Fuzzy Rule based Unsupervised Sentiment Analysis from Social Media Posts

Article *in* Expert Systems with Applications · July 2019

DOI: 10.1016/j.eswa.2019.112834

2 authors:

Srishti Vashishtha
Bharati Vidyapeeth College of Engineering, Delhi

Seba Susan
Delhi Technological University

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

Fuzzy Rule based Unsupervised Sentiment Analysis from Social Media Posts

Srishti Vashishtha * [1] and Seba Susan[1]

[1] Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi-110042, India
srishtidtu@gmail.com and seba_406@yahoo.in
+919953812446

Abstract. In this paper, we compute the sentiment of social media posts using a novel set of fuzzy rules involving multiple lexicons and datasets. The proposed fuzzy system integrates Natural Language Processing techniques and Word Sense Disambiguation using a novel unsupervised nine fuzzy rule based system to classify the post into: positive, negative or neutral sentiment class. We perform 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. Traditionally, Sentiment Analysis of twitter data is performed using a single lexicon. Our results can give an insight to researchers to choose which lexicon is best for social media. The fusion of fuzzy logic with lexicons for sentiment classification provides a new paradigm in Sentiment Analysis. Our method can be adapted to any lexicon and any dataset (two-class or three-class sentiment). The experiments on benchmark datasets yield higher performance for our approach as compared to the state-of-the-art.

Keywords: Social Media, Twitter, Sentiment Analysis, Fuzzy Rule, Lexicon.

1 Introduction

Sentiment Analysis is a challenging research problem especially on social media. Users can freely express their views, opinions and feelings on different trending events, topics, etc. via social media posts. These posts need to be analysed to know what sentiment is conveyed through these posts. Sentiment Analysis, also referred as emotion AI, involves analyzing views from the written text so as to understand and gauge human emotions. The social media allows world-wide users to connect and interact with each other and express the opinions on general topics. Social Sentiment Analysis can be used to improve customer service and marketing and also serves as a measure of social media performance. In recent years, the impact of social media websites on daily life has become so considerable that even information on large and small incidents or disasters is

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

gathered via social media sites. The users portray not only the content about events but also their feelings (Yoo *et al.*,2018). The automated extraction of sentiment from these posts and classifying them into different polarities--positive, negative or neutral-- has received extensive attention from researchers during the past decade.

Twitter is one of the popular social media and boasts of a respectful 255 million active monthly users. Some of the challenges in analysing tweets are: use of informal language, short forms, abbreviations, heavy use of emoticons and slangs. Twitter, also known as microblogging, has limited size of tweets that makes it difficult to compute the polarity. In this paper, we apply fuzzy rule-based unsupervised approach to process the tweets in such a way as to overcome the above challenges. We have implemented our approach on multiple public twitter datasets using multiple lexicons. The proposed fuzzy rule-based approach can compute sentiment for two-class and three-class sentiment datasets. Two-class datasets have only positive and negative sentiment while three-class have neutral sentiment as well.

Fuzzy logic is an extension of deterministic logic, i.e. the truth value has range from 0 to 1 rather than a binary value. The primary aim of the theory of fuzzy logic is turning a black and white problem into a grey problem (Zadeh, 2015). In the field of artificial intelligence, possibly the easiest way to represent the human knowledge is to transform it into natural language expressions in the format of IF-THEN rules. These rules are based on natural language representations and models, which are themselves based on fuzzy sets and fuzzy logic (Ross, 2010). Classification systems based on fuzzy rules are powerful and acknowledged tools for pattern recognition and classification. These systems can handle uncertainty, ambiguity or vagueness in a very efficient way due to the presence of fuzziness (López *et al.*, 2015). We have used the concept of fuzzy rule-based system to create our own nine fuzzy rules to determine the sentiment of each tweet.

The main contributions of this paper 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 dataset (two-class or three-class) 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 rest of the paper is organized as follows. Section 2 describes the state-of-the-art on Sentiment Analysis from social media, while our proposed fuzzy rule-based system is presented in Section 3. Section 4 is about the experimental setup & implementation. Results are discussed in Section 5. The overall conclusions are drawn in Section 6.

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

2 Related Work

In recent years, a lot of 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 methods for Sentiment Analysis of social media posts can be classified as supervised, semi-supervised and unsupervised approach. In social media, to keep track of user opinion behavior, historical information about users can be used to develop a content-based supervised model to predict the sentiment. These models are developed in (Chen *et al.*, 2018) using recurrent neural network in order to explore the expression styles of users which give useful information to marketing companies. Models have been developed in (Liu *et al.*, 2015) for the sentiment classification of tweets specific to a topic. These classifiers are supervised and built on common features and mixed labeled 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 *et al.*, 2013; Jain *et al.*, 2015; Parveen *et al.*, 2015; Yan *et al.*, 2017; Saleena *et al.*, 2018; Barnaghi *et al.*, 2016; Hamdan *et al.*, 2013) and Support Vector Machines (SVM) (Neethu *et al.*, 2013; Saleena *et al.*, 2018; Hamdan *et al.*, 2013) for Sentiment Analysis from tweets. Windasari *et al.* 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 *et al.*, 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 *et al.*,2018). This work focuses on the sentiments about these AI assistants to ascertain which assistant is statistically better than the other with the help of VADER lexicon (Gilbert *et al.*, 2014) and T-test, Kruskal-Wallis test, and Mann-Whitney test (Park *et al.*,2018). In another recent work (Montoro *et*

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

al.,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, 2010). Tsukamoto fuzzy rule-based system has been used in (Liu *et al.*,2017; Jefferson *et al.*,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 negative 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 sentiments of tweets. This Naïve Bayes classifier is trained with the help of several sentiment lexicons (Siddiqua *et al.*,2016).

In 1975, Mamdani and Assilian's influential work (Mamdani *et al.*,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 (Márquez *et al.*, 2007; Duţu *et al.*, 2018) 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 parametric 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 (López *et al.*, 2015; Chang *et al.*, 2008; Sanz *et al.*, 2013; Ishibuchi *et al.*, 2001; Ishibuchi *et al.*, 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 *et al.*, 2001; Ishibuchi *et al.*, 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 *et al.*, 2014) while another paper compares the machine learning based, lexicon based and graph-based classification methods (Giachanou *et al.*, 2016). Table 1 presents a summary of few papers based on

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

Twitter Sentiment Analysis indicating their references, algorithm names and learning paradigms they tackle.

