Neuro-fuzzy network incorporating multiple lexicons for social sentiment analysis

Srishti Vashishtha¹ · Seba Susan¹

Accepted: 2 November 2021 / Published online: 29 November 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

We have proposed MultiLexANFIS which is an adaptive neuro-fuzzy inference system (ANFIS) that incorporates inputs from multiple lexicons to perform sentiment analysis of social media posts. We classify tweets into two classes: neutral and non-neutral; the latter class includes both positive and negative polarity. This type of classification will be considered for applications that aim to test the neutrality of content posted by the users in social media platforms. In our proposed model, features are extracted by integrating natural language processing with fuzzy logic; hence, it is able to deal with the fuzziness of natural language in a very efficient and automatic manner. We have proposed a novel set of 64 rules for the proposed neuro-fuzzy network that can classify tweets correctly by working on fuzzy features fetched from VADER, AFINN and SentiWordNet lexicons. The proposed novel rules are domain independent, i.e., we can extend these rules for any textual data that employs lexicons. The antecedent and consequent parameters of the ANFIS are optimized by gradient descent and least squares estimate algorithms, respectively, in an iterative manner. The key contributions of this paper are: (1) a novel neuro-fuzzy system: MultiLexANFIS that takes as its input the positive and negative sentiment scores of tweets computed from multiple lexicons—VADER, AFINN and SentiWordNet, in order to classify the tweets into neutral and non-neutral content, (2) a novel set of 64 rules for the Sugeno-type fuzzy inference system—MultiLexANFIS, (3) single-lexicon-based ANFIS variants to classify tweets when multiple lexicons are not available and (4) comparison of MultiLexANFIS with different fuzzy, non-fuzzy and deep learning state of the art on various benchmark datasets revealing the superiority of our proposed neuro-fuzzy system for social sentiment analysis.

Keywords ANFIS · Tweets · Lexicon · Neuro-fuzzy network · Sentiment analysis · Social media

1 Introduction

Social sentiment analysis has a huge impact on daily life, in modern times, due to the widespread use of social media for expressing public opinion. People prefer to communicate on the online social network such as Twitter, Facebook, WhatsApp, Instagram and YouTube; among them, Twitter is the most popular microblogging forum for expressing public sentiment. Sentiment analysis (SA) is used to inspect the polarity or intention of tweets about any

entity. The content of tweets could be movie reviews (Vashishtha and Susan 2021; Parveen and Pandey 2016; Rahman et al. 2016; Anagha et al. 2015; Kulkarni et al. 2014; Ghiassi et al. 2013), product reviews (Ghiassi et al. 2013; Indhuja and Reghu 2014), stock market trends and political events (Khan et al. 2020), political views (Katta and Hegde 2019), COVID-19 (Nemes and Kiss 2021) or about any other topic that attracts public opinion.

Tweets can be classified as positive, negative or neutral using either supervised or unsupervised machine learning algorithms. Generally, the twitter data contains a major portion of neutral tweets compared to positive and negative tweets. It has been proved that neutral samples are vital for learning polarity in order to obtain remarkable classification results (Koppel and Schler 2005, 2006). In this research work, classification of tweets is not the stereotype—positive, negative or neutral; rather, it is neutral and

✉ Srishti Vashishtha
srishtidtu@gmail.com

Seba Susan
seba_406@yahoo.in

¹ Delhi Technological University, Shahbad Daultpur, Main Bawana Road, Delhi 110042, India

non-neutral, and the latter class includes both positive and negative polarity. The purpose of this type of classification will be beneficial for such systems that check neutral content posted by the users. There are many intelligence agencies that aim to track an individual's social media activities for national security reasons, in search of anti-nationalists or terrorist activities, etc. Social media is widely used by people, and they tend to post, like and share content of their interest on a regular basis. These agencies analyze the content posted by individuals to determine whether the content is neutral or biased toward any political event or leader, or generally indicative of any malicious intent. Hence, such tasks need to deploy SA on online social network in order to classify the social media posts into neutral and non-neutral category related to an individual or group. To the best of our knowledge, none of the previous works on SA have concentrated on solely classifying tweets into neutral and non-neutral and none of them have targeted to build models that would aid security applications based on sentiment scores. Also, mostly the authors use one lexicon at a time and have not simultaneously incorporated multiple lexicons into their learning framework as proposed in our work. Two types of ANFIS variants—single lexicon and multiple lexicons, are presented in this work. The architecture of the proposed MultiLexANFIS model consists of fuzzy-inspired feature extraction phase where the input features are fetched from multiple lexicons—VADER (Hutto and Gilbert 2014), AFINN (Nielsen 2011) and SentiWordNet (Baccianella et al. 2010), and fuzzy inference phase is constructed using a novel set of 64 rules devised specifically for these three lexicons. The rules can be easily extended for adaptation to a greater number of lexicons.

The major contributions of this research are: (1) development of a novel neuro-fuzzy system: MultiLexANFIS that integrates sentiment scores from multiple lexicons to classify tweets into neutral and non-neutral. These sentiment scores are computed from pre-processed textual data and are interpreted as a fuzzy membership values. (2) A novel set of 64 rules for Sugeno-type fuzzy inference system (FIS)—MultiLexANFIS, is constructed. Each lexicon provides two sentiment scores—positive and negative, so a total of six sentiment scores are derived from the VADER, AFINN and SentiWordNet lexicons for each tweet. Each of these six inputs is assigned two membership functions (MFs): *Low* and *High*. (3) We have also constructed single-lexicon-based ANFIS variants with nine rules each to classify tweets. Three variants that we call as VADER, AFINN and SentiWordNet-ANFIS consider the two sentiment scores of positive and negative as input at a time. Each of these inputs is assigned three membership functions (MFs): *Low*, *Medium* and *High*. These variants can be deployed when multiple lexicons are not available.

The experiments are conducted on MultiLexANFIS and all three single-lexicon ANFIS variants; the results reveal that MultiLexANFIS performs better than all single-lexicon variants. (4) A comparative analysis of proposed MultiLexANFIS with fuzzy approaches and non-fuzzy approaches including deep learning models on various benchmark datasets reveals the supremacy of our proposed neuro-fuzzy system. The performance of all models and the state of the art is measured in terms of RMSE (root mean square error).

This paper is organized as follows. Section 2 describes several recent works related to ANFIS and sentiment analysis. The proposed MultiLexANFIS is explained in Sect. 3. The experimental setup and implementation details are discussed in Sect. 4. The results are demonstrated in Sect. 5. Finally, Sect. 6 draws the overall conclusion.

2 Related work

Sentiment analysis is a process that is involved in the determination and classification of opinions or feelings expressed in text, through the use of computing machines. The major challenge is to extract information from unstructured data. Most methodologies in this area focus on polarity or opinion classification into positive or negative classes (Siddiqua et al. 2016; Vashishtha and Susan 2020b; Jefferson et al. 2017; Vashishtha and Susan 2020a, b); some researchers also include a neutral class (Parveen and Pandey 2016; Anagha et al. 2015; Indhuja and Reghu 2014; Vashishtha and Susan 2018, 2019a). Others detect neutral opinions and filter out them to enhance binary polarity classification (Valdivia et al. 2017, 2018). Sentiment analysis models utilizing part-of-speech tags and n-gram feature extraction are gaining popularity. These models deploy either unsupervised algorithms (Vashishtha and Susan 2021) or supervised machine learning algorithms involving SVM, naïve Bayes, etc. (Tripathy et al. 2016). One notable supervised technique is the combination of hidden Markov model for feature extraction and SVM and ANN for the classification process (Kalarani and Brunda 2019). Machine learning classifier models have proved to perform well for sentiment classification, and some of the popular algorithms are naïve Bayes (Parveen and Pandey 2016), support vector machines (SVM) (Katta and Hegde 2019), (Valdivia et al. 2017, 2018), long short-term memory (LSTM) (Rahman et al. 2016; Vashishtha and Susan 2019), recurrent neural network (RNN) (Nemes and Kiss 2021), maximum entropy with probabilistic latent semantic analysis (Xie et al. 2019) and ANFIS. Out of all these options, neuro-fuzzy networks perform generally well in classification tasks involving ambiguous and

uncertain systems since they combine learning with reasoning which is a human trait (Susan et al. 2013).

