Deep Sentiment Extraction using Fuzzy-Rule based Deep Sentiment Analysis

SIREESHA JASTI[1]
Research Scholar, Computer Science and Engineering
GITAM School of Technology
Visakhapatnam, India

G.V.S. RAJ KUMAR[2]
Professor, Computer Science and Engineering
GITAM School of Technology
Visakhapatnam, India

***Abstract*—In the world of social media, the amount of textual data is increasing exponentially on the internet, and a large portion of it expresses subjective opinions. Sentiment Analysis (SA) also named as Opinion mining, which is used to automatically identify and extract the subjective sentiments from text. In recent years, the research on sentiment analysis started taking off because of a huge of amount of data is available on the social media like twitter, machine learning algorithms popularity is increased in IR (Information Retrieval) and NLP (Natural Language Processing). In this work, we proposed three phase systems for sentiment classification in twitter tweets task of SemEval competition. The task is predicting the sentiment like negative, positive or neutral of a twitter tweets by analyzing the whole tweet. The first system used Artificial Bee Colony (ABC) optimization technique is used with Bag-of-words (BoW) technique in association with Naive Bayes (NB) and k-Nearest Neighbor (kNN) classification techniques with combination of various categories of features in identifying the sentiment for a given twitter tweet. The second system used to preserve the context a Rider Feedback Artificial Tree Optimization-enabled Deep Recurrent neural networks (RFATO-enabled Deep RNN) is developed for the efficient classification of sentiments into various grades. Further to improve the accuracy of classification on n-valued scale Adaptive Rider Feedback Artificial Tree (Adaptive RiFArT)-based Deep Neuro fuzzy network is devised for efficient sentiment grade classification. Finally, this research work proposed a Fuzzy-Rule Based Deep Sentiment Extraction (FBDSE) Algorithm with Deep Sentiment Score computation. Accuracy measure is considered to test the proposed systems performance. It was observed that the fuzzy-rule based system achieved good accuracy compared with machine learning and deep learning based approaches.**

Keywords*—*Sentiment analysis; SemEval; recurrent neural networks; LSTM; word embeddings; accuracy; f1-score; fuzzy –rule; deep sentiment extraction

I. Introduction

Nowadays, online social media platforms like Twitter or Face book provide a valuable framework for individuals to share and discuss ideas and opinions regarding various topics including health related issues. Within the past decade these kinds of platforms have spread globally and witnessed a rapid growth in the number of users reaching people from various demographic groups, ethnicity and occupations. Twitter is a social media platform that was used by millions of active users to share their opinion, sentiments and their thoughts on any issue. In this platform, the users are interacted through messages which are called tweets. The estimation of the organizers of the twitter is around 500 million tweets are forwarded in every day. These tweet messages are restricted in size which can contain at most 140 characters. Twitter allow for an exhaustive aggregation and analysis of user generated content which facilitates monitoring of public opinions and sentiment over time towards certain topics of interest. Keeping track of these developments can be crucial for health professionals in order to understand and address the public opinion and behavior with regard to health related topics.

Due to the importance of sentiment analysis to business, the interest has shifted from computer science to management, economics and to the whole society. Nowadays, almost all of the big companies and organizations are having a sort of voice of the customer channel such as emails or call centers and a mean to analyze it. This will help organizations to reshape their services and re-engineer their business processes into the best practice ones. Recently, business has realized the important role that voice of the customer plays in their organizations. The need to have classification systems that are able to handle their feedback data efficiently is emerged. Sentiment analysis is considered as branch of social information where retrieved text is intended to be classified into many classes depending on the detected emotions [1].

