Tamil Using R.
[159] Shi, Y., & Song, W. (2010, August). Speech emotion recognition based on data mining technology. In *2010 Sixth International Conference on Natural Computation* (Vol. 2, pp. 615-619). IEEE.
[160] Shivaprasad, T. K., & Shetty, J. (2017, March). Sentiment analysis of product reviews: a review. In *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)* (pp. 298-301). IEEE.
[161] Siddiqua, U. A., Ahsan, T., & Chy, A. N. (2016, October). Combining a rule-based classifier with weakly supervised learning for twitter sentiment analysis. In *2016 International Conference on Innovations in Science, Engineering and Technology (ICISET)* (pp. 1-4). IEEE.
[162] Siganos, A., Vagenas-Nanos, E., & Verwijmeren, P. (2014). Facebook's daily sentiment and international stock markets. *Journal of Economic Behavior & Organization*, 107, 730-743.
[163] Singh, P., Sawhney, R. S., & Kahlon, K. S. (2018). Sentiment analysis of demonetization of 500 & 1000 rupee banknotes by Indian government. *ICT Express*, 4(3), 124-129.
[164] Sinha, P. P., Mishra, R., Sawhney, R., Mahata, D., Shah, R. R., & Liu, H. (2019, November). # suicidal-A multipronged approach to identify and explore suicidal ideation in twitter. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management* (pp. 941-950).
[165] Sinith, M. S., Aswathi, E., Deepa, T. M., Shameema, C. P., & Rajan, S. (2015, December). Emotion recognition from audio signals using Support Vector Machine. In *2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS)* (pp. 139-144). IEEE.
[166] Sirbu, D., Secui, A., Dascalu, M., Crossley, S. A., Ruseti, S., & Trausan-Matu, S. (2016, September). Extracting Gamers' Opinions from Reviews. In *2016 18th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)* (pp. 227-232). IEEE.
[167] Songpan, Wararat (2017). "The analysis and prediction of customer review rating using opinion mining." In *2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA)*, pp. 71-77. IEEE, 2017.
[168] Soleymani, M., Garcia, D., Jou, B., Schuller, B., Chang, S. F., & Pantic, M. (2017). A survey of multimodal sentiment analysis. *Image and Vision Computing*, *65*, 3-14.
[169] Srivastava, R., & Bhatia, M. P. S. (2013, August). Quantifying modified opinion strength: A fuzzy inference system for sentiment analysis. In *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 1512-1519). IEEE.