2.1 Fuzzy logic-based sentiment analysis

In recent years, several fuzzy approaches have started to emerge for text processing. Fuzzy logic is a concept used in computing based on the “degree of truth” instead of true or false labels (0 or 1). It uses fuzzy sets, where each element belongs to the set with a certain degree; this degree is determined by a particular membership function. It has been meticulously exploited as a methodology useful both for representing information associated with sentiment and for executing various sentiment analysis tasks. The exploitation of fuzzy concepts in the field of SA can be incorporated in feature extraction, in fuzzy-based model architecture or in both (Serrano-Guerrero et al. 2021). An unsupervised fuzzy logic-based method for sentiment classification of textual reviews was proposed in Vashishtha and Susan (2020b). The authors formulated fuzzy cardinality, as the measure, from fuzzy sets for the evaluation of word polarity scores, and the proposed model has two versions based on the sentiment lexicon utilized in the model. Many research works have implemented fuzzy membership functions (MFs): triangular (Vashishtha and Susan 2019a), trapezoidal (Liu and Cocca 2017; Jefferson et al. 2017), Gaussian and generalized bell MFs (Gupta et al. 2014; Talpur et al. 2017). These MFs for a basic atomic term can be modified using linguistic Hedges, which are adverbs or adjectives. Linguistic Hedges with fuzzy logic have been successfully used to compute the sentiment of text (Vashishtha and Susan 2018, 2021; Khattak et al. 2020). A combination of fuzzy entropy, k-means clustering and sentiment lexicon shortlists words that help in sentiment cognition in Vashishtha and Susan (2019b). In the field of social media SA, there is also a shift toward multimodal social web, where users post their opinion in the form of text in Facebook and Twitter, audio or video clips in YouTube. The social media reviews are either classified using only fuzzy logic-based text classification (Wu et al. 2017) or can be classified by multimodal sentiment classification using a fusion of fuzzy logic with audio and text features (Vashishtha and Susan 2020a). Sentiment classification can be deployed using rule-based systems, a special type of expert system which uses a set of pre-defined rules for the classification task. Rule-based classifier can be supervised (Siddiqua et al. 2016) or may use fuzzy rules constructed in an unsupervised manner (Anagha et al. 2015; Liu and Cocca 2017; Jefferson et al. 2017; Vashishtha and Susan 2019a; Talpur et al. 2017). The rules can be further applied to supervised machine learning algorithms: naïve Bayes (Siddiqua et al. 2016) and SVM (Katta and Hegde 2019) and can also be learned in

ANFIS (Jang 1993). Ahuja et al. (2015) implemented a SA model on public opinion data from Twitter to find a relationship between the stock market and public mood, using a self-organizing fuzzy neural network. An optimized fuzzy technique—shuffled frog leaping algorithm (SFLA), developed by Madhusudhanan et al. (2019) is deployed for improving the SA process. The SFLA is an optimization technique which imitates the mimetic evolutionary activity of frogs searching for a place with most quantity of food; along with SFLA, a local search algorithm is implemented for performing SA on book reviews. Phan et al. (2019) proposed a method that detects polarity of only specific kind of tweets that contain fuzzy sentiment phrases (FSPs). The first step of this method is to detect FSPs and 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 multi-layer perceptron model. In a recent research work (Yakubu et al. 2021), customer satisfaction model based on online reviews has been developed, based on the sentiment scores derived by applying a novel multigene genetic programming-based fuzzy regression (MGFP-FR) approach. To address the fuzziness of customer opinions, the sentiment scores are transformed into asymmetrical fuzzy numbers. Then, the defuzzified sentiment scores are applied to MGFP; further, Tanaka’s fuzzy regression is employed to determine whether the customer is satisfied or not.

2.2 Adaptive neuro-fuzzy inference system (ANFIS)

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 has been applied for classification (Žunić et al. 2016), prediction (Cosma and Acampora 2016; Kulkarni et al. 2012, 2014; Mazumder et al. 2018), subjectivity analysis (Rustamov et al. 2013a; b; Rustamov 2018; Kamil et al. 2018; Padmaja and Hegde 2019) and SA (Katta and Hegde 2019), (Mazumder et al. 2018), (Kamil et al. 2018; Padmaja and Hegde 2019; Acampora and Cosma 2014; Çakıt et al. 2019). Kulkarni et al. (2012) developed an ANFIS based on sentiment information from movie blogs and reviews, for predicting sales performance in movie domain. The proposed predictive model is built to incorporate the auto-regressive method and ANFIS technique to handle the fuzzy demand with incomplete information. Another work by Kulkarni et al. (2012) proposed neuro-fuzzy methodology—ANFIS, for predicting sales performance of movies by analyzing sentiments in online reviews. The ANFIS model here has three inputs: review

ratings, revenue budget and type of the movies. 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 Hegde 2019), and using principal component analysis with ANFIS to predict political sentiments (Mazumder et al. 2018). Another hybrid classifier of ANFIS with genetic algorithm (GA) has shown good results in twitter SA (Padmaja and Hegde 2019; Acampora and Cosma 2014). GA extracts the important and semantic features from each tweet and optimizes the fuzzy principles in ANFIS. The authors in Asimuzzaman et al. (2017) implemented ANFIS to predict the polarity of Bangla tweets. The ANFIS utilizes semantic rules which are simple yet have a greatly affect the actual polarity of the text. Since the model is designed for classifying tweets into positive and negative polarity, the neutral tweets were ignored in pre-processing phase. Each and every word of Bangla text is translated into English text, so that the polarity value of each word can be extracted from SentiWordNet lexicon. The experiments of this model are conducted on a small twitter samples training set, and it has not been compared with recent state of the art for SA. Kamil et al. (2018) applied ANFIS for executing three different classification problems: (1) sentence-level subjectivity detection, (2) sentiment analysis of texts and (3) detecting user intention in natural language call routing system. In their work, the main goal is to avoid using human annotation or lexical knowledge; the membership degree of each term is calculated by pruned ICF (inverse-class frequency). The fuzzification operations applied involve the computation of the maximum of membership degree with respect to the classes for every term and mean of maxima for all classes. A recent study compared 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 media networks (Çakit et al. 2019). Combination of ANFIS and hidden Markov model has been deployed for SA in (Rustamov et al. 2013a; b) 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) and this motivated us to use these MFs in our proposed MultiLexANFIS. A multi-modal convolutional neuro-fuzzy framework has been developed by Nguyen et al. (2019) that integrates the concepts of fuzzy logic with deep neural network for emotion classification of videos. Convolutional neural network is deployed to extract features from text, audio and visual modalities; these are sent as input to ANFIS that generates new rules to correctly classify emotions. An improved ANFIS (IANFIS) model has been proposed by Sasikala and Sheela (2020) aimed for future prediction of

online products. The authors have also developed a deep learning modified neural network (DLMNN) for SA of online products. A food review dataset being analyzed, initially the dataset is divided into three groups: grade based, content based and collaboration based. For the grade-based data, polarity scores are determined and the review is classified as positive, negative or neutral using DLMNN, ANN and hybrid dragonfly-genetic algorithm (HDF-GA). Social media platform provides abundant and huge amount of data that can be analyzed for different purposes. A hybrid machine learning model: bi-objective optimization and LAN²FIS (logistic adaptive network based on neuro-fuzzy inference system), proposed by Nagamanjula et al. (2020) aims to utilize and handle this vast amount of data for sentiment classification of tweets. The bi-objective optimization implements minimum redundancy and maximum relevancy for selecting optimal set of features from tweets. These features are fed as input to logistic regression-based ANFIS for initially classifying tweets into five classes: positive, negative, strongly positive, strongly negative and neutral; further, the classes are reduced to positive, negative and neutral.

3 Proposed ANFIS architecture for social sentiment analysis

In this paper, we have proposed a novel ANFIS-based system: MultiLexANFIS that integrates inputs from multiple sentiment lexicons to compute the sentiment of social media posts. We have also proposed single-lexicon-based ANFIS variants that require input from one sentiment lexicon at a time. These variants can be deployed when multiple lexicons are not available. The datasets chosen in this research work comprise tweets that are categorized into two classes: neutral and non-neutral. We have merged the positive and negative polarities into one class: non-neutral.