Sentiment analysis is the science of extracting, studying and investigating people's sentiments, experience and point of view that documented into a piece of text called, *review*. This extracted information could express general feelings of the authors or extracting sentiments regarding entities such as service, or a product [2]. In computer science, it is about modeling a system that classifies the polarity of a given review. Sentiment analysis model trying to classify reviews into labeled polarities like negative or positive and some adding neutral class [3]. The significant definition for an opinion is a quadruple of four components (G, S, H, T). G is sentiment target, also known as entity or aspect such as camera, product or service under review. S is the sentiment with respect to a target. Opinion's holder is representing by H, whereas T is the time that the opinion was given in. Analyzing the sentiment of given input could be done on three different levels. In the literature, document level sentiment analysis is investigated when calculating the sentiment of the total document, having that the document is about single entity. Another level of a scope is the sentiment level when each sentence is analyzed and given a polarity [3]. In this level, it is assumed that the review consists of several opinionated targets, aspects or entities and their polarities. Finally, aspect level sentiment analysis, where the sentiment is determined on the aspects of an entity. Sentiment analysis has many end-user applications such as monitoring news and social media to look for biased opinions, monitor attitudes towards political candidates or controversial topics during an election, or keeping track of company reputation and consumer response. Sentiment analysis can also be useful for researchers in other fields such as political science and media studies by providing a quantitative method of analyzing public opinion. The internet is growing by the minute and most of the content is in the form of unstructured text. A lot of this data, like blogs, social media and product reviews, are subjective and opinionated. Market analysis is one of the most prominent use cases of SA. Companies used sentiment analysis to keep track of customer relationships, company reputation, consumer response to products and more. Another use of SA is intelligence, where governments can monitor and flag potential threats to national security. There are also many political uses like identifying ideological shifts on social media or identifying which topics engage the public. SA is also used to analyze twitter for public opinion on some topic analysing the plot of fiction, tracking how emotions change throughout the story.

Many of sentiment analysis systems focus on a single language, typically English. However, as the Internet spreads around the world, users are leaving comments in a variety of their native languages. Sentiment analysis in a single language increases the possibility of missing sensitive details in texts written in other languages. Multi-lingual sentiment classification techniques were developed to evaluate data in multiple languages [10]. Sentiment classification frameworks and techniques for various languages are being developed as a result of this. The majority of research in sentiment classification based on the usage of machine learning techniques, unsupervised and supervised. The supervised methods used machine learning algorithms to train with labeled sentiment data and to determine the sentiment of unlabeled test documents. A machine learning method is used in the training phase to learn a prediction model, which is then utilized in the testing or prediction phase to classify documents that the model has never seen before. Feature engineering is likely the most important of these components for classification. People express their negative feelings in sarcastic text by using positive words. Because of this, sarcasm can easily fool sentiment analysis models unless they are specifically designed to account for its possibility. English is considered Majority of the digital platforms and articles are using.

Traditional approaches take the input text, process it and find the sentiment classifications like positive and/or negative based on users one-dimensional reviews. But, these are failed to look into multidimensional perspective of the users review like sarcastic reviews. So, there is very much need for the approaches which are capable to understand the user’s multidimensional perspective in giving the reviews.

In this work, machine learning based approaches are used for sentiment analysis at the initial phases. Deep leaning based architectures are developed in the next phases to tokenize the input as word embeddings for the sentiment analysis task. Finally, this research demonstrates the fuzzy rule based approaches to compute the degree of sentiment using polarity and deep sentiment score. This study is experimented on SemEval competition sentiment classification dataset contains small micro-blog twitter messages.

This work is structured in six sections. The Section 2 explains the existing work of the sentiment analysis. The dataset details and evaluation measures are presented in Section 3. The machine learning based approach described in Section 4 with experiment results. The Section 5 explains the deep neural architectures and experimental results for sentiment classification. Section 6 concludes this work.

II. Related Work

The automated process of splitting and categorizing units of texts into separate, specified categories, also known as classes, is referred to as text categorization. Text classification can be used to extract a text's topic, but it can also be used to classify sentiment. Sentiment classification, often known as sentiment analysis, is the process of determining whether a text is negative, positive, or neutral. NLP methods are used to systematically analyze and evaluate the sentiment expressed in text and assign it to a sentiment class.

Topic classification, i.e. determining whether a text is about politics or sports, is usually done using machine-learning approaches. Pang et al. (2002) adopted [1] these machine-learning methods and regarded positive or negative sentiment as topics of their own. With this approach they managed to achieve performances hovering around 80% when analyzing movie reviews. However, the experiments conducted only contained reviews that were considered positive or negative. Agarwal et al. conducted [4] the accuracy of sentiment analysis on Tweets using a Support Vector Machine (SVM) was around 60% depending on the features employed. They also pre-processed the Tweets by replacing acronyms with full meanings and emoticons with their emotional states. When doing binary classification, the proposed method achieved an accuracy of roughly 75% when omitting the neutral class and just having positive or negative classes.