[170] Sun, L., Fu, S., & Wang, F. (2019). Decision tree SVM model with Fisher feature selection for speech emotion recognition. *EURASIP Journal on Audio, Speech, and Music Processing*, *2019*(1), 1-14.
[171] Suresh, H. (2016, October). An unsupervised fuzzy clustering method for twitter sentiment analysis. In *2016 International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)* (pp. 80-85). IEEE.
[172] Susan, S., & Hanmandlu, M. (2013, July). A novel Fuzzy Entropy based on the Non-Extensive entropy and its application for feature selection. In *2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (pp. 1-8). IEEE.
[173] Susan S, Khowal SK, Kumar A, Kumar A, Yadav AS (2013). Fuzzy min-max neural networks for business intelligence. In 2013 International Symposium on Computational and Business Intelligence, pp. 115-118. IEEE, 2013.
[174] Susan, S. , & Keshari, J. (2019). Finding significant keywords for document databases by two-phase maximum entropy partitioning. *Pattern Recognition Letters* .
[175] Susan, S., & Sharma, S. (2012, November). A fuzzy nearest neighbor classifier for speaker identification. In *2012 Fourth International Conference on Computational Intelligence and Communication Networks* (pp. 842-845). IEEE.
[176] Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *IEEE transactions on systems, man, and cybernetics*, (1), 116-132.
[177] Ton-That, A. H., & Cao, N. T. (2019). Speech emotion recognition using a fuzzy approach. *Journal of Intelligent & Fuzzy Systems*, *36*(2), 1587-1597.
[178] Tripathy, A., Agrawal, A., & Rath, S. K. (2016). Classification of sentiment reviews using n-gram machine learning approach. *Expert Systems with Applications*, *57*, 117-126.
[179] Trupthi, M., Pabboju, S., & Narasimha, G. (2017, January). Sentiment analysis on twitter using streaming API. In *2017 IEEE 7th International Advance Computing Conference (IACC)* (pp. 915-919). IEEE.
[180] Turney, P. D. (2002, July). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. *In Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 417-424). Association for Computational Linguistics.
[181] Valdivia A, Luzíón MV, Herrera F (2017). Neutrality in the sentiment analysis problem based on fuzzy majority. In 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1-6. IEEE, 2017.
[182] Valdivia A, Luzón MV, Cambria E, Herrera F (2018). Consensus vote models for detecting and filtering neutrality in sentiment analysis. Information Fusion 44 (2018): 126-135.
[183] Valsamidis, S., Theodosiou, T., Kazanidis, I., & Nikolaidis, M. (2013). A framework for opinion mining in blogs for agriculture. *Procedia Technology*, *8*, 264-274.
[184] Vapnik, V., Guyon, I., & Hastie, T. (1995). Support vector machines. *Mach. Learn*, *20*(3), 273-297.
[185] Vashishtha, S., & Susan, S. (2018, December). Fuzzy logic based dynamic plotting of mood swings from tweets. In *International Conference on Innovations in Bio-Inspired Computing and Applications* (pp. 129-139). Springer, Cham.
[186] Vashishtha, S., & Susan, S. (2019 a). Fuzzy rule based unsupervised sentiment analysis from social media posts. *Expert Systems with Applications*, *138*, 112834.
[187] Vashishtha, S., & Susan, S. (2019 b). Sentiment Cognition from Words Shortlisted by Fuzzy Entropy. *IEEE Transactions on Cognitive and Developmental Systems*, *12*(3), 541-550.
[188] Vashishtha, S., & Susan, S. (2020 a). Inferring Sentiments from Supervised Classification of Text and Speech cues using Fuzzy Rules. *Procedia Computer Science*, *167*, 1370-1379.
[189] Vashishtha, S., & Susan, S. (2020 b). Fuzzy Interpretation of Word Polarity Scores for Unsupervised Sentiment Analysis. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1-6). IEEE.
[190] Vashishtha, S., & Susan, S. (2020 c). Unsupervised fuzzy inference system for speech emotion recognition using audio and text cues (workshop paper). In *2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM)* (pp. 394-403). IEEE.
[191] Vashishtha, S., & Susan, S. (2021 a). Highlighting keyphrases using senti-scoring and fuzzy entropy for unsupervised sentiment analysis. *Expert Systems with Applications*, *169*, 114323.
[192] Vashishtha, S., & Susan, S. (2021 b). Neuro-fuzzy network incorporating multiple lexicons for social sentiment analysis. *Soft Computing*, 1-21.
[193] Vu, P. M., Pham, H. V., Nguyen, T. T., & Nguyen, T. T. (2016, August). Phrase-based extraction of user opinions in mobile app reviews. *In Proceedings of the 31st IEEE/ACM International Conference on Automated Software Engineering* (pp. 726-731).
[194] Wang, B., Huang, Y., Wu, X., & Li, X. (2015). A fuzzy computing model for identifying polarity of chinese sentiment words. *Computational intelligence and neuroscience, 2015*.
[195] Wang, S., & Manning, C. D. (2012, July). Baselines and bigrams: Simple, good sentiment and topic classification. *In Proceedings of the 50th annual meeting of the association for computational linguistics:* Short papers-volume 2 (pp. 90-94). Association for Computational Linguistics.
[196] Wang, W., Chen, L., Thirunarayan, K., & Sheth, A. P. (2012, September). Harnessing twitter" big data" for automatic emotion identification. In *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing* (pp. 587-592). IEEE.
[197] Wang, X., Zhang, H., & Xu, Z. (2016 a). Public sentiments analysis based on fuzzy logic for text. *International Journal of Software Engineering and Knowledge Engineering*, 26(09n10), 1341-1360.
[198] Wang, J., Yu, L. C., Lai, K. R., & Zhang, X. (2019, November). Investigating dynamic routing in tree-structured LSTM for sentiment analysis. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 3432-3437).
[199] Wilson, T., Wiebe, J., & Hoffmann, P. (2005, October). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing* (pp. 347-354).
[200] Windasari, I. P. , Uzzi, F. N. , & Satoto, K. I. (2017). Sentiment analysis on Twit- ter posts: An analysis of positive or negative opinion on GoJek. In *Information Technology, Computer, and Electrical Engineering (ICITACEE), 2017 4th International Conference on* (pp. 266–269). IEEE. 2017 .
[201] Wöllmer, M., Weninger, F., Knaup, T., Schuller, B., Sun, C., Sagae, K., & Morency, L. P. (2013). Youtube movie reviews: Sentiment analysis in an audio-visual context. *IEEE Intelligent Systems*, *28*(3), 46-53.
[202] Wu, T. F., Lin, C. J., & Weng, R. C. (2004). Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research*, *5*(Aug), 975-1005.
[203] Xi, D., Zhuang, F., Zhou, G., Cheng, X., Lin, F., & He, Q. (2020, April). Domain adaptation with category attention network for deep sentiment analysis. In *Proceedings of The Web Conference 2020* (pp. 3133-3139).