3.1 Motivation

In the last decade, fuzzy systems and their hybrid derivations have proved that they can emulate the typical human reasoning ability in a computationally efficient 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 incorporates the power of artificial neural network and fuzzy logic. Our ANFIS model combines natural language processing (NLP) with fuzzy logic, which boosts the performance of SA. The importance of fuzziness comes into play while dealing with natural language due to the presence of ambiguity in word

senses and the semantic interpretation of sentences. The amalgamation of fuzzy logic concepts into SA helps in dealing with the fuzziness of natural language in a very efficient and automatic manner. Neuro-fuzzy systems have been applied, before, in the field of SA or opinion mining, but these systems were not able to handle the uncertainty and vagueness present in the natural language. Moreover, these systems usually classified the data into positive or negative class; our goal is to classify data into neutral or non-neutral class that would aid organizations to test for the neutrality of content posted by users in social media platforms about persons, events, services, etc. To the best of our knowledge, this novel sentiment classification problem—segregation of neutral and non-neutral classes, has not been explicitly handled by any research work. We have designed two types of ANFIS variants: for single lexicon and multiple lexicons. The single-lexicon variant is based on the type of sentiment lexicon chosen. The experiments have been designed to figure out which lexicon is best suited for ANFIS-based sentiment classification in the field of social SA. Since single-lexicon variant proved to be inconclusive about 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, the MultiLexANFIS architecture has been proposed. This system integrates all the lexicons into a single unit and produces lower error 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 for any textual data that employs lexicons to extract positive and negative score for each textual document or sentence or review.

3.1.1 Motivation behind segregating neutral and non-neutral classes

The motivation behind segregating neutral and non-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 as compared to a training set of the same size that includes only positives and negatives samples (Koppel and Schler 2005, 2006). The use of neutral training samples in learning not only promotes better distinction 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 which deal with visa applications may require the candidates to have neutral behavior for their country; those candidates who are too biased can pose a threat for that country. Such systems can also be used by military services and intelligence agencies for tracking anti-nationalists or terrorist activities.

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

Katta and Hegde (2019) proposed a hybrid ANFIS that integrates nonlinear SVM for SA of political twitter data. In 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 is expected to yield better results because sentiment scores extracted from lexicons would provide better information for sentiment classification. In Katta and Hegde (2019), authors used only one fuzzy MF: generalized bell, while in our proposed model we have explored three fuzzy MFs: triangular (trimf), generalized bell (gbellmf) and Gaussian (gaussmf), in order to select the best performing MF. The authors in Katta and Hegde (2019) have conducted experiments and compared their model's different hybrid ANFIS-SVM variants but not with other state-of-the-art methods, whereas we have conducted experiments and compared 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 takes into consideration the context of the word to determine which is the intended meaning the user is looking for. This process allows to decrease noise and speed up the user's task. 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 Kamil et al. (2018), the membership degree of each term is calculated by pruned ICF (inverse-class frequency), whereas in our model, three popular fuzzy MFs are applied for fuzzification of sentiment scores.

Padmaja and Hegde (2019) developed a hybrid classifier based on ANFIS and genetic algorithm (GA) to classify the tweets into binary sentiment classes—positive or negative class, while our proposed model classifies tweets into neutral or non-neutral sentiment class. The ANFIS classifier integrates the benefits of both artificial neural network and fuzzy logic 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 scores of tweets are computed by considering adjectives, nouns, verbs and adverbs. The inclusion of other part of speech enhances the sentiment evaluation process. Moreover, they have not specified how the words are classified as positive or negative. In our ANFIS model, we have deployed sentiment lexicons to compute the positive and negative scores of tweets. The authors in Padmaja and Hegde (2019) have compared with 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 eight existing SA approaches, and the experiments are conducted on ten benchmark datasets.

3.2 Processing of tweets

In this section, we will discuss 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 ten publicly available datasets: Apple Twitter Dataset (2019), Nuclear Twitter Dataset (2019), Sanders Twitter Dataset (2019), SemEval 2017, SemEval 2016, SemEval 2015, Stanford Twitter Sentiment Test Set (STS-Test) (Go et al. 2009), Airline Twitter Dataset (2019), Twitter_2019 (2019) dataset and Reddit dataset. All the datasets contain three types of tweets: positive, negative and neutral. The class distribution of all datasets is represented in Table 3. We have combined the positive and negative tweets into one class: non-neutral; our goal is to classify the tweets by ANFIS into two classes: neutral and non-neutral. We change the labels of positive and negative tweets to non-neutral; thus, we have now two types of classes: neutral and non-neutral. The labels, 1: neutral and 0: non-neutral, are

the class labels in ANFIS. The various steps in tweet processing are outlined in the following.

3.2.1 Data pre-processing

The first 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 which can lead to incorrect results.

3.2.2 Feature extraction using sentiment lexicons

We have to compute the sentiment of this pre-processed tweet which is required to be in a numerical format so that it can be used as an input to ANFIS. Sentiment lexicons are employed for this purpose, and these lexicons assign scores to either words or the whole tweet. The three sentiment lexicons are VADER (Hutto and Gilbert 2014), AFINN (Nielsen 2011) and SentiWordNet (Baccianella et al. 2010). These sentiment lexicons are popularly used 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 the positive (*TweetPos*) and negative (*TweetNeg*) scores of a tweet as the 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 employs the AFINN method to fetch the score of each word a as represented in Eq. (1). If the score is greater than 0, it is a positive word (Eq. 2). If the score is less than 0, it is a negative word (Eq. 3). The positive score of the tweet (*TweetPos*) is calculated by adding up the scores of all the positive words in the tweet as shown in Eq. (4). The negative score of the tweet (*TweetNeg*) is calculated by adding up the scores of all the negative words in the tweet as shown in Eq. (5).

$$afn = \text{afinn.score}(a) \quad (1)$$

$$\text{if}(afn > 0) \text{then}(af_{\text{Pos}}(a) = afn) \quad (2)$$

$$\text{if}(afn < 0) \text{then}(af_{\text{Neg}}(a) = -afn) \quad (3)$$

$$\text{TweetPos} = \sum_{a=1}^m \text{af}_{\text{Pos}}(a) \quad (4)$$

$$\text{TweetNeg} = \sum_{a=1}^m \text{af}_{\text{Neg}}(a) \quad (5)$$

where a is a word in a tweet and m is the number of selected words.

SentiWordNet lexicon is one of the most popular sentiment lexicons used by researchers. It assigns 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 for this lexicon in the following manner: We remove the stopwords and punctuations, perform lemmatization, part-of-speech (POS) tagging and word sense disambiguation (WSD) by Lesk (Banerjee and Pedersen 2002). WSD is the method of determining which “sense” (meaning) of a word is activated 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 positive (Eq. (8)) and negative score (Eq. (9)) computed using WSD, which can be interpreted as a fuzzy membership belonging to the fuzzy sets Pos (Eq. (6)) and Neg (Eq. (7)). The words, which are having higher positive score than negative score in a tweet, are summed up to evaluate the positive score (TweetPos) of the tweet (Eq. (10)). Similarly, words which are having higher negative score than positive score in a tweet are summed up to evaluate the negative score (TweetNeg) of the tweet (Eq. (11)). These scores are computed for all tweets.

$$\text{Pos} = \{(a, \mu_{\text{Pos}}(a)), a \in X_i\} \quad (6)$$

$$\text{Neg} = \{(a, \mu_{\text{Neg}}(a)), a \in X_i\} \quad (7)$$

$$\mu_{\text{Pos}}(a) = \text{syn.pos_score}() \quad (8)$$

$$\mu_{\text{Neg}}(a) = \text{syn.neg_score}() \quad (9)$$

$$\begin{aligned} &\text{if}(\mu_{\text{Pos}}(a) > 0 \& \mu_{\text{Pos}}(a) > \mu_{\text{Neg}}(a)) \\ &\text{then} \left(\text{TweetPos} = \sum_{a=1}^m \mu_{\text{Pos}}(a) \right) \end{aligned} \quad (10)$$

$$\begin{aligned} &\text{if}(\mu_{\text{Neg}}(a) > 0 \& \mu_{\text{Neg}}(a) > \mu_{\text{Pos}}(a)) \\ &\text{then} \left(\text{TweetNeg} = \sum_{a=1}^m \mu_{\text{Neg}}(a) \right) \end{aligned} \quad (11)$$

where a is the 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 represent the numerical score of that tweet; these scores are given as input in ANFIS.