Table 1. List of few works on Twitter Sentiment Analysis

Ref.	Algorithm Name	Learning Paradigm	Sentiment Polarity
(Go *et al.*, 2009)	<no name>	Machine Learning based classification. Naïve Bayes, Max Entropy, Support Vector Machines	2 class-Positive, Negative
(Pak *et al.*, 2010)	New Classifier based on N-gram and Part of Speech tags using new metric: salience	Multinomial Naïve Bayes Classifier for Text Classification	3 class- Positive, Negative, Neutral
(Agarwal *et al.*, 2011)	Tree Kernel	Text Classification	2class- Positive, Negative 3 class- Positive, Negative, Neutral
(Kouloumpis *et al.*, 2011)	<no name>	Supervised Text Classification	3 class- Positive, Negative, Neutral
(Wang *et al.*, 2011)	Loopy Belief Propagation (LBP), Relaxation Labeling (RL) and Iterative Classification Algorithm (ICA)	Hashtag Graph Model using Support Vector Machine for Hashtag text Classification	2 class- neutral, subjective 2 class- positive, negative
(Bae *et al.*, 2012)	PN influence measure	Measuring Influence, Time Series	2 class-Positive, Negative, Neutral
(Gokulakrishnan *et al.*, 2012)	<no name>	Naïve Bayes, Random Forest, Support Vector Machines, Sequential Mining Optimization, J48 decision tree	2 class- Positive, Negative, 2 class – Relevant/Irrelevant
(Kumar *et al.*, 2012)	<no-name>	Semantic Orientation, Text Classification	3 class- Positive, Negative, Neutral
(Liu *et al.*, 2012)	Emoticon Smoothed Language Model	Maximum Likelihood Estimate (Probability)	3 class- Positive, Negative, Neutral
(Saif *et al.*, 2012)	New feature: Semantics	Multinomial Naïve Bayes Classifier for Text Classification	Aspect Based 2 class: Positive, Negative
(Ghiassi *et al.*, 2013)	Twitter specific lexicon	Supervised Text Classification	5 class-Strongly Positive, Mildly Positive, Neutral, Mildly Negative, Strongly Negative
(Hassan *et al.*, 2013)	Bootstrapping ensemble framework	Time Series, Text Classification	3 class- Positive, Negative, Neutral
(Kontopoulos *et al.*, 2013)	Ontology	Formal Concept Analysis	Aspect Based
(Neethu *et al.*, 2013)	New feature vector <no-name>	Naïve Bayes, Max Entropy, Support Vector Machines, Ensemble Classifiers for Text Classification	2 class-Positive, Negative
(Srivastava *et al.*, 2013)	Opinion Word Lexicon (OWL), Fuzzy Inference System	Fuzzy Inferencing, linguistic Hedges	2 class- positive, negative
(Haque *et al.*, 2014)	<no-name>	Fuzzy Logic for Text Classification	2 class- objective, subjective (positive, negative)

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

			6 class-Strong Pos, Pos, Weak Pos,Weak Neg, Neg, Strong Neg.
(Gautam *et al.*, 2014)	Semantic Analysis (WordNet)	Machine Learning based classification	2 class-Positive, Negative
(Liu *et al.*, 2015)	Topic Adaptive Sentiment Classification	Decision Tree, Support Vector Machines, Random Forest	3 class- Positive, Negative, Neutral
(Rosenthal *et al.*, 2015)	Subtask A: Contextual Polarity Disambiguation, B: Message Polarity Classification, C: Topic-Based Message Polarity Classification, D: Detecting Trend Towards a Topic, E: Degree of Prior Polarity	Machine Learning approaches for Text Classification	3 class- Positive, Negative, Neutral 5 class-Strongly Positive, Weakly Positive, Neutral, Weakly Negative, Strongly Negative
(Severyn *et al.*, 2015)	Subtask A: Contextual Polarity Disambiguation, B: Message Polarity Classification	Deep Neural Network for Text Classification	3 class- Positive, Negative, Neutral
(Nakov *et al.*, 2016)	Subtask A: Message Polarity Classification, B: Tweet classification (2-point scale), C: Tweet classification (5point scale), D: Tweet quantification (2-point scale), E: Tweet quantification (5-point scale)	Machine Learning approaches for Text Classification	2 class-Positive, Negative 3 class- Positive, Negative, Neutral 5 class-Highly Positive, Positive, Neutral, Negative, Highly Negative
(Saif *et al.*, 2016)	SentiCircles, Created Stanford Sentiment- Gold Stanford dataset	Text Classification	3 class- Positive, Negative, Neutral
(Howells *et al.*, 2017)	<no-name>	Fuzzy Logic for Text Classification	5 class-strongly positive, positive, neutral, negative, strongly negative
(Chen *et al.*, 2018)	Content Based Sequential Opinion Influence Framework	Prediction Models: Degroot, Flocking, AsLM, Voter, Coupled Markov Chain.	3 class- Positive, Negative, Neutral

3 Proposed Fuzzy Rule System for Sentiment Analysis

In this section, we present the details of the proposed fuzzy logic-based model. Fig. 1 describes the framework of a fuzzy logic-based model. Fuzzification is the process of making a crisp quantity 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 commonly used.

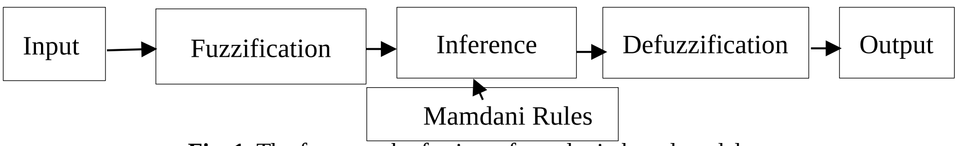


Fig. 1. The framework of using a fuzzy logic-based model

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

In the field of artificial intelligence, possibly the easiest way to represent the human knowledge is to transform it into natural language expressions in the format of IF-THEN rules (Zadeh, 1975). The fuzzy rule-based system is most useful in modeling some complex systems that can be observed by humans because they make use of linguistic variables as their antecedents and consequents. These linguistic variables can be naturally represented by 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:

$$\text{Rule } R_j: \text{ IF A is } A_1^j \text{ and B is } B_1^j \text{ THEN C is } C_1^j,\ j = 1,2,..,r \tag{1}$$

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 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 of a fuzzy quantity to a precise quantity, opposite to fuzzification. Some of the defuzzification methods are: centroid, bisector, mean of maximum (MOM), smallest of maximum (SOM) 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 preprocessing, use of sentiment lexicon and fuzzy rule system for sentiment polarity classification.

3.1 Text Pre-processing

The social media text is of limited size. In 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 containing "#" (hashtags), usually represent an emotion, thought or opinion about the tweet's topic, so we remove only the "#".

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

3.2 Use of sentiment lexicon

A sentiment lexicon is a list of lexical features (e.g., words) which are generally labeled according to their semantic orientation as either positive or negative. In our paper, we have investigated three different sentiment lexicons: SentiWordNet (Baccianella *et al.*, 2010), AFINN (Nielsen *et al.*, 2011) and VADER (Gilbert *et al.*, 2014) in isolation with each other. The pre-processed text is used along with these lexicons to compute the positive and negative score 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 ranges from 0.0 to 1.0, and their sum 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). Word-sense disambiguation refers to the process of identifying which sense of a word is used in a sentence when the word has multiple meanings (i.e. its contextual meaning). The aim of the WSD process consists of determining 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 positive and negative score (eq. (2) and eq. (3)) computed using WSD, that can be interpreted as a fuzzy membership pertaining to the fuzzy sets *Pos* and *Neg* (eq. (4) and eq. (5)). The words which are having higher positive score than negative score in a tweet are summed up to compute the positive score (*TweetPos*) of the tweet (eq. (6)). Similarly, words which are having higher negative score than positive score in a tweet are summed up to compute the negative score (*TweetNeg*) of the tweet (eq. (7)). These scores are computed for all tweets.

$$\mu_{Pos}(a) = syn.pos\_score() \tag{2}$$

$$\mu_{Neg}(a) = syn.neg\_score() \tag{3}$$

$$Pos = \{(a, \mu_{Pos}(a)\}, a \in X_i \tag{4}$$

$$Neg = \{(a, \mu_{Neg}(a)\}, a \in X_i \tag{5}$$

$$\begin{aligned} &if\ (\mu_{Pos}(a) > 0\ \&\ \mu_{Pos}(a) > \mu_{Neg}(a)) \\ &then(TweetPos = \sum_{a=1}^{m} \mu_{Pos}(a)) \end{aligned} \tag{6}$$

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

$$\begin{aligned}&if\ (\mu_{Neg}(a) > 0\ \&\ \mu_{Neg}(a) > \mu_{Pos}(a))\\&then(TweetNeg = \sum_{a=1}^{m} \mu_{Neg}(a))\end{aligned} \tag{7}$$

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 (Nielsen *et al.*, 2011). This lexicon is equipped to handle modern day tweets due to its inclusion of Internet slang and obscene words. It has been created specially 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 (eq. (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 positive and negative score can be interpreted as a fuzzy membership pertaining to the fuzzy sets Pos and Neg (eq. (4) and eq. (5)). The positive words are summed up to compute the positive score of the tweet (eq. (11)); similarly, the negative score is computed for each tweet (eq. (12)).

$$\mu(a) = af.score(a) \tag{8}$$

$$if\ (\mu(a) > 0)then(\mu_{Pos}(a) = \mu(a)) \tag{9}$$

$$if\ (\mu(a) < 0)then(\mu_{Neg}(a) = -\mu(a) \tag{10}$$

$$TweetPos = \sum_{a=1}^{m} \mu_{Pos}(a) \tag{11}$$

$$TweetNeg = \sum_{a=1}^{m} \mu_{Neg}(a) \tag{12}$$

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based Sentiment Analysis tool that is specifically accustomed to sentiments expressed in social media, it was created in 2014 (Gilbert *et al.*, 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*) score of a tweet as output.