Bhayani et al. conducted [5] experiments that utilized distant supervision on data extracted from Twitter. Distant supervision relates to gathering and labelling training data automatically. They worked on the notion that any Tweet with a positive emoticon likewise has positive emotion, and any Tweet with a negative emoticon has negative sentiment. They could construct an annotated classifier training set without hand-labeling data using these assumptions. They regarded emoticons to be noise; therefore they removed them from the data set. Given the importance of emoticons in expressing moods and their widespread use in sentiment analysis, a feasible upgrade to this approach would be to appropriately integrate emoticons and their emotional connotations.

Edilson Anselmo et al., explained [6] a multi-view ensemble approach for the task of SA also specifically named as Message Polarity Classification in SemEval-2017 challenge. In this ensemble approach, different types of features are used to train every base classifier. The first base classifier is Linear SVM which uses bag-of-words model as a feature space. The second base classifier is another Linear SVM which uses the averaging of word embeddings of tweets. The third base classifier is Logistic Regressor where in the tweets are represented as averaging of weighted word embeddings. In the first classifier, the tweets are represented with the TF-IDF (Term Frequency and Inverse Document Frequency) weights of bag-of-words. In the third classifier the word embeddings weight is represented with TF-IDF measure. In case of F1-score and recall, their proposed system got an 18th rank and 20th respectively among 38 participants in the competition.

Tzu-Hsuan Yang et al., implemented [7] a system for SA task in twitter dataset. This system is a combination of two deep neural networks based models such as LSTM Recurrent Neural Network and convolution neural network (CNN) through interpolation. These models take the input as distributional words representations of tweets as vectors and the output is the sentiment of a tweet. They observed that the word embeddings performance is good in RNN for computing Recall and accuracy when compared with one-hot vectors. They identified that LSTM model performance is good for predicting all classes but the RNN fails to predict the negative class. They also observed that the performance of LSTM is good on neutral and positive classes and the performance of CNN is good on negative class compared with LSTM. The proposed system obtained an average F1 score of 0.587, average recall rate of 0.618 and the accuracy of 0.618 for sentiment analysis task of SemEval 2017 Competition.

Recently, the implementation of SA in twitter using neural networks become one of the state of the art techniques because they took less number of features when compared with traditional techniques. Yichun Yin et al., proposed [8] an effective and simple ensemble method to boost the performance of neural network models. They collected the several sets of word embeddings which are constructed by using skip-gram model or released publicly or learned from different corpus. They assume that the usage of these embeddings increases the performances and generalizations of neural network models. They identified different types of neural networks like RNN, CNN, LSTM and GRU for implementing their method, but they used RCNN [9] in their method. The aim of RCNN is to capture non-consecutive and longer range patterns in a weighted manner by using adaptive gated decay and non-consecutive convolution. The proposed method achieved 1st rank in Accuracy and 5th rank in average recall in the SemEval 2017 sentiment analysis competition.

Traditional techniques used a different type of hand crafted features such as semantic, surface-form and sentiment lexicons for sentiment classification of twitter [10]. The performance of these techniques mainly depends on the quality work of feature construction and developing a popular system for these features. Moreover, the one-hot representation of these features proposes the sparsity problem in the representation and the semantic information also not captured in the representation. To overcome this problem, Tang et al. induced [11] real-valued, sentiment-specific and low-dimensional embedding features for sentiment classification of twitter which encode the combination of sentiments and semantics of words. They identified that the hand-crafted features and embedding features achieved similar results.

AbeedSarker et al., presented [12] a system by combining the dense and sparse vector representations of words and clusters are used for generalized words representations for supervised text classification. The neural networks are trained by using the dense vectors which are represented with word embeddings in a large unlabeled dataset to predict the neighboring words. The sequences of word n-grams (n range is 1-3) are used for generating the sparse vectors. The Support Vector Machines (SVMs) with an RBF kernel is used for classification of a text segment by concatenating the different vector representations. This system is specifically proposed for non-experts of machine learning and natural language processing and doesn't require any manual tuning of weights or parameters. The system generates the classification model by automatically optimizing the relevant hyper-parameters and producing the training vectors for SVM classifier from a given training dataset. This system is evaluated on the sentiment analysis task of SemEval 2017 English dataset. The proposed system achieved F1-score of 0.632, accuracy of 0.646 and average recall of 0.637.