3.3 ANFIS processing of tweets

ANFIS was introduced by Jang in (1993); it takes crisp inputs and fuzzifies these inputs using membership functions. ANFIS is a method of artificial intelligence that uses a hybrid way of learning by connecting fuzzy logic and neural networks into a single unit. It is based on specific “human features,” because fuzzy logic has a human subjectivity feature, while the neural network has a feature of human way of learning from past experience. ANFIS network is appropriate for solving function problems in various fields. Generally, the problem to be solved 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 k th rule, $k = 1 \dots M$, has the following form:

$$\begin{aligned} &\text{if } x_1 \text{ is } B_1^{(k)}, \text{ and } x_2 \text{ is } B_2^{(k)} \dots x_n \text{ is } B_n^{(k)}, \text{ then } y^{(k)} \\ &= \sum_{j=1}^n a_j^{(k)} x_j + a_0^{(k)} \end{aligned} \quad (12)$$

where $x = [x_1 x_2 \dots x_n]$ 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 $B_j^{(k)}$, $j = 1 \dots n$, and by the coefficients $a_j^{(k)}$, $j = 1 \dots n$, of the crisp output. There are numerous options available for the fuzzification of crisp inputs, the composition of input MFs, and the way rule outputs are combined. 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:

$$\tilde{y} = \frac{\sum_{k=1}^M \mu_{B^{(k)}}(x) \cdot y^{(k)}}{\sum_{k=1}^M \mu_{B^{(k)}}(x)} \quad (13)$$

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 that 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: non-neutral), for one sentiment lexicon–dataset combination at a time. MultiLexANFIS has six inputs; it includes the positive and negative score of a tweet from all the three lexicons: VADER, AFINN and SentiWordNet, and one sentiment label as output for each dataset.

3.3.1 Single-lexicon-based ANFIS

The fuzzy inference system (FIS) is of Sugeno first-order type with two inputs TweetPos and TweetNeg and one output label *sentiment*. A common rule set with two inputs and three 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 , where $i = 1, 2, 3$. The fuzzy if-then rules are then defined as follows:

$$\text{Rule1 : If } x \text{ is } A_1 \text{ and } y \text{ is } B_1 \text{ then } f_1 = p_1x + q_1y + r_1 \quad (14)$$

$$\text{Rule2 : If } x \text{ is } A_2 \text{ and } y \text{ is } B_1 \text{ then } f_2 = p_2x + q_2y + r_2 \quad (15)$$

$$\text{Rule3 : If } x \text{ is } A_3 \text{ and } y \text{ is } B_1 \text{ then } f_3 = p_3x + q_3y + r_3 \quad (16)$$

$$\text{Rule4 : If } x \text{ is } A_1 \text{ and } y \text{ is } B_2 \text{ then } f_4 = p_4x + q_4y + r_4 \quad (17)$$

$$\text{Rule5 : If } x \text{ is } A_2 \text{ and } y \text{ is } B_2 \text{ then } f_5 = p_5x + q_5y + r_5 \quad (18)$$

$$\text{Rule6 : If } x \text{ is } A_3 \text{ and } y \text{ is } B_2 \text{ then } f_6 = p_6x + q_6y + r_6 \quad (19)$$

$$\text{Rule7 : If } x \text{ is } A_1 \text{ and } y \text{ is } B_3 \text{ then } f_7 = p_7x + q_7y + r_7 \quad (20)$$

$$\text{Rule8 : If } x \text{ is } A_2 \text{ and } y \text{ is } B_3 \text{ then } f_8 = p_8x + q_8y + r_8 \quad (21)$$

$$\text{Rule9 : If } x \text{ is } A_3 \text{ and } y \text{ is } B_3 \text{ then } f_9 = p_9x + q_9y + r_9. \quad (22)$$

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 2019a). 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:

$$\text{Rule } R_j : \text{IF } A \text{ is } A_1^j \text{ and } B \text{ is } B_1^j \text{ THEN } C \text{ is } C_1^j, \quad (23)$$

$$j = 1, 2, \dots, r$$

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

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

The ANFIS architecture comprises five layers (Jang et al. 1997). Our single-lexicon ANFIS architecture is shown in Fig. 1. We denote the output of the i th node in layer 1 as $O_{1,i}$. Each layer is described next.

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

$$O_{1,i} = \mu_{A_i}(x), \quad \text{for } i = 1, 2, 3 \text{ or} \quad (24)$$

$$O_{1,i} = \mu_{B_{i-3}}(y), \quad \text{for } i = 4, 5, 6 \quad (25)$$

where $O_{1,i}$ 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). Here, the MF for A and B can be any parametrized MF: triangular (trimf), generalized bell (gbellmf) and Gaussian (gaussmf) described in Eqs. (26), (27) and (28) respectively.

$$\mu_t(x) = \begin{cases} 0, & x \leq d \\ (x-d)/(e-d), & d < x \leq e \\ (f-x)/(f-e), & e < x \leq f \\ 0, & x \geq f \end{cases} \quad (26)$$

$$\mu_{gb}(x) = \frac{1}{1 + \left| \frac{x-w}{u} \right|^{2v}} \quad (27)$$

$$\mu_{ga}(x) = e^{-\frac{1}{2} \left(\frac{x-c}{\sigma} \right)^2}. \quad (28)$$

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 Eq. (26). Gbellmf is a symmetrical shape similar to a bell. As expressed by Eq. (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 Eq. (28). The values of these parameters are attuned in the learning phase of ANFIS. The parameters in this layer are referred as premise parameters.

Layer 2 Every node in this layer is a fixed node labeled as Π , whose output is the product of all the incoming signals shown in Eq. (29). Each node output represents the firing strength of a rule, where $i = 1, 2, \dots, 9$.

Fig. 1 Single-lexicon-based ANFIS architecture with nine rules

$$O_{2,i} = w_i = \mu_{A_k}(x)\mu_{B_j}(y), \quad k, j = 1, 2, 3. \quad (29)$$

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 Eq. (30). Outputs of this layer are called the normalized firing strengths.

$$O_{3,i} = \bar{w} = \frac{w_i}{\sum_i w_i}. \quad (30)$$

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

$$O_{4,i} = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i) \quad (31)$$

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 The single node in this layer is a fixed node labeled $\sum$, which computes the overall output as the summation of all incoming signals:

$$\text{overall output} = O_{5,i} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}. \quad (32)$$

3.3.2 Multiple-lexicon-based ANFIS: MultiLexANFIS

MultiLexANFIS combines all the three lexicons into a single integrated unit. MultiLexANFIS has six inputs; it includes positive and negative scores 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 $sneg$. $vpos$ 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 the sentiment label, which has value 1 for neutral tweet and 0 for non-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 are a total of 64 rules ($2^6 = 64$). These rules are depicted in Tables 1 and 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 scores for each textual document or sentence or review. MultiLexANFIS architecture has five layers similar to the single-lexicon ANFIS discussed in Sect. 3.3.1, except for the fact that the number of inputs has increased from two to six and the number of MFs for each input is two. MultiLexANFIS architecture is depicted in Fig. 2.

ANFIS learns by updating tunable parameters which are the premise parameters and the consequent parameters. The first and fourth layers 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 estimate method to update the consequent parameters before calculating the final output. In the second pass or backward pass, the error is propagated

Table 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 2 MultiLexANFIS rules (33–64)

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

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 observed values, as shown

in Eq. (33). In other words, it measures the quality of the fit between the actual data and the predicted model.

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^n (\text{actual}(t) - \text{predicted}(t))^2}{p}} \quad (33)$$

where $\text{actual}(t)$ and $\text{predicted}(t)$ are the actual and estimated values, respectively, and n is the total number of records in the testing data, $t = 1, 2, 3, \dots, p$. The overall process flow of our proposed ANFIS for social SA is summarized in Fig. 3.

Fig. 2 MultiLexANFIS architecture with 64 rules

Fig. 3 Overall process of our proposed ANFIS for social SA

4 Experimental setup and implementation

In the experimental phase, the proposed ANFIS was simulated in MATLAB 2020b version with Intel Core i5 processor, 64-bit operating system and 8 GB RAM. We have used a total of ten publicly available datasets: Apple Twitter Dataset (2019), Nuclear Twitter Dataset (2019), Sanders Twitter Dataset (2019), SemEval 2017, SemEval 2016, SemEval 2015, Stanford Twitter Sentiment Test Set (STS-Test) (Go et al. 2009), Airline Twitter Dataset (2019), Twitter_2019 (2019) dataset and Reddit dataset. All the datasets contain three types of tweets: positive, negative and neutral. The class distribution of all datasets is represented in Table 3. We have combined the positive and negative tweets into one class: non-neutral; our goal is to classify the tweets by ANFIS into two classes: neutral and non-neutral. The training and testing sets of these datasets are constructed using 70–30% split ratio. Our proposed model uses inputs derived from sentiment lexicons. The three sentiment lexicons used in this study are: VADER (Hutto and Gilbert 2014), AFINN (Nielsen 2011) and SentiWordNet (Baccianella et al. 2010). Our proposed model is compared with eight state-of-the-art methods that include both fuzzy and non-fuzzy approaches; their description is mentioned in the next section.