[204] Yadav, P., & Pandya, D. (2017, February). SentiReview: Sentiment analysis based on text and emoticons. In *2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)* (pp. 467-472). IEEE.
[205] Yoon, S., Byun, S., & Jung, K. (2018, December). Multimodal speech emotion recognition using audio and text. In *2018 IEEE Spoken Language Technology Workshop (SLT)* (pp. 112-118). IEEE.
[206] Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning-III. *Information sciences*, *9*(1), 43-80.
[207] Zadeh, L. A. (1996). Fuzzy logic = computing with words. *IEEE Transactions on Fuzzy Systems, 4* (2), 103–111 May, 1996
[208] Zadeh, L. A. (2015). Fuzzy logic: A personal perspective. *Fuzzy Sets and Systems, vol.281* , 4–20 Dec., 2015 .
[209] Zadeh, A., Zellers, R., Pincus, E., & Morency, L. P. (2016). Mosi: multimodal corpus of sentiment intensity and subjectivity analysis in online opinion videos. *arXiv preprint arXiv:1606.06259*.
[210] Zadeh, A., Chen, M., Poria, S., Cambria, E., & Morency, L. P. (2017). Tensor fusion network for multimodal sentiment analysis. *arXiv preprint arXiv:1707.07250*.
[211] Zhang, Z., & Lim, J. S. (2015, November). Emotion Recognition Algorithm Based on Neural Fuzzy Network and the Cloud Technology. In *2015 10th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA)* (pp. 576-579). IEEE.
[212] Zhao, C., Wang, S., & Li, D. (2014, December). Fuzzy sentiment membership determining for sentiment classification. In *2014 IEEE International Conference on Data Mining Workshop* (pp. 1191-1198). IEEE.
[213] Zhao, R., & Mao, K. (2017). Fuzzy bag-of-words model for document representation. *IEEE Transactions on Fuzzy Systems,* 26(2), 794-804.
[214] Zhao, J., Mao, X., & Chen, L. (2019). Speech emotion recognition using deep 1D & 2D CNN LSTM networks. *Biomedical Signal Processing and Control*, *47*, 312-323.
[215] Žunić E, Djedović A, Avdagić Z (2016). Decission support system for candidates classification in the employment process based on ANFIS method. In 2016 XI International Symposium on Telecommunications (BIHTEL), pp. 1-6. IEEE, 2016.

BRIEF BIO-DATA

Name: Ms. Srishti Vashishtha

Date of Birth: 22-08-1991

Place of Residence: New Delhi

Email: srishtidtu@gmail.com

Educational Qualification:
B. Tech in Information Technology from Guru Gobind Singh Indraprastha University, Delhi in 2013 with 77.57% marks

M. Tech. in Computer Science and Engineering from Guru Gobind Singh Indraprastha University, Delhi in 2015 with 86.4%

Joined Ph.D in Information Technology Department at Delhi Technology University Delhi in July 2017.

Cleared GATE and NET examinations.

Received DTU Research Commendable Research Excellence Award 3 times

Research areas of interest:

Fuzzy logic, Data Mining, Machine Learning and Natural language Processing.