Raphael Troncy et al., proposed [13] a SentiME++ system which is an ensemble approach for the SA task of SemEval 2017 competition. The aim of this system is to classify the English tweets based on the type of sentiment like negative, positive or neutral sentiments they have. SentiME++ merge the predictions of five popular sentiment classifiers such as NRC-Canada [10], GU-MLT-LT [14], KLUE [14], TeamX [15] and Stanford Sentiment System (subsystem of the Stanford NLP Core toolkit). In SentiME++ approach, a bootstrap sampling (a uniform random sampling with replacement) technique is used to produce four different training sets from the initial training set T. These training sets are used to train the four sub classifiers separately. SentiME++ trains four classifiers separately and fifth classifier Stanford Sentiment System is not trained with training data. The outputs of five classifiers are represented as a feature vector and this is directed to stack supervised learner. This system produced a F1-score of 0.613 for sentiment classification and obtained 12th rank in the competition.

Chukwuyem Onyibe explained [16] a supervised system which uses lexical features (Uni-grams, Tweet length, Tweet length binned, Bi-grams, SentiStrength, Removed URL, Stop words) and optimized Conditional Random Fields to predict the tweet sentiment. We used CRF++which is implemented with the primary machine learning component as Conditional Random Fields (CRF). They were inspired by the work of Yang et al. (2007) who used CRFs to identify the sentiment in web blogs by giving training at sentence level and classifying at the document level by considering the sequences of sentences. In their work, they optimized the parameters of CRF++ as well as lexical features for the task of sentiment classification in twitter of SemEval 2017. The combination of SentiStrength and unigrams performed well to obtain good results. They obtained good results for sentiment classification when the parameters of CRF are f value 1, c value 8.5 and features are uni-grams. The proposed system achieved an average F1-score of 0.54226, average recall of 0.59024 and accuracy of 0.61519.

Mohammed Jabreel et al., proposed [17] a system named as SiTAKA for SemEval 2017 twitter SA task for both Arabic and English languages tweets. In this system, the tweets are represented with a set of novel features which includes the information generated by five lexicons (NRC hashtags lexicon [10], General Inquirer [T8], TSLex [18], Hu-Liu opinion lexicon (HL) [19] and SenticNet [20]) and a bag of negated words. It was observed that the combination of these features with some basic features (syntactic, basic text, lexicon, Word Embeddings and cluster) increases the performance of the classification. The SVM classifier is used to identify the sentiment of tweets. The proposed system achieved 2nd rank in Arabic and 8th rank English language tweets.

SymeonSymeonid is proposed [21] an approach based on a Majority Vote scheme and combined the classical linguistic resources such as sentiment lexicon features and bag-of-words with supervised machine learning methods for sentiment classification task. The usage of lexicons and bag-of-words representation has a predefined sentiment for each uni-gram and bi-gram. We used different types of classifiers such as Ridge, Logistic Regression, Stochastic Gradient Descent, Nearest Centroid, Bernoulli Naive Bayes, Linear SVC, Passive-Aggressive for testing the performance of the proposed approach. It was observed that among the set of all three combination classifiers one set of classifiers such as the Nearest Centroid, the Stochastic Gradient Descent (SGD) and Bernoulli Naive Bayes achieved best results. They also identified that Nearest Centroid is a weak classifier when alone but it gives best contribution when combined with other two classifiers.

Sentiment classification is one of the major issues of NLP and become more popular among many research fields. Typically, sentiment classification predicts the sentiment of a text into various discrete classes like negative, positive or neutral. Ming Wang et al., explained [22] a deep learning system to classify 2-polarity, 3-polarity and 5-polarity in tweets by combining SVM with GRU. In their system, first they used pre-trained word embeddings to train a gated recurrent neural network, and then they extracted features from GRU layer and forwarded these features to SVM to perform quantification and classification sub-tasks. Joosung Yoon et al., proposed [23] a sentiment analyzer for sentiment classification to predict the sentiment at document level of English tweets for SemEval 2017 competition. This method is based on lexicon integrated CNNs with attention (LCA). The proposed method achieved an average recall of 58.9%, an average F1-score of 55.2% and an accuracy of 61.4% for the task of sentiment classification.