In our single-lexicon-based ANFIS variants, there are two inputs: positive and negative scores, and one output for sentiment. Class 1 indicates neutral tweets, and Class 0 indicates non-neutral tweets. The number of fuzzy rules in grid partitioning is number of MFs $\wedge$ number of input variables. In single-lexicon-based ANFIS variant, since

Table 3 Class distribution of different datasets

	Sanders	Nuclear	Apple	SemEval 2017	SemEval 2016	SemEval 2015	STS-Test (Go et al. 2009)	Airline	Twitter	Reddit
Positive	519	10	423	2375	5157	4377	182	2363	72,249	15,829
Negative	572	19	1219	3972	1225	1745	177	9178	35,509	8278
Neutral	2333	161	2162	5937	2667	5593	139	3099	55,212	12,909
Total	3424	190	3804	12,284	9049	11,715	498	14,640	162,970	37,014

each input has three membership functions (MFs), so in total, ANFIS has nine rules ($3^2 = 9$). In MultiLexANFIS, there are six inputs (two from each of the three lexicons) and the same output variable for sentiment, and each input has two MFs. Hence, it has a total of 64 rules ($2^6 = 64$). Optimal number of inputs and MFs are selected for both the 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. The “curse of dimensionality” refers to such situation where the number of fuzzy rules, when the grid partitioning is used, increases exponentially. Grid partitioning generates rules by enumerating all possible combinations of membership functions of all inputs; this leads to an exponential explosion even when the number of inputs is moderately large. 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 combines the gradient descent method 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 remains constant after 500 epochs, while in MultiLexANFIS, the training error remains constant after 100 epochs. Hence, the number of epochs required by single lexicon is 500 epochs, while MultiLexANFIS requires 100 epochs during training phase.

5 Results and discussion

In this paper, we have proposed two ANFIS 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 ten datasets. 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.

Our ANFIS has been compared with eight state-of-the-art methods that include various fuzzy, non-fuzzy and deep

learning models. The eight state-of-the-art methods are: (1) A deep learning model—recurrent neural network (RNN), developed by (Nemes and Kiss 2021) for classifying tweets. (2) A deep learning model—long short-term memory (LSTM) model, developed by (Rahman et al. 2016) for text classification. (3) An n-gram model developed by (Tripathy et al. 2016) that converts text reviews into numeric matrices using count vectorizer and TF-IDF, which are then given as input to machine learning algorithm—support vector machine (SVM) for classification. (4) An unsupervised fuzzy rule-based system for sentiment analysis is a published previous work of authors (Vashishtha and Susan 2019a). This fuzzy rule-based classifier has nine rules implemented in a Mamdani fuzzy inference system. (5) VADER SA by Hutto and Gilbert (2014); this approach yields positive, negative, neutral scores for a given text by deploying a number of rules. (6) AFINN SA by Nielsen (2011) approach uses AFINN dictionary to assign scores to each word; the sentiment is computed for text by summing up all scores. (7) SA using SentiWordNet lexicon by Cavalcanti et al. (2011) approach uses SentiWordNet to assign scores to selected words—adjectives, verbs, adverbs and nouns, and then, final sentiment is computed by considering the average score of all words. (8) SA using SentiWordNet lexicon by Ortega et al. (2013) approach uses SentiWordNet to check the polarity of words, assigns fixed values for each polarity and then computes the sentiment by summing up the values.

5.1 Single-lexicon-based ANFIS

In this research work, we have considered three sentiment lexicons, so there are three single-lexicon-based ANFIS: VADER (Hutto and Gilbert 2014), AFINN (Nielsen 2011) and SentiWordNet (Baccianella et al. 2010) single-lexicon-based ANFISs. These systems have nine rules. We have performed experiments with each of these ANFISs for all the datasets and three MFs; the aim is to find out which MF is best suitable for the given ANFIS. Table 4 depicts the outcomes for VADER lexicon-based ANFIS, and the bold values highlight the lowest RMSE achieved for each dataset. Both trimf and gaussmf membership functions

Table 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
Airline	0.39421	0.38901	0.388
Twitter	0.44973	0.43362	0.43371
Reddit	0.43592	0.39184	0.39223

Bold values are the lowest RMSE values for each dataset

Table 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
Airline	0.38994	0.38732	0.38982
Twitter	0.42823	0.42691	0.42991
Reddit	0.43172	0.42713	0.45053

Bold values are the lowest RMSE values for each dataset

Table 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
Airline	0.39752	0.39614	0.39623
Twitter	0.43542	0.43032	0.43214
Reddit	0.45023	0.46493	0.45302

Bold values are the lowest RMSE values for each dataset

(MFs) have lower RMSE in a greater number of datasets compared to gbellmf. The least RMSE, 0.33174, is scored in nuclear dataset under gaussmf, while in the case of AFINN-based ANFIS in Table 5, gbellmf has performed better than trimf and gaussmf; it has the lowest RMSE for three datasets. The least RMSE, 0.3873, is scored in Airline Twitter dataset under gbellmf. In Table 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 0.3635 under the trimf. Hence, we cannot identify a single MF which works well for all the three variants. It can be deduced that trimf and gbell MFs are better as in majority cases they yield lower RMSE.

5.1.1 Comparison of single-lexicon-based ANFIS with other approaches

Comparison of VADER lexicon-based ANFIS with VADER-specific approaches is shown in Table 7. It has been compared with two approaches—fuzzy rule approach (Vashishtha and Susan 2019) and the non-fuzzy approach (Hutto and Gilbert 2014). It can be clearly observed that our ANFIS outperforms all methods for all datasets. For instance, in Apple dataset RMSE of ANFIS is 0.45668, while other methods have 0.6127 and 0.6223. Table 8 presents the achievement of AFINN lexicon-based ANFIS with AFINN-specific methods. It has been compared with two approaches—fuzzy rule approach (Vashishtha and Susan 2019) and the non-fuzzy approach (Nielsen 2011). Our ANFIS has the lowest RMSE as compared to both approaches for all datasets. For SemEval 2017 dataset, ANFIS scored 0.47428, while others scored 0.6127 and 0.7046, respectively. Similar observations can be made for SentiWordNet lexicon-based ANFIS with SentiWordNet-specific approaches in Table 9. It has been compared with three approaches—fuzzy rule approach (Vashishtha and Susan 2019) and two non-fuzzy approaches: Cavalcanti (Cavalcanti et al. 2011) and Ortega (Ortega et al. 2013). Here also, ANFIS surpasses all methods, for example, for the SemEval 2016 dataset, 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 ANFISs have lowest RMSE for all datasets compared to state-of-the-art SA models. But it is difficult to decide the best membership function (MF) for each lexicon among all datasets. In other words, we cannot conclude which is the best MF among trimf, gbellmf and gaussmf for the VADER, AFINN or SentiWordNet lexicons. In VADER, there is a confusion 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. The execution

Table 7 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 2019a)	0.6223
	Vader-ANFIS	0.45668
Nuclear	VADER SA (Hutto and Gilbert 2014)	0.8013
	Fuzzy rule (Vashishtha and Susan 2019a)	0.4168
	Vader-ANFIS	0.33174
Sanders	VADER SA (Hutto and Gilbert 2014)	0.6353
	Fuzzy rule (Vashishtha and Susan 2019a)	0.5593
	Vader-ANFIS	0.43453
SemEval (2017)	VADER SA (Hutto and Gilbert 2014)	0.6137
	Fuzzy rule (Vashishtha and Susan 2019a)	0.6892
	Vader-ANFIS	0.47814
SemEval (2016)	VADER SA (Hutto and Gilbert 2014)	0.6309
	Fuzzy rule (Vashishtha and Susan 2019a)	0.8078
	Vader-ANFIS	0.45195
SemEval (2015)	VADER SA (Hutto and Gilbert 2014)	0.5764
	Fuzzy rule (Vashishtha and Susan 2019a)	0.689
	Vader-ANFIS	0.45257
STS (Go et al. 2009)	VADER SA (Hutto and Gilbert 2014)	0.4251
	Fuzzy Rule (Vashishtha and Susan 2019a)	0.7525
	Vader-ANFIS	0.3686
Airline	VADER SA (Hutto and Gilbert 2014)	0.5012
	Fuzzy rule (Vashishtha and Susan 2019a)	0.8552
	Vader-ANFIS	0.388
Twitter	VADER SA (Hutto and Gilbert 2014)	0.5296
	Fuzzy rule (Vashishtha and Susan 2019a)	0.7949
	Vader-ANFIS	0.4336
Reddit	VADER SA (Hutto and Gilbert 2014)	0.4648
	Fuzzy rule (Vashishtha and Susan 2019a)	0.7765
	Vader-ANFIS	0.3918

Bold values are the lowest RMSE values for each dataset

time for single-lexicon-based ANFIS is about an average of 10–15 min.