3.3 Fuzzy Rule System

We have used one of the popularly used fuzzy inference technique called Mamdani fuzzy model. The Mamdani style fuzzy inference process is performed in four steps:

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

Fuzzification of input variables, Rule evaluation (inference), Aggregation of the rule outputs and Defuzzification.

3.3.1 Fuzzification: The positive and negative score of each tweet obtained from the second phase is fuzzified using 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 triangular function defined by a lower limit *d*, an upper limit *f*, and an intermediate value *e*, where $d < e < f$:

$$\mu_S(x) = \begin{cases} 0, x <= d \\ (x-d)/(e-d), d < x <= e \\ (f-x)/(f-e), e < x <= f \\ 0, x >= f \end{cases} \tag{13}$$

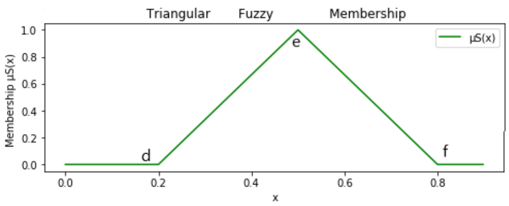


Fig. 2. Triangular Fuzzy Membership

The triangular fuzzy membership is graphically presented in Fig.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:

$$mid = (\min + \max)/2 \tag{14}$$

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.3.

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

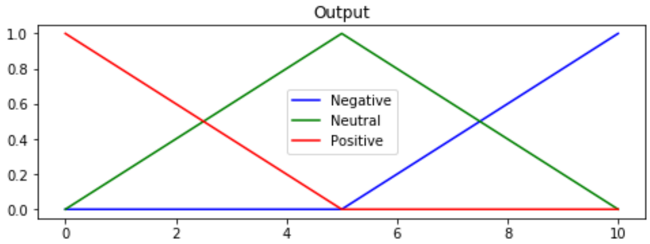


Fig. 3. Triangular Fuzzy Membership sets for output variables

3.3.2 Formulating the Rule-Base: The novelty of this paper is the proposal of nine rules, described in Table 2. Fig. 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 the higher score (positive or negative) indicates the sentiment. In case of common scores, the sentiment is neutral. The rule evaluation is done on the basis of the Table 2 and Fig .4.

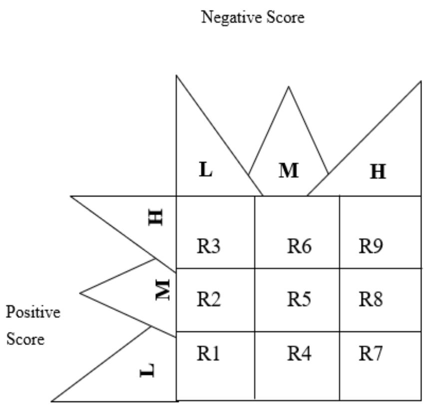


Fig. 4. Visualization of nine rules

Table 2. 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

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

R5	Medium	Medium	Neutral
R6	High	Medium	Positive
R7	Low	High	Negative
R8	Medium	High	Negative
R9	High	High	Neutral

$$w_{R1} = pos\_low \wedge neg\_low \quad (15)$$
$$w_{R2} = pos\_med \wedge neg\_low \quad (16)$$
$$w_{R3} = pos\_high \wedge neg\_low \quad (17)$$
$$w_{R4} = pos\_low \wedge neg\_med \quad (18)$$
$$w_{R5} = pos\_med \wedge neg\_med \quad (19)$$
$$w_{R6} = pos\_hi \wedge neg\_med \quad (20)$$
$$w_{R7} = pos\_low \wedge neg\_high \quad (21)$$
$$w_{R8} = pos\_med \wedge neg\_high \quad (22)$$
$$w_{R9} = pos\_high \wedge neg\_high \quad (23)$$

These equations depict the nine rules $w_{R1} \ldots w_{R9}$ 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.

3.3.3 Aggregation of Rule outputs:

$$w_{neg} = w_{R4} \vee w_{R7} \vee w_{R8} \quad (24)$$

$$w_{neu} = w_{R1} \vee w_{R5} \vee w_{R9} \quad (25)$$

$$op\_activation\_low = w_{neg} \wedge op\_neg \quad (27)$$

$$op\_activation\_med = w_{neu} \wedge op\_neu \quad (28)$$

$$op\_activation\_high = w_{pos} \wedge op\_pos \quad (29)$$

$$aggregated = op\_activation\_low \cup op\_activation\_med \cup op\_activation\_high \quad (30)$$

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

In equations (24-26) w_{neg} depicts the overall firing strength or degree of fulfillment 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 degree to which the antecedent part of the rule is satisfied (Jang *et al.*, 1997). In equations (27-29) o*p_neg*, *op_neu* and *op_pos* are the MFs of consequent parts of respective rules (equations 15-23). 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 equations (27-29). Overall output MF is obtained by aggregating the induced consequent MFs using union operator in eq. (30).

3.3.4 Defuzzification: The last step in fuzzy rule system is defuzzification. In our paper, we have implemented the centroid defuzzification method as it yields reliable results (Jang *et al.*, 1997). It returns the center of area (COA) under the curve (Hellendoorn *et al.*, 1993). This method provides a crisp value based on the center of gravity of the fuzzy set. The total area of the membership function distribution used to represent the combined control action is divided into a number of sub-areas. The area and the center of gravity or centroid of each sub-area is calculated and then the summation of all these sub-areas is taken to find the defuzzified value for a discrete fuzzy set. The aggregated output (μ_A) computed in eq. (30) is used to calculate the defuzzified output in eq. (31), where z indicates sample value in output variable, *x_op* described in section 3.3.1.

$$COA = \frac{\sum z\mu_A(z)}{\sum \mu_A(z)} \tag{31}$$

Finally, the defuzzified output is checked for different ranges to classify the tweet according to its polarity: Negative, Neutral or Positive class in eq. (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*.

$$Output = \begin{Bmatrix} Negative, 0 < COA < 3.3 \\ Neutral, 3.3 < COA < 6.7 \\ Positive, 6.7 < COA < 10 \end{Bmatrix} \tag{32}$$

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

1. In (Srivastava *et al*., 2013) they have constructed their own lexicon Opinion Words Lexicon (OWL) by performing some modifications on SentiWordNet data. This

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

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 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 union of output activation level: 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.

4 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 given at: https://github.com/SrishtiVashishtha. Most of the papers use the Twitter API to extract tweets but we have used publicly available datasets. In this paper we have used a total of nine benchmark datasets: The Sanders Twitter Dataset, The Nuclear

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

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 Gilbert & Hutto (Gilbert *et al.*, 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 introduced 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 3. Furthermore, we have used various SemEval twitter datasets: SemEval-2017 Task 4, subtask A decides whether a given tweet expresses 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, determine whether it expresses a positive, a negative, or a neutral/objective sentiment (Rosenthal *et al*.,2015). The last twitter dataset is obtained from (Gilbert *et al.*, 2014). Table 5 represents the distribution of tweets in these datasets.