5.2 Multiple-lexicon-based ANFIS: MultiLexANFIS

To solve the above confusion of choosing the best membership function (MF) for a given lexicon, the 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 Sect. 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 10 displays the performance of MultiLexANFIS for all datasets. This system has been compared with

single-lexicon-based ANFIS as well; the best (lowest) RMSE—combination of lexicon—membership function for each dataset, is represented in the second column of Table 10. For instance, for the Apple dataset, among single-lexicon ANFIS, VADER variant has achieved the lowest RMSE of 0.457 using the gbell MF, while MultiLexANFIS has attained 0.448 RMSE using the gauss MF, which is the lowest value. Similar analysis is accomplished for all the datasets. This integrated MultiLexANFIS achieves lower RMSE as compared to the single-lexicon-based ANFIS for almost all datasets, though for some datasets the results are comparable. We also note down that trimf: triangular membership function performs the best in MultiLexANFIS (7 out of 10 datasets) and gaussmf: Gaussian membership function is at second position. Thus, the problem of confusion among the MFs for all datasets has been resolved. Comparison amidst the single-lexicon

Table 8 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 2019a)	0.6402
	AFINN-ANFIS	0.46623
Nuclear	AFINN SA (Nielsen 2011)	0.846
	Fuzzy rule (Vashishtha and Susan 2019a)	0.4413
	AFINN-ANFIS	0.39682
Sanders	AFINN SA (Nielsen 2011)	0.6235
	Fuzzy rule (Vashishtha and Susan 2019a)	0.5634
	AFINN-ANFIS	0.44414
SemEval (2017)	AFINN SA (Nielsen 2011)	0.6127
	Fuzzy rule (Vashishtha and Susan 2019a)	0.7046
	AFINN-ANFIS	0.47428
SemEval (2016)	AFINN SA (Nielsen 2011)	0.6591
	Fuzzy rule (Vashishtha and Susan 2019a)	0.8303
	AFINN-ANFIS	0.45331
SemEval (2015)	AFINN SA (Nielsen 2011)	0.5649
	Fuzzy rule (Vashishtha and Susan 2019a)	0.718
	AFINN-ANFIS	0.44535
STS (Go et al. 2009)	AFINN SA (Nielsen 2011)	0.4321
	Fuzzy rule (Vashishtha and Susan 2019a)	0.8287
	AFINN-ANFIS	0.38946
Airline	AFINN SA (Nielsen 2011)	0.532
	Fuzzy rule (Vashishtha and Susan 2019a)	0.8819
	AFINN-ANFIS	0.3873
Twitter	AFINN SA (Nielsen 2011)	0.5444
	Fuzzy rule (Vashishtha and Susan 2019a)	0.8106
	AFINN-ANFIS	0.4269
Reddit	AFINN SA (Nielsen 2011)	0.483
	Fuzzy rule (Vashishtha and Susan 2019a)	0.8001
	AFINN-ANFIS	0.4271

Bold values are the lowest RMSE values for each dataset

ANFISs reveals that VADER lexicon performs best. This is observed in the second column of Table 10; among ten datasets, VADER lexicon has the lowest RMSE for six datasets. The execution time of MultiLexANFIS is about 30 min. The training error graph for MultiLexANFIS for the Apple dataset with gaussmf is shown in Fig. 4. The computed test error, 0.448, is depicted in Fig. 5.

5.2.1 Comparison of multiple-lexicon-based ANFIS: MultiLexANFIS with state of the art

The comparative analysis of our proposed MultiLexANFIS with eight state-of-the-art methods that include various fuzzy, non-fuzzy and deep learning models is presented in Table 11. The following are the state-of-the-art methods: (1) a deep learning model—recurrent neural network (RNN) developed by Nemes and Kiss (2021) for

classifying tweets, (2) a deep learning model—long short-term memory (LSTM) model developed by Rahman et al. (2016) for text classification, (3) an n-gram model with SVM classifier developed by c, (4) an unsupervised fuzzy rule-based system for sentiment analysis which is a published previous work of authors (Vashishtha and Susan 2019a), (5) VADER SA by Hutto and Gilbert (2014), (6) AFINN SA by Nielsen (2011) approach, (7) SA using SentiWordNet lexicon by Cavalcanti et al. (2011) approach and (8) SA using SentiWordNet lexicon by Ortega et al. (2013) approach. The deep learning models—RNN and LSTM, are given the same input features that are fed to MultiLexANFIS. The models are built using Keras and Tensorflow tools; RNN model consists of embedding layer, followed by bidirectional RNN wrapper, dropout layer and dense layer, while LSTM model contains LSTM layer instead of bidirectional RNN layer. The error produced by

Table 9 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 2019a)	0.655
	SENTIWORDNET-ANFIS	0.48503
Nuclear	Cavalcanti (Cavalcanti et al. 2011)	0.9205
	Ortega (Ortega et al. 2013)	0.8553
	Fuzzy rule (Vashishtha and Susan 2019a)	0.4168
	SENTIWORDNET-ANFIS	0.36351
Sanders	Cavalcanti (Cavalcanti et al. 2011)	0.795
	Ortega (Ortega et al. 2013)	0.6202
	Fuzzy rule (Vashishtha and Susan 2019a)	0.5655
	SENTIWORDNET-ANFIS	0.45318
SemEval (2017)	Cavalcanti (Cavalcanti et al. 2011)	0.6892
	Ortega (Ortega et al. 2013)	0.6693
	Fuzzy rule (Vashishtha and Susan 2019a)	0.7176
	SENTIWORDNET-ANFIS	0.49243
SemEval (2016)	Cavalcanti (Cavalcanti et al. 2011)	0.554
	Ortega (Ortega et al. 2013)	0.7164
	Fuzzy rule (Vashishtha and Susan 2019a)	0.8115
	SENTIWORDNET-ANFIS	0.45204
SemEval (2015)	Cavalcanti (Cavalcanti et al. 2011)	0.6839
	Ortega (Ortega et al. 2013)	0.6619
	Fuzzy rule (Vashishtha and Susan 2019a)	0.7222
	SENTIWORDNET-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 2019a)	0.7418
	SENTIWORDNET-ANFIS	0.42304
Airline	Cavalcanti (Cavalcanti et al. 2011)	0.4629
	Ortega (Ortega et al. 2013)	0.6804
	Fuzzy rule (Vashishtha and Susan 2019a)	0.8552
	SENTIWORDNET-ANFIS	0.3961
Twitter	Cavalcanti (Cavalcanti et al. 2011)	0.5654
	Ortega (Ortega et al. 2013)	0.7703
	Fuzzy rule (Vashishtha and Susan 2019a)	0.7949
	SENTIWORDNET-ANFIS	0.4321
Reddit	Cavalcanti (Cavalcanti et al. 2011)	0.5224
	Ortega (Ortega et al. 2013)	0.7256
	Fuzzy rule (Vashishtha and Susan 2019a)	0.7765
	SENTIWORDNET-ANFIS	0.4502

Bold values are the lowest RMSE values for each dataset

these models is higher than MultiLexANFIS, and this is because the parameter tuning and optimization in ANFIS are better than the tuning of weights during backpropagation in deep learning models. In Table 11, for instance we can observe for the Airline Twitter dataset, MultiLexANFIS has the lowest RMSE value—0.3843, while RNN has 0.4417 RMSE and LSTM has 0.4383 RMSE.