We have used three different sentiment lexicons: SentiWordNet (Baccianella *et al.*,2010), AFINN (Nielsen *et al.*, 2011) and VADER (Gilbert *et al.*, 2014); these are described in section 3.2. We have compared our model with a supervised approach involving classification using Support Vector Machines (SVM) classifier (Vapnik *et al.*, 1995). SVM is implemented using sklearn package in python. The parameters of SVM: C=1.0, auto mode for gamma and linear kernel are selected for best results. We have executed 5-fold SVM with 70% training and 30% test sets of datasets.

Table 3. Dataset Distribution of different datasets.

	Sanders Twitter[1]	**Nuclear Twitter[1]**	**Apple Twitter[1]**	**STS-Test[1]**	**Sentiment140[1]**
Positive	519	10	423	182	248576
Negative	572	19	1219	177	799999
Neutral	2333	161	2162	139	Null
Total	3424	190	3804	498	1048575

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

Table 4. Dataset Distribution of SemEval and Gilbert datasets.

	SemEval 2017[1]	SemEval 2016[1]	SemEval 2015[1]	Gilbert Tweets[1]
Positive	2375	5157	4377	2742
Negative	3972	1225	1745	1219
Neutral	5937	2667	5593	239
Total	12284	9049	11715	4200

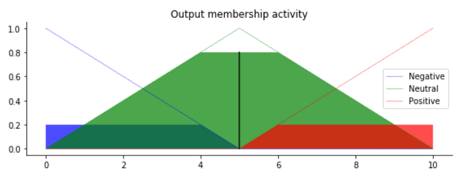
5 Results and Discussion

5.1 Processing of a single tweet

In this section, we present how 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 (Gilbert *et al.*, 2014) is shown in Fig 5. 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. 5. Processing of a sample tweet using VADER lexicon



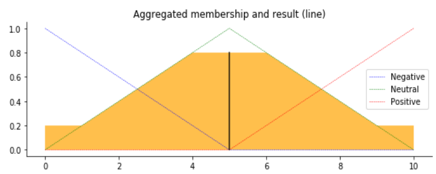


Fig. 6. Output showing different emotions of the tweet **Fig. 7.** Aggregated Output of the tweet

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 (equations 15-23) are applied. The overall firing

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

strength of tweet for different emotion classes are evaluated using equations (24-26). Fig.6 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 equations (27-29). Fig.7 depicts the aggregated output membership ($\mu_A(x)$) computed in eq. (30). The area under the aggregated output is used for centroid defuzzification in eq. (31). The defuzzified output equal to 4.81 is shown as bold straight line. Finally, the sentiment of tweet is evaluated as 'Neutral' using eq. (32). We can check the polarity of the tweet from dataset, and it turns out to be same.

5.2 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.8 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.9-11) detect the correct sentiment (neutral) by the proposed scheme. Another sample tweet (3420) of Sanders dataset is depicted in Fig.12. Here with the help of VADER lexicon (Fig.15) correct sentiment class is detected (positive) while SentiWordNet (Fig.13) and AFINN (Fig.14) detect the wrong sentiment class (neutral). Further we have displayed the execution time of all methods with different lexicon-dataset combinations in Table 9 and Table 10. We can observe 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 9 and Table 10.

```
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. 8. 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. 9. 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. 10. 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. 11. Tweet Processing by VADER

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

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

Fig. 12. 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. 13. 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. 14. 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. 15. Tweet Processing by VADER

5.3 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) Gilbert *et al.* iv) Nielsen *et al.* and one state-of-the-art-method for supervised machine learning involving Support Vector Machines (SVM) classifier. The first two methods have used SentiWordNet (Baccianella *et al.*, 2010) lexicon, the third method has implemented simple Sentiment Analysis using VADER lexicon (Gilbert *et al.*, 2014) and the last method has used AFINN lexicon (Nielsen *et al.*, 2011) to perform Sentiment Analysis. Classifying sentiment of tweets using supervised learning SVM method was investigated by (Go *et al.*, 2009). We have executed SVM using Term Frequency- Inverse Document Frequency (TF-IDF) as text features. We have implemented our fuzzy rule based-method using all the lexicons 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 5 and Table 6.

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 (Gilbert *et al.*, 2014) 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 (Gilbert *et al.*, 2014) 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

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

(Nielsen *et al.*, 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 Gilbert *et al.* 's method has highest F1-scores.

VADER lexicon (Gilbert *et al.*, 2014) has performed the best because this lexicon is best suited for social media posts. It handles emojis, slangs, emoticons, acronyms very well and evaluates the emoticons contained in 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 but is constructed from a generalizable, valence-based, human-curated gold standard sentiment lexicon, and hence evaluates tweets more accurately. The next best lexicon is AFINN and then SentiWordNet lexicon.

Table 5. F1- Scores of Different methods, Lexicons and Twitter datasets.

Lexicons	**Methods**	**Sanders**[54]		**Nuclear**[45]		**Apple**[1]		**STS Test**[17]		**Sentiment 140**[17]	
		F1-Micro	**F1-Macro**	**F1-Micro**	**F1-Macro**	**F1-Micro**	**F1-Macro**	**F1-Micro**	**F1-Macro**	**F1-Micro**	**F1-Macro**
SentiWordNet[4]	Cavalcanti[9]	0.255	0.266	0.074	0.110	0.307	0.287	0.502	0.423	0.600	0.38
	Ortega[46]	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[44]	Simple SA[44]	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[16]	Simple SA[16]	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

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

Table 6. F1- Scores of Different methods, Lexicons and SemEval-Gilbert datasets.

Lexicons	Methods	SemEval 2017[51]		SemEval 2016[42]		SemEval 2015[50]		Gilbert Tweets[16]	
		F1-Micro	F1-Macro	F1-Micro	F1-Macro	F1-Micro	F1-Macro	F1-Micro	F1-Macro
SentiWordNet[4]	Cavalcanti[9]	0.358	0.334	0.436	0.314	0.372	0.309	0.549	0.406
	Ortega[46]	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[44]	Simple SA[44]	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[16]	Simple SA[16]	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 7. Precision Recall Macro Scores of Different methods, Lexicons and Twitter datasets.

Lexicons	Methods	Sanders[54]		Nuclear[45]		Apple[1]		STS Test[17]		Sentiment 140[17]	
		Preci-sion	Recall	Preci-sion	Recall	Preci-sion	Recall	Precision	Recall	Precision	Recall
SentiWordNet[4]	Cavalcanti[9]	0.410	0.430	0.079	0.372	0.409	0.457	0.524	0.469	0.4	0.425
	Ortega[46]	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[44]	Simple SA[44]	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[16]	Simple SA[16]	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 8. Precision Recall Macro Scores of Different methods, Lexicons and SemEval-Gilbert datasets.

Lexicons	Methods	SemEval 2017[51]		SemEval 2016[42]		SemEval 2015[50]		Gilbert Tweets[16]	
		Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

SentiWordNet[4]	Cavalcanti[9]	0.458	0.454	0.345	0.392	0.457	0.458	0.467	0.447
	Ortega[46]	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[44]	Simple SA[44]	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[16]	Simple SA[16]	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 7 and Table 8 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 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 highest precision in Sanders, SemEval 2017 and SemEval 2015 datasets; and highest recall in SemEval 2016. Nielson *et al.* 's method performed the best in STS test dataset in both recall and precision; highest recall in Nuclear, SemEval 2017 and SemEval 2015 datasets; highest precision in SemEval 2016. In sentiment 140 dataset, the highest scores were scored by Gilbert *et al.* 's method. This method gained highest recall in Sanders and Apple dataset as well. We can conclude that AFINN and VADER lexicon performed better compared to SentiWordNet lexicon.

Table 9. Execution Time (in sec) of Different methods, Lexicons and Twitter datasets.

Lexicons	**Methods**	**Sanders[54]**	**Nuclear[45]**	**Apple[1]**	**STS Test[17]**	**Sentiment 140[17]**
SentiWordNet[4]	Cavalcanti[9]	7.86	0.55	19.79	1.077	2042.85
	Ortega[46]	222.87	2.19	393.5	6.32	15000
	Fuzzy Rules	15.81	1.02	18.46	2.87	8924.22
AFINN[44]	Simple SA[44]	4.91	0.51	5.106	0.61	2511.15
	Fuzzy Rules	9.95	0.68	11.42	1.32	2834.90
VADER[16]	Simple SA[16]	1.75	0.27	2.36	0.328	936.93
	Fuzzy Rules	7.65	0.46	5.6	1.18	2111.85

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

Table 10. Execution Time (in sec) of Different methods, Lexicons & SemEval-Gilbert datasets.

Lexicons	Methods	SemEval 2017[51]	SemEval 2016[42]	SemEval 2015[50]	Gilbert Tweets[16]
SentiWordNet[4]	Cavalcanti[9]	31.87	23.22	44.89	10.04
	Ortega[46]	4599	1584.10	2781	277.5
	Fuzzy Rules	54.9	40.75	57.61	18.43
AFINN[44]	Simple SA[44]	4.80	27.8	31.96	14.3
	Fuzzy Rules	11.7	12.8	40.2	9.62
VADER[16]	Simple SA[16]	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 9 and Table 10. Execution time depends upon various factors: size of 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 the maximum time compared to other methods. Our fuzzy rule method with AFINN or VADER lexicon performs faster compared to methods which implement SentiWordNet lexicon.

Table 11. Performance of Supervised Method (SVM) on some Twitter datasets.

Datasets	F1- Micro	F1-Macro	Precision	Recall	Execution Time
Sanders[54]	0.694	0.822	0.361	0.327	5500
Nuclear[45]	0.831	0.309	0.366	0.335	58.13
Apple[1]	0.733	0.793	0.547	0.554	7390
STS Test[17]	0.529	0.697	0.489	0.441	293
Gilbert Tweets[16]	0.732	0.549	0.422	0.418	3029

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