Overall, we can deduce that after MultiLexANFIS, RNN is second best followed by LSTM. The third approach (Tripathy et al. 2016) creates different n-grams—unigrams, bigrams and trigrams from text, and assigns term frequency-inverse document frequency (TF-IDF) values to each n-gram; these numerical vectors are sent as input to SVM classifier. The errors produced by the n-gram model

Table 10 MultiLexANFIS: multiple-lexicon-based ANFIS

Datasets	Least among Lex Lex MF RMSE	TRIMF RMSE	GBELL RMSE	GAUSS 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
Airline	A Gbell—0.3873	0.3843	0.3860	0.3859
Twitter	A Gbell—0.4269	0.4082	0.4159	0.4127
Reddit	V Gbell—0.4036	0.3918	0.4104	0.4377

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

Fig. 4 Training error for Apple dataset with gaussmf in MultiLexANFIS **Fig. 5** Test error for Apple dataset with gaussmf in MultiLexANFIS (red dots indicate the predicted values for testing data) (color figure online)

are higher than MultiLexANFIS for all datasets, and this is because of missing fuzzy factor in the extracted features and in the model. The selected n-grams are based on frequency, while in our model we use the polarity and sentiment score of each word derived from lexicons; this way we are able to deal with natural language in a better way for SA. The proposed MultiLexANFIS is also compared to fuzzy rule classifier's (Vashishtha and Susan 2019a) three variants—VADER, AFINN and SentiWordNet. Our neuro-fuzzy model yields better results by using scores from

sentiment lexicon as features for subjectivity classification (neutral non-neutral) compared to the unsupervised fuzzy rule classifier which uses the same scores for polarity classification (positive, negative or neutral). In fuzzy rule classifier, triangular (trimf) fuzzy membership is used and its parameters are fixed, while in our ANFIS model these parameters are tuned during each epoch, thereby producing lower errors. The next four state-of-the-art methods are unsupervised and non-fuzzy algorithms. They produce very high errors compared to all other approaches, and this is because they are simple and naïve in nature. Finally, we can conclude that the proposed MultiLexANFIS produces lower error for all datasets compared to other approaches and also it takes less execution time in contrast to single-lexicon-based ANFIS.

6 Conclusion

In this paper, a novel MultiLexANFIS has been proposed for social sentiment analysis. This architecture is an adaptive neuro-fuzzy inference system (ANFIS) that incorporates inputs from VADER, AFINN and SentiWordNet lexicons to perform sentiment analysis 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 into neutral and non-neutral tweets, (2) a novel set of 64 rules for Sugeno-type fuzzy inference system—MultiLexANFIS, (3) single-lexicon-based ANFIS variants to classify tweets and (4) comparison of fuzzy approaches, non-fuzzy approaches and deep learning models with MultiLexANFIS that affirms the supremacy of our proposed neuro-fuzzy system. ANFIS has deployed optimization with the help of a hybrid learning algorithm; that uses least square method and gradient descent to tune parameters. We can deduce that MultiLexANFIS has the

Table 11 Comparison of MultiLexANFIS with the state of the art

Datasets	Least among Single-Lex ANFIS	MultiLexANFIS	RNN (Nemes and Kiss 2021)	LSTM (Rahman et al. 2016)	n-gram SVM (Tripathy et al. 2016)	VADER—fuzzy rule (Vashishtha and Susan 2019a)	AFINN—fuzzy rule (Vashishtha and Susan 2019a)	SWN—fuzzy rule (Vashishtha and Susan 2019a)	VADER SA (Hutto and Gilbert 2014)	AFINN SA (Nielsen 2011)	SWN SA (Cavalcanti et al. 2011)	SWN SA (Ortega et al. 2013)
	Lex MF RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE
Apple	V Gbell—0.457	0.448	0.5176	0.5309	0.6689	0.6223	0.6402	0.655	0.6127	0.6112	0.7437	0.6449
Nuclear	V Gauss—0.332	0.320	0.4699	0.456	0.3245	0.4168	0.4413	0.4168	0.8013	0.846	0.9205	0.8553
Sanders	V Trimf—0.435	0.440	0.4856	0.4911	0.5527	0.5593	0.5634	0.5655	0.6353	0.6235	0.795	0.6202
SemEval (2017)	A Trimf—0.474	0.473	0.514	0.5254	0.6978	0.6892	0.7046	0.7176	0.6137	0.6127	0.6892	0.6693
SemEval (2016)	V Trimf—0.452	0.454	0.4632	0.4661	0.5462	0.8078	0.8303	0.8115	0.6309	0.6591	0.554	0.7164
SemEval (2015)	A Gauss—0.445	0.433	0.5615	0.5452	0.6909	0.6890	0.7180	0.7222	0.5764	0.5649	0.6839	0.6619
STS (Go et al. 2009)	V Gauss—0.369	0.366	0.5262	0.5244	0.5416	0.7525	0.8287	0.7418	0.4251	0.4321	0.5226	0.6555
Airline	A Gbell—0.3873	0.3843	0.4417	0.4383	0.4569	0.8552	0.8819	0.8816	0.5012	0.5320	0.4629	0.6804
Twitter_2019	A Gbell—0.4269	0.4082	0.5241	0.5275	0.5629	0.7949	0.8106	0.794	0.5296	0.5444	0.5654	0.7703
Reddit	V Gbell—0.4036	0.3918	0.5662	0.5627	0.5723	0.7765	0.8001	0.762	0.4648	0.483	0.5224	0.7256

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

lowest RMSE for the ten benchmark datasets, as compared to other approaches. Results reveal that among the single-lexicon ANFISs, 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 resolved using MultiLexANFIS. 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 score 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 sarcasm and mixed tweets, i.e., the tweets that contain both positive and negative opinions. In future, we will compute the sentiment of mixed tweets by enhancing our model using advanced fuzzy deep neural networks.

Funding This study was not funded by any agency.

Declarations

Conflict of interest Both authors are associated with the Department of Information Technology, Delhi Technological University.

Ethical approval This article does not contain any studies with animals performed by any of the authors.

Human and animal rights This article does not contain any studies with human participants or animals performed by any of the authors.

References

- 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). IEEE, pp 73–80
- Airline Twitter Dataset (2019) <https://www.kaggle.com/crowdflower/twitter-airline-sentiment>. Accessed on 29 Aug 2021
- Ahuja R, Rastogi H, Choudhuri A, Garg B (2015) Stock market forecast using sentiment analysis. In: IEEE 2nd international conference on computers for sustainable global development, pp 1008–1010
- Anagha M, Kumar RR, Sreetha K, Raj PR (2015) Fuzzy logic based hybrid approach for sentiment analysis of Malayalam movie reviews. In: 2015 IEEE international conference on signal processing, informatics, communication and energy systems (SPICES). IEEE, pp 1–4
- Apple Twitter Dataset (2019). <https://data.world/crowdflower/apple-twitter-sentiment>. Accessed 29 Aug 2021
- Asimuzzaman M, Nath PD, Hossain F, Hossain A, Rahman RM (2017) Sentiment analysis of bangla microblogs using adaptive neuro fuzzy system. In: 2017 13th international conference on natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD). IEEE, pp 1631–1638
- Baccianella S, Esuli A, Sebastiani F (2010) Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. *Lrec* 10(2010):2200–2204
- Banerjee S, Pedersen T (2002) An adapted Lesk algorithm for word sense disambiguation using WordNet. In: International conference on intelligent text processing and computational linguistics. Springer, Berlin, Heidelberg, pp 136–145
- Çakıt E, Karwowski W, Servi L (2019) Application of soft computing techniques for estimating emotional states expressed in Twitter® time series data. *Neural Comput Appl* 1–14:2019
- Cavalcanti DC, Prudêncio RBC, Pradhan SS, Shah JY, Pietrobon RS (2011) Good to be bad? Distinguishing between positive and negative citations in scientific impact. In: 2011 IEEE 23rd international conference on tools with artificial intelligence. IEEE, pp 156–162
- Cosma G, Acampora G (2016) Neuro-fuzzy sentiment analysis for customer review rating prediction. In: Sentiment analysis and ontology engineering. Springer, Cham, pp 379–397
- Ghiassi M, Skinner J, Zimbra D (2013) Twitter brand sentiment analysis: a hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Syst Appl* 40(16):6266–6282
- Guerra PHC, Veloso A, Meira Jr W, Almeida V (2011) From bias to opinion: a transfer-learning approach to real-time sentiment analysis. In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp 150–158
- Gupta N, Sharma G, Sharma RS (2014) A comparative study of ANFIS membership function to predict ERP user satisfaction using ANN and MLRA. *Int J Comput Appl* 105(5)
- Go A, Bhayani R, Huang L (2009) Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford 1, no. 12 (2009)
- Hutto CJ, Gilbert E (2014) Vader: a parsimonious rule-based model for sentiment analysis of social media text. In: Eighth international AAAI conference on weblogs and social media
- Indhuja K, Reghu RP (2014) Fuzzy logic based sentiment analysis of product review documents. In: 2014 first international conference on computational systems and communications (ICCSC). IEEE, pp 18–22
- Jang JSR (1993) ANFIS: adaptive-network-based fuzzy inference system. *IEEE Trans Syst Man Cybern* 23(3):665–685
- Jang JSR, Sun CT, Mizutani E (1997) Neuro-fuzzy and soft computing—a computational approach to learning and machine intelligence [Book Review]. *IEEE Trans Autom Control* 42(10):1482–1484
- Jefferson C, Liu H, Cocea M (2017) Fuzzy approach for sentiment analysis. In: 2017 IEEE international conference on fuzzy systems (FUZZ-IEEE). IEEE, pp 1–6
- 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. Springer, Cham, pp 63–70
- Kalarani P, Brunda SS (2019) Sentiment analysis by POS and joint sentiment topic features using SVM and ANN. *Soft Comput* 23(16):7067–7079
- Katta P, Hegde NP (2019) A hybrid adaptive neuro-fuzzy interface and support vector machine based sentiment analysis on political twitter data. *Int J Intell Eng Syst* 12(1):165–173