We have implemented one state-of-the-art-method for supervised machine learning algorithm: Support Vector Machines (SVM). We have executed 5-fold SVM with 70% training and 30% test sets of datasets. We have not used any lexicon in the supervised approach: SVM classification. We have used TF-IDF vectorization features for classification. We have executed SVM on some of the datasets, the F1- scores, Precision, Recall and Execution time of these datasets are presented in Table 11. We can compare and observe that our unsupervised fuzzy rule-based method with VADER lexicon has scored 0.865 in Gilbert Tweets, 0.842 in Nuclear and 0.642 in STS Test, these scores are higher than the supervised SVM classifier. SVM method scored 0.732 in Gilbert Tweets, 0.831 in Nuclear and 0.529 in STS Test datasets. While our method with VADER lexicon scored 0.686 F1-Micro score in Sanders dataset which is comparable to 0.694 for SVM classifier. Our fuzzy rule method with VADER lexicon has scored higher Precision- Recall scores for Sanders (0.583, 0.421), Nuclear (0.616, 0.347), Apple (0.69, 0.426), STS-Test (0.705, 0.655) and Gilbert Tweets (0.748, 0.881) compared to SVM. SVM has scored following Precision-Recall scores: Sanders (0.361, 0.327), Nuclear (0.366, 0.335), Apple (0.547, 0.554), STS-Test (0.489, 0.441) and Gilbert Tweets (0.422, 0.418). Comparison of execution time for both these methods reveal that our method with VADER lexicon takes only 7.65 secs, 0.46 sec, 5 secs, 1.18 sec and 6.89 secs for Sanders, Nuclear, Apple, STS Test and Gilbert Tweets while SVM takes more time: 5500secs, 58.13 secs, 7390 secs, 293 secs and 3092 secs respectively. We can conclude that our unsupervised fuzzy rule-based method with VADER lexicon has performed much better than supervised machine learning (SVM). It is acknowledged that though supervised learning using deep neural networks may result in higher classification scores (for SemEval-2017 dataset: 0.685 F1-Macro score (Cliche, 2017) and 0.675 F1-Macro score (Baziotis *et al*., 2017)), they involve huge training time and large number of training samples. This criterion may not be met by some of the datasets used in our experiments. Our method is unsupervised and requires no training time and is not dependent on size of 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. A dataset that takes a fraction of a second to analyze with VADER Lexicon can take hours when using more complex supervised models like SVM.

6 Conclusion

In this paper, we have proposed a fuzzy rule-based approach for Sentiment Analysis of social media posts specifically for twitter datasets. The novelty of this paper is i) the formulation of nine fuzzy rules to evaluate the sentiment class of tweets, ii) the proposed approach is unsupervised and can be adapted to any lexicon and iii) to any dataset (two-class or three-class). 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) Deals with reasoning and gives closer views to the exact sentiment values.

We have implemented our proposed method using three different lexicons: SentiWordNet (Baccianella *et al.*, 2010), AFINN (Nielsen *et al.*, 2011) and VADER (Gilbert *et al.*, 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, reveal 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 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 is achieved by VADER lexicon in Gilbert Tweets and Nuclear Twitter datasets respectively. Moreover, the methods which implement VADER lexicon execute in least time while the methods which implement SentiWordNet lexicon take maximum time in execution. In terms of the precision and recall scores for 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, slangs, emoticons, acronyms very well and evaluates the emoticons contained in text. Tremendous benefits can be obtained by using VADER in micro-blogging websites wherein the text data is of complex nature. We have summarized our fuzzy rule based proposed approach for a single tweet in a flowchart in Fig.16.

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. We can extend our method by incorporating fuzzy inferencing into deep neural network models with comparison to state-of-the-art in deep learning.

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

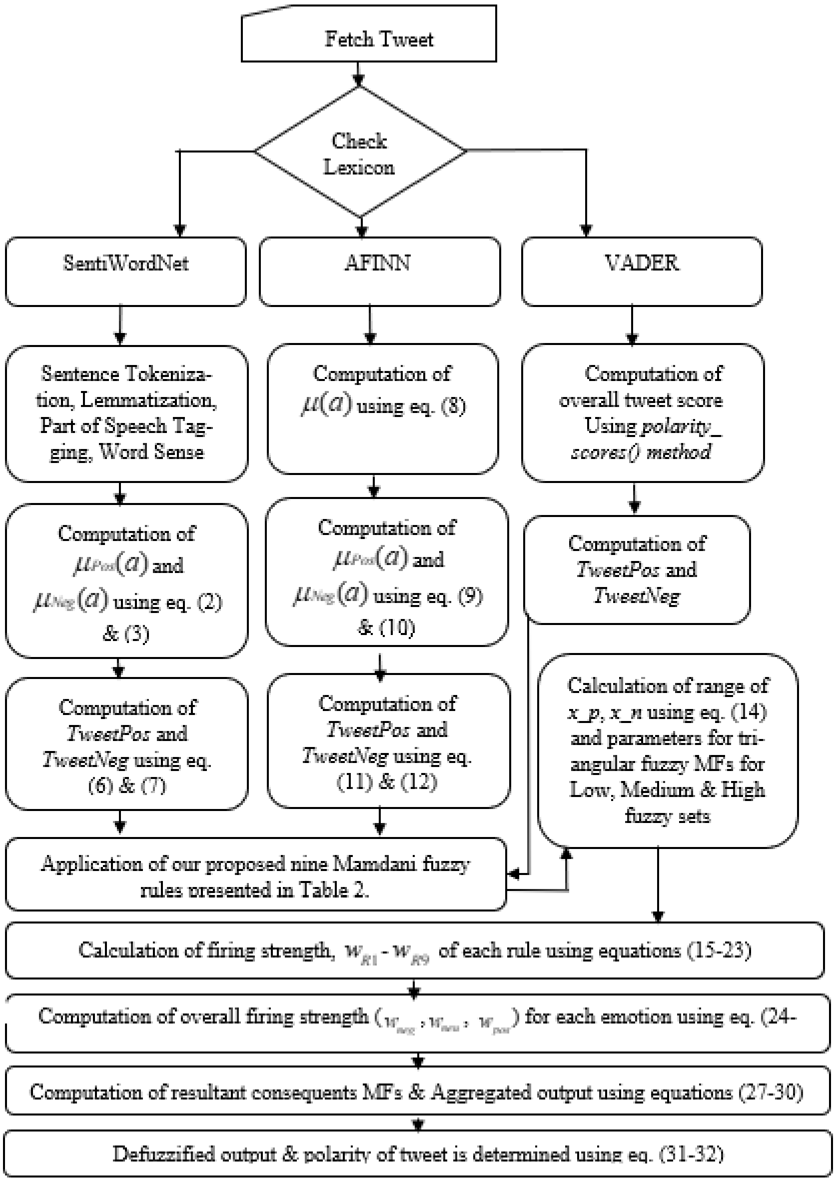


Fig. 16. Processing of a single tweet by our fuzzy system

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

References

1. Apple Twitter Dataset, Retrieed Jan 31,2019 from https://data.world/crowdflower/apple-twitter-sentiment.
2. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of twitter data. In Proceedings of the Workshop on Language in Social Media (LSM 2011) (pp. 30-38).
3. Banerjee, Satanjeev, and Ted Pedersen (2002). "An adapted Lesk algorithm for word sense disambiguation using WordNet." In International conference on intelligent text processing and computational linguistics, pp. 136-145. Springer, Berlin, Heidelberg, 2002.
4. Baccianella, S., A. Esuli, and F. Sebastiani, 2010, May. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In Lrec (Vol. 10, No. 2010, pp. 2200-2204).
5. Bae, Y., and Lee, H. (2012). Sentiment analysis of twitter audiences: Measuring the positive or negative influence of popular twitterers. Journal of the American Society for Information Sci C:\Users\srish\Desktop\New folderence and Technology, 63(12), 2521-2535.
6. Barnaghi, Peiman, Parsa Ghaffari, and John G. Breslin(2016). "Opinion mining and sentiment polarity on twitter and correlation between events and sentiment." In Big Data Computing Service and Applications (BigDataService), 2016 IEEE Second International Conference on, pp. 52-57. IEEE, 2016.
7. Baziotis, C., Pelekis, N. and Doulkeridis, C. (2017, August). Datastories at semeval-2017 task 4: Deep lstm with attention for message-level and topic-based sentiment analysis. In Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017) (pp. 747-754).
8. Bird, S., E. Klein, and E. Loper (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc.", 2009.
9. Cavalcanti, D. C., Prudêncio R. B. C., Pradhan S.S., Shah J. Y., Pietrobon R. S (2011).: "Good to be bad? Distinguishing between positive and negative citations in scientific impact." In: *Tools with Artificial Intelligence (ICTAI), 2011 23rd IEEE International Conference on*, pp. 156-162. IEEE, 2011.
10. Chang, Yu-Chuan, Shyi-Ming Chen, and Churn-Jung Liau (2008). "Fuzzy interpolative reasoning for sparse fuzzy-rule-based systems based on the areas of fuzzy sets." IEEE Transactions on Fuzzy Systems 16, no. 5 (2008): 1285.
11. Chen, Chengyao, Zhitao Wang, and Wenjie Li ( ). "Tracking Dynamics of Opinion Behaviors with a Content-based Sequential Opinion Influence Model." IEEE Transactions on Affective Computing 1 (2018): 1-1.
12. Cliche, M. (2017). BB_twtr at SemEval-2017 task 4: twitter sentiment analysis with CNNs and LSTMs. arXiv preprint arXiv:1704.06125.
13. Duţu, Liviu-Cristian, Gilles Mauris, and Philippe Bolon (2018). "A fast and accurate rule-base generation method for Mamdani fuzzy systems." IEEE Transactions on Fuzzy Systems 26, no. 2 (2018): 715-733.
14. Gautam, G., & Yadav, D. (2014, August). Sentiment analysis of twitter data using machine learning approaches and semantic analysis. In 2014 Seventh International Conference on Contemporary Computing (IC3) (pp. 437-442). IEEE.

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

15. 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.
16. Giachanou, A., & Crestani, F. (2016). Like it or not: A survey of twitter sentiment analysis methods. ACM Computing Surveys (CSUR), 49(2), 28.
17. Gilbert, CJ Hutto Eric (2014). "VADER: A parsimonious rule-based model for sentiment analysis of social media text." In Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. VADER. hutto. pdf. 2014.
18. Go, A., Bhayani, R., Huang, L (2009).: Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford (2009).
19. Gokulakrishnan, B., Priyanthan, P., Ragavan, T., Prasath, N., & Perera, A. (2012, December). Opinion mining and sentiment analysis on a twitter data stream. In International Conference on Advances in ICT for Emerging Regions (ICTer2012) (pp. 182-188). IEEE.
20. Hamdan, Hussam, Frederic Béchet, and Patrice Bellot (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), vol. 2, pp. 455-459. 2013.
21. Haque, M., 2014. Sentiment analysis by using fuzzy logic. arXiv preprint arXiv:1403.3185.
22. Hassan, A., Abbasi, A., & Zeng, D. (2013, September). Twitter sentiment analysis: A bootstrap ensemble framework. In 2013 International Conference on Social Computing (pp. 357-364). IEEE.
23. Hellendoorn, H. and C. Thomax (1993), "Defuzzification in fuzzy controllers," Journal of Intelligent & Fuzzy Systems, vol. 1, no. 2,pp.109-123, 1993.
24. Howells, K. and Ertugan, A., 2017. Applying fuzzy logic for sentiment analysis of social media network data in marketing. Procedia computer science, 120, pp.664-670.
25. Ishibuchi, Hisao, and Tomoharu Nakashima (2001). "Effect of rule weights in fuzzy rule-based classification systems." IEEE Transactions on Fuzzy Systems 9, no. 4 (2001): 506-515.
26. Ishibuchi, Hisao, and Takashi Yamamoto (2005). "Rule weight specification in fuzzy rule-based classification systems." IEEE transactions on fuzzy systems 13, no. 4 (2005): 428-435.
27. Jain, Anurag P., and Vijay D. Katkar (2015). "Sentiments analysis of Twitter data using data mining." In Information Processing (ICIP), 2015 International Conference on, pp. 807-810. IEEE, 2015.
28. Jang, J.S.R., Sun, C.T. and 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), pp.1482-1484.
29. Jefferson, C., Liu, H. and Cocea, M., (2017, July). Fuzzy approach for sentiment analysis. In 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (pp. 1-6). IEEE.
30. Kontopoulos, E., Berberidis, C., Dergiades, T., & Bassiliades, N. (2013). Ontology-based sentiment analysis of twitter posts. Expert systems with applications, 40(10), 4065-4074.
31. Kouloumpis, E., Wilson, T., & Moore, J. (2011, July). Twitter sentiment analysis: The good the bad and the omg!. In Fifth International AAAI conference on weblogs and social media.
32. Kumar, A., & Sebastian, T. M. (2012). Sentiment analysis on twitter. International Journal of Computer Science Issues (IJCSI), 9(4), 372.

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

33. Liu, K. L., Li, W. J., & Guo, M. (2012, July). Emoticon smoothed language models for twitter sentiment analysis. In Twenty-sixth aAAI conference on artificial intelligence.
34. Liu, Shenghua, Xueqi Cheng, Fuxin Li, and Fangtao Li(2015). "TASC: topic-adaptive sentiment classification on dynamic tweets." IEEE transactions on knowledge and data engineering 27, no. 6 (2015): 1696-1709.
35. Liu, Han, and Mihaela Cocea (2017). "Fuzzy rule based systems for interpretable sentiment analysis." In Advanced Computational Intelligence (ICACI), 2017 Ninth International Conference on, pp. 129-136. IEEE, 2017.
36. López, Victoria, Sara del Río, José Manuel Benítez, and Francisco Herrera (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.
37. Lucas, Gale M., Jonathan Gratch, Nikolaos Malandrakis, Evan Szablowski, Eli Fessler, and Jeffrey Nichols (2017). "GOAALLL!: Using sentiment in the world cup to explore theories of emotion." Image and Vision Computing 65 (2017): 58-65.
38. Mamdani, E.H. and S. Assilian (1975), "An experiment in linguistic synthesis with a fuzzy logic controller," Int. J. Man-Mach. Stud., vol. 7, no. 1, pp. 1–13, 1975.
39. Márquez, Francisco Alfredo, Antonio Peregrín, and Francisco Herrera (2007). "Cooperative evolutionary learning of linguistic fuzzy rules and parametric aggregation connectors for Mamdani fuzzy systems." IEEE Transactions on Fuzzy Systems 15, no. 6 (2007): 1162-1178.
40. Martínez-Cámara, E., Martín-Valdivia, M. T., Urena-López, L. A., & Montejo-Ráez, A. R. (2014). Sentiment analysis in Twitter. Natural Language Engineering, 20(1), 1-28.
41. Mejova, Y. (2009), "Sentiment analysis: An overview," Comprehensive exam paper, available on http://www. cs. uiowa. edu/˜ymejova/publications/CompsYelenaMejova. pdf [2010-02-03]
42. Montoro, Andres, Jose A. Olivas, Arturo Peralta, Francisco P. Romero, and Jesus Serrano-Guerrero (2018). "An ANEW based Fuzzy Sentiment Analysis Model." In 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1-7. IEEE, 2018.
43. Nakov, P., Ritter, A., Rosenthal, S., Sebastiani, F., & Stoyanov, V. (2016). SemEval-2016 task 4: Sentiment analysis in Twitter. In Proceedings of the 10th international workshop on semantic evaluation (semeval-2016) (pp. 1-18).
44. Neethu, M. S., and R. Rajasree (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.
45. Nielsen, Finn Årup (2011). "Afinn." Richard Petersens Plads, Building 321 (2011).
46. Nuclear Twitter Dataset, Retrieved Jan 31, 2019 from https://data.world/crowdflower/emotions-about-nuclear-energy.
47. Ortega R., Fonseca A., Gutierrez Y., Montoyo A (2013).: "Ssa-uo: unsupervised twitter sentiment analysis," In: *Second joint conference on lexical and computational semantics (* SEM)*, vol. 2, 2013, pp. 501–507.
48. Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In LREc (Vol. 10, No. 2010, pp. 1320-1326).
49. Park, Chae Won, and Dae Ryong Seo (2018). "Sentiment analysis of Twitter corpus related to artificial intelligence assistants." In 2018 5th International Conference on Industrial Engineering and Applications (ICIEA), pp. 495-498. IEEE, 2018.

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

50. Parveen, Huma, and Shikha Pandey (2016). "Sentiment analysis on Twitter Data-set using Naive Bayes algorithm." In Applied and Theoretical Computing and Communication Technology (iCATccT), 2016 2nd International Conference on, pp. 416-419. IEEE, 2016.
51. Rosenthal, S., Nakov, P., Kiritchenko, S., Mohammad, S., Ritter, A., & Stoyanov, V. (2015). Semeval-2015 task 10: Sentiment analysis in twitter. In Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015) (pp. 451-463).
52. Rosenthal, S., Farra, N. and Nakov, P. (2017, August). SemEval-2017 task 4: Sentiment analysis in Twitter. In Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017) (pp. 502-518).
53. Ross, T. (2010), Fuzzy Logic with Engineering Applications. West Sussex: Wiley, 2010.
54. Saleena, Nabizath (2018). "An Ensemble Classification System for Twitter Sentiment Analysis." Procedia Computer Science 132 (2018): 937-946.
55. Sanders Twitter Dataset, Retrieved Jan 31, 2019 from http://www.sananalytics.com/lab.
56. Saif, H., He, Y., & Alani, H. (2012, November). Semantic sentiment analysis of twitter. In International semantic web conference (pp. 508-524). Springer, Berlin, Heidelberg.
57. Saif, H., He, Y., Fernandez, M., & Alani, H. (2016). Contextual semantics for sentiment analysis of Twitter. Information Processing & Management, 52(1), 5- 19.
58. Sanz, Jose Antonio, Alberto Fernandez, Humberto Bustince, and Francisco Herrera (2013). "IVTURS: A linguistic fuzzy rule-based classification system based on a new interval-valued fuzzy reasoning method with tuning and rule selection." IEEE Transactions on Fuzzy Systems 21, no. 3 (2013): 399-411.
59. Severyn, A., & Moschitti, A. (2015, August). Twitter sentiment analysis with deep convolutional neural networks. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 959-962). ACM.
60. Siddiqua, Umme Aymun, Tanveer Ahsan, and Abu Nowshed Chy (2016). "Combining a rule-based classifier with weakly supervised learning for twitter sentiment analysis." In Innovations in Science, Engineering and Technology (ICISET), International Conference on, pp. 1-4. IEEE, 2016.
61. Srivastava, R. and 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.
62. Susan, S. and Keshari, J. (2019). Finding significant keywords for document databases by two-phase Maximum Entropy Partitioning. Pattern Recognition Letters.