- Khan W, Malik U, Ghazanfar MA, Azam MA, Alyoubi KH, Alfakeeh AS (2020) Predicting stock market trends using machine learning algorithms via public sentiment and political situation analysis. *Soft Comput* 24(15):11019–11043
- Khattak A, Paracha WT, Asghar MZ, Jillani N, Younis U, Saddozai FK, Hameed IA (2020) Fine-grained sentiment analysis for measuring customer satisfaction using an extended set of fuzzy linguistic Hedges. *Int J Comput Intell Syst* 13(1):744–756
- Kulkarni S, Vidhate A, Anuradha G (2014) Neuro fuzzy approach for predicting sales performance of movies considering sentiments in online reviews
- 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. *Int J Multidiscip Cryptol Inf Secur* 1(2)
- Koppel M, Schler J (2005) Using neutral examples for learning polarity. In: *International joint conference on artificial intelligence*. Lawrence Erlbaum Associates Ltd, vol 19, p 1616
- Koppel M, Schler J (2006) The importance of neutral examples for learning sentiment. *Comput Intell* 22(2):100–109
- Liu H, Cocea M (2017) Fuzzy rule based systems for interpretable sentiment analysis. In: *2017 ninth international conference on advanced computational intelligence (ICACI)*. IEEE, pp 129–136
- Madhusudhanan S, Moorthi M (2019) Optimized fuzzy technique for enhancing sentiment analysis. *Clust Comput* 22(5):11929–11939
- 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*. Springer, Cham, pp 63–74
- Nagamanjula R, Pethalakshmi A (2020) A novel framework based on bi-objective optimization and LAN 2 FIS for Twitter sentiment analysis. *Soc Netw Anal Min* 10:1–16
- Nemes L, Kiss A (2021) Social media sentiment analysis based on COVID-19. *J Inf Telecommun* 5(1):1–15
- Nguyen TL, Kavuri S, Lee M (2019) A multimodal convolutional neuro-fuzzy network for emotion understanding of movie clips. *Neural Netw* 118:208–219
- Nielsen FÅ (2011) AFINN. Richard Petersens Plads, Building 321 (2011)
- Nuclear Twitter Dataset (2019) <https://data.world/crowdfunder/emotions-about-nuclear-energy>. Accessed 29 Aug 2021
- Ortega R, Fonseca A, Montoyo A (2013) SSA-UO: unsupervised Twitter sentiment analysis. In: *Second joint conference on lexical and computational semantics (* SEM)*, vol 2, pp 501–507
- 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)*. IEEE, pp 498–503
- 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)*. IEEE, pp 416–419
- Phan HT, Nguyen NT, Van Cuong T, Hwang D (2019) 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)*. IEEE, pp 1–6
- Rahman L, Mohammed N, Al Azad AK (2016) A new LSTM model by introducing biological cell state. In: *2016 3rd international conference on electrical engineering and information communication technology (ICEEICT)*. IEEE, pp 1–6
- Rustamov S, Mustafayev E, Clements M (2013) Sentiment analysis using Neuro-Fuzzy and Hidden Markov models of text. In: *2013 Proceedings of IEEE Southeastcon*. IEEE, pp 1–6
- 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
- Rustamov S (2018) A hybrid system for subjectivity analysis. *Advances in Fuzzy Systems* 2018
- Sanders Twitter Dataset (2019) <http://www.sananalytics.com/lab>. Accessed 29 Aug 2021
- Sasikala P, Sheela LMI (2020) Sentiment analysis of online product reviews using DLMNN and future prediction of online product using IANFIS. *J Big Data* 7(1):1–20
- Siddiqua UA, Ahsan T, Chy AN (2016) Combining a rule-based classifier with weakly supervised learning for twitter sentiment analysis. In: *2016 international conference on innovations in science, engineering and technology (ICISSET)*. IEEE, pp 1–4
- SemEval (2017), <http://alt.qcri.org/semeval2017/task4/>. Accessed 29 Aug 2021
- SemEval (2016), <http://alt.qcri.org/semeval2016/task4/>. Accessed 29 Aug 2021
- SemEval (2015), <http://alt.qcri.org/semeval2015/task10/>. Accessed 29 Aug 2021
- Serrano-Guerrero J, Romero FP, Olivas JA (2021) Fuzzy logic applied to opinion mining: a review. *Knowl-Based Syst*:107018
- Sugeno M, Kang GT (1988) Structure identification of fuzzy model. *Fuzzy Sets Syst* 28(1):15–33
- 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*. IEEE, pp 115–118
- Talpur N, Salleh MNM, Hussain K (2017) An investigation of membership functions on performance of ANFIS for solving classification problems. In: *IOP conference series: materials science and engineering*. IOP Publishing, vol 226, no 1, p 012103
- Tripathy A, Agrawal A, Rath SK (2016) Classification of sentiment reviews using n-gram machine learning approach. *Expert Syst Appl* 57:117–126
- Twitter_2019 (2019) <https://www.kaggle.com/cosmos98/twitter-and-reddit-sentimental-analysis-dataset>. Accessed 29 Aug 2021
- Vashishtha S, Susan S (2018) Fuzzy logic based dynamic plotting of mood swings from tweets. In: *International conference on innovations in bio-inspired computing and applications*, Springer, Cham, pp 129–139
- Vashishtha S, Susan S (2019a) Fuzzy rule based unsupervised sentiment analysis from social media posts. *Expert Syst Appl* 138:112834
- Vashishtha S, Susan S (2019b) Sentiment cognition from words shortlisted by fuzzy entropy. *IEEE Trans Cogn Dev Syst*. <https://doi.org/10.1109/TCDS.2019.2937796>
- Vashishtha S, Susan S (2020a) Inferring sentiments from supervised classification of text and speech cues using fuzzy rules. *Procedia Comput Sci* 167:1370–1379
- Vashishtha S, Susan S (2020b) Fuzzy interpretation of word polarity scores for unsupervised sentiment analysis. In: *2020 11th international conference on computing, communication and networking technologies (ICCCNT)*. IEEE (in press)
- Vashishtha S, Susan S (2021) Highlighting keyphrases using sentiment scoring and fuzzy entropy for unsupervised sentiment analysis. *Expert Syst Appl* 169:114323
- Valdivia A, Luzón MV, Herrera F (2017) Neutrality in the sentiment analysis problem based on fuzzy majority. In: *2017 IEEE international conference on fuzzy systems (FUZZ-IEEE)*. IEEE, pp 1–6
- Valdivia A, Luzón MV, Cambria E, Herrera F (2018) Consensus vote models for detecting and filtering neutrality in sentiment analysis. *Inf Fusion* 44(2018):126–135

- Wu KY, Zhou MC, Lu XS, Huang L (2017) A fuzzy logic-based text classification method for social media data. In: 2017 IEEE international conference on systems, man, and cybernetics (SMC). IEEE, pp 1942–1947
- Xie X, Ge S, Hu F, Xie M, Jiang N (2019) An improved algorithm for sentiment analysis based on maximum entropy. *Soft Comput* 23(2):599–611
- Yakubu H, Kwong CK, Lee CK (2021) A multigene genetic programming-based fuzzy regression approach for modelling customer satisfaction based on online reviews. *Soft Comput* 25(7):5395–5410
- Žunić E, Djedović A, Avdagić Z (2016) Decision support system for candidates classification in the employment process based on ANFIS method. In: 2016 XI International symposium on telecommunications (BIHTEL). IEEE, pp 1–6

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.