63. Tan, Shulong, Yang Li, Huan Sun, Ziyu Guan, Xifeng Yan, Jiajun Bu, Chun Chen, and Xiaofei He (2014). "Interpreting the public sentiment variations on twitter." IEEE transactions on knowledge and data engineering 26, no. 5 (2014): 1158-1170.
64. Vapnik, V., Guyon, I. and Hastie, T. (1995). Support vector machines. Mach. Learn, 20(3), pp.273-297.
65. Vashishtha, S. and 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.
66. Windasari, Ike Pertiwi, Fajar Nurul Uzzi, and Kodrat Iman Satoto (2017). "Sentiment analysis on Twitter 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.

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

67. Yan, Yeqing, Hui Yang, and Hui-Ming Wang (2017). "Two simple and effective ensemble classifiers for Twitter sentiment analysis." In Computing Conference, 2017, pp. 1386-1393. IEEE, 2017.
68. Yoo, SoYeop, JeIn Song, and OkRan Jeong. (2018) "Social media contents based sentiment analysis and prediction system." Expert Systems with Applications 105 (2018): 102-111.
69. Wang, X., Wei, F., Liu, X., Zhou, M., & Zhang, M. (2011, October). Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. In Proceedings of the 20th ACM international conference on Information and knowledge management (pp. 1031-1040). ACM.
70. Zadeh, L.A. (1975), "The concept of a linguistic variable and its application to approximate reasoning," Inf. Sci., vol. 8, pp. 301–357, 1975.
71. Zadeh, L.A (1996), "Fuzzy logic = computing with words," IEEE transactions on fuzzy systems, vol. 4, no. 2, pp.103-111, May, 1996.
72. Zadeh, L.A. (2015). "Fuzzy logic: A personal perspective," *Fuzzy Sets Syst.*, vol.281, pp. 4–20, Dec., 2015.

Appendix

In this section, we present how a single tweet is being processed by our proposed fuzzy rule based unsupervised Sentiment Analysis model. A sample tweet of Nuclear twitter dataset ("Nuclear Twitter Dataset", 2019) is being processed using the three lexicons.

SentiWordNet Lexicon

Initially text preprocessing is done and list of tokens is generated. Using WSD, eq. (2) and eq. (3), positive score ($\mu_{Pos}(a)$) and negative score ($\mu_{Neg}(a)$) for each token is calculated. The positive score of tweet (*TweetPos*) is equal to 0.5 and negative score (*TweetNeg*) is equal to 0, computed using eq. (6) and eq. (7) respectively. The fuzzy sets Low, Medium and High are created using triangular fuzzy membership for universe variables: positive (*x_p)* is (0-5), negative (x*_n)* is (0-5) and output (*x_op)* is (0-10). The fuzzy rules (equations 15-23) are applied. The overall firing strength of tweet for different emotion classes are evaluated using equations (24-26). Fig.17 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 equations (27-29). Fig.18 depicts the aggregated output membership ($\mu_A(x)$) computed in eq. (30). The area under the aggregated output is used for centroid defuzzification in eq. (31). The defuzzified output equal to 5.31 is shown as bold straight line. Finally, the sentiment of tweet is evaluated as 'Neutral' using eq. (32). We can check the polarity of the tweet from dataset, and it turns out to be same.

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

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

Tokens: ['mention', 'us', 'nuclear', 'industry', 'taking', 'steps', 'reconfirm', 'safety', 'amp', 'emergency', 'preparedness', 'nuclear', 'plants', 'learn', 'link']

Positive Score for each token in tweet:
[0.5, 0.0, 0.0, 0.0, 0.0, 0, 0.0, 0.0, 0.0, 0, 0.0, 0.0, 0.0, 0.0]

Negative Score for each token in tweet:
[0.0, 0.0, 0.0, 0.0, 0.0, 0, 0.0, 0.0, 0.0, 0, 0.0, 0.0, 0.0, 0.0]

Selected Positive Score: [0.5]
Selected Negative Score: [ ]

Positive *(TweetPos)* Score for each tweet :0.5
Negative *(Tweet Neg)* Score for each tweet :0

Firing Strength of Negative (w_{neg}): 0.0
Firing Strength of Neutral (w_{neu}): 0.75
Firing Strength of Positive (w_{pos}): 0.25

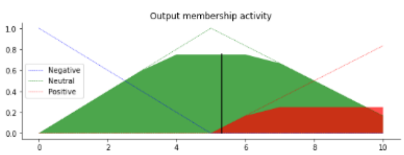
Resultant consequents MFs:
op_activation_low: [0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
op_activation_med: [0. 0.2 0.4 0.6 0.75 0.75 0.75 0.67 0.5 0.34 0.17]
op_activation_high: [0. 0. 0. 0. 0. 0. 0.17 0.25 0.25 0.25 0.25]

Aggregated Output: [0. 0.2 0.4 0.6 0.75 0.75 0.75 0.67 0.5 0.34 0.25]

Defuzzified Output: 5.31

Output after Defuzzification: **Neutral**

Doc sentiment: **neutral**



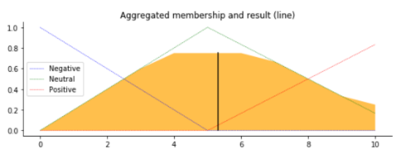


Fig. 17. Output showing different emotions of the tweet **Fig.18.**Aggregated Output of the tweet

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

AFINN Lexicon

Initially text preprocessing is done and list of tokens is generated. The AFINN method fetches the score, $\mu(a)$, of each token using AFINN lexicon (eq. (8)) Using eq. (9) and eq. (10), positive score ($\mu_{Pos}(a)$) and negative score ($\mu_{Neg}(a)$) for each token is calculated. The positive score of tweet (*TweetPos*) is equal to 1.0 and negative score (*TweetNeg*) is equal to 2.0, computed using eq. (11) and eq. (12) respectively. The fuzzy sets Low, Medium and High are created using triangular fuzzy membership for universe variables: positive (*x_p)* is (0-9), negative (x*_n)* is (0-9) and output (*x_op)* is (0-10). The fuzzy rules (equations 15-23) are applied. The overall firing strength of tweet for different emotion classes are evaluated using equations (24-26). Fig. 19 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 equations (27-29). Fig. 20 depicts the aggregated output membership ($\mu_A(x)$) computed in eq. (30). The area under the aggregated output is used for centroid defuzzification in eq. (31). The defuzzified output equal to 4.8 is shown as bold straight line. Finally, the sentiment of tweet is evaluated as 'Neutral' using eq. (32). We can check the polarity of the tweet from dataset, and it turns out to be same.

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

Tokens:['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 token in tweet: [1.0]

Negative Score for each token in tweet: [-2.0]

Positive Score (***TweetPos***) for each tweet: 1.0

Negative Score (***TweetNeg***) for each tweet: 2.0

Defuzzified Output: 4.8

Firing Strength of Negative (w_{neg}): 0.5
Firing Strength of Neutral (w_{neu}): 0.75
Firing Strength of Positive (w_{pos}): 0.25

Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

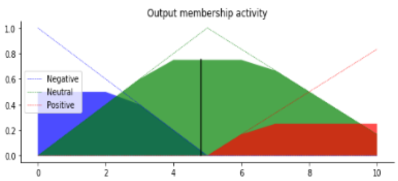
Resultant consequents MFs:
op_activation_low: [0.5 0.5 0.5 0.4 0.2 0. 0. 0. 0. 0. 0.]
op_activation_med: [0. 0.2 0.4 0.6 0.75 0.75 0.75 0.67 0.5 0.34 .167]
op_activation_high: [0. 0. 0. 0. 0. 0. 0.167 0.25 0.25 0.25 0.25]

Aggregated Output: [0.5 0.5 0.5 0.6 0.75 0.75 0.75 0.67 0.5 0.34 0.25]

Defuzzified Output: 4.8

Output after Defuzzification: **Neutral**

Doc sentiment: **neutral**



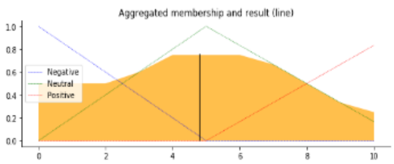


Fig. 19. Output showing different emotions of the tweet **Fig.20.**Aggregated Output of the tweet

VADER Lexicon

The processing of tweet using VADER lexicon has been explained in section 5.1.
Tweet: rt mention the us nuclear industry is taking 7 steps to reconfirm safety & emergency preparedness at nuclear plants. learn more: {link} {'neg': 0.107, 'neu': 0.779, 'pos': 0.115, 'compound': 0.0516}

Positive Score (***TweetPos***) for each tweet :0.1
Negative Score (***TweetNeg***) for each tweet :0.1

Firing Strength of Negative (w_{neg}): 0.2
Firing Strength of Neutral (w_{neu}): 0.8
Firing Strength of Positive (w_{pos}): 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]

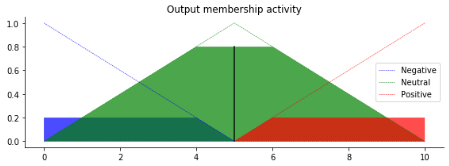
Citation information: https://doi.org/10.1016/j.eswa.2019.112834
Expert Systems with Applications, Volume 138, 30 December 2019, 112834

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**



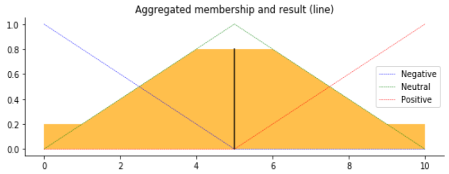


Fig. 21. Output showing different emotions of the tweet **Fig.22.**Aggregated Output of the tweet