MickaelRouvier et al., explained [24] a system which is an ensemble model of Deep Neural Networks (DNN) such as RNN-LSTM and CNN. We used four different types of pre-trained embeddings such as three different sentiment embeddings and one lexical embedding on large datasets to initialize the input representation of DNN. These models can capture word level information only. We injected some sentence level features like Emoticons, Elongated units, Lexicons, All-caps and Punctuation into the system. They used a score-level fusion approach to combine the ensemble of DNNs. The proposed system got 2nd rank at SemEval 2017 competition and achieved an average recall of 67.6%. It was identified that the CNN model obtained best results for sentiment classification when compared with RNN-LSTM models. Haowei Zhang et al., proposed [25] a multichannel model named as CNN-LSTM to predict the sentiment of English language twitter tweets and this model is a combination of two parts such as multi-channel LSTM and CNN. Unlike a CNN, a multi-channel strategy is applied in the proposed model to extract different scales of active local n-gram features and various filters of different length are used in this strategy. Then LSTM is used to compose the information sequentially. In the classification process, they considered both long distance dependencies across tweets and local information within tweets by combining both CNN and LSTM. This multi-channel approach achieved an accuracy of 0.640 for sentiment classification.

AmitAjitDeshmane et al., proposed [26] a system for SemEval 2017 twitter sentiment analysis task. The system is an ensemble of three different deep learning architectures. The first architecture used CNNto perform the text classification. The second architecture implemented with gated RNN. In third subsystem, the opinion lexicons are integrated directly with CNN architecture. The proposed ensemble system obtained a macro-averaged recall of 64.3%. It was observed that the first architecture is crucial to improve the results of sentiment classification. Iv´an Castro et al., proposed [27] a system for SemEval-2017 twitter sentiment analysis task. In this system, they studied about how the relationships among sense n-grams and sentiment polarities like negative, neutral or positive are contributed to this task. They also tested the effect of removing a large set of char n-grams features reported in previous works. Based on these observations, they explore a SentiWordNet as a polarity lexicon and constructed a SVM system. The proposed system got 10th rank in the competition and achieved an F1-score of 0.624.

Yunxiao Zhouet al., reported [28] a system for SA in twitter task of SemEval-2017 competition. They investigated various traditional Natural Language Processing (NLP) features (Word RF n-grams, POS tag, Negation), domain specific features (All-caps, Bag-of-Hashtags, Elongated, Emoticon, Punctuation) and word embedding features (GoogleW2V, GloVe, sentiment word vector (SWV), sentiment-specific word embedding (SSWE)) alongwith supervised machine learning techniques ((SVM), AdaBoost, Logistic Regression (LR) and SGD) to address this task.

III. Proposed Fuzzy-Rule based System and Methodology

In this section an automated framework for deep sentiment extraction is proposed with the components described in (Fig. 1).

Initially, the framework will accept the reviews from any dataset and then text processing steps like stemming and stop word removal are performed. The process of feature selection is carried out as second step for the process using novel machine learning methods, which is elaborated in this work. Pre-processed data is given as input in the feature extraction stage, where extraction of significant features, namely spam words-based features, SentiWordNet features, emoticon-based features and TF-IDF features takes place for decreasing the processing of data. Further, the key phrases from the review text are extracted using feature fusion methods. This section elaborates the procedure of identification of parameter μ based on decision tree. At first, review data is trained for finding the parameter after that, rule is induced and categorized. Feature fused output is taken as input for classification of sentiment grade. Here, a classifier named Deep learning network is used to perform the classification of the sentiment grade. Moreover, deep learning network is trained based on devised Adaptive RFATO technique. This proposed Adaptive RFATO algorithm is devised by combining ROA [29] and FAT scheme [30] with adaptive concept.

On the other hand, ROA involves four riders, namely follower, bypass rider, attacker and overtake racing besides others for reaching destination. Finally, fuzzy rule based approaches are used to analyze the degree of sentiment using polarity and to extract the opinion. The combination of ROA and FAT scheme, named as RFATO technique offers best solution to solve optimization issues. However, this method consumes more computational time. Hence, this research included the adaptive mechanism with RFATO method for obtaining less computational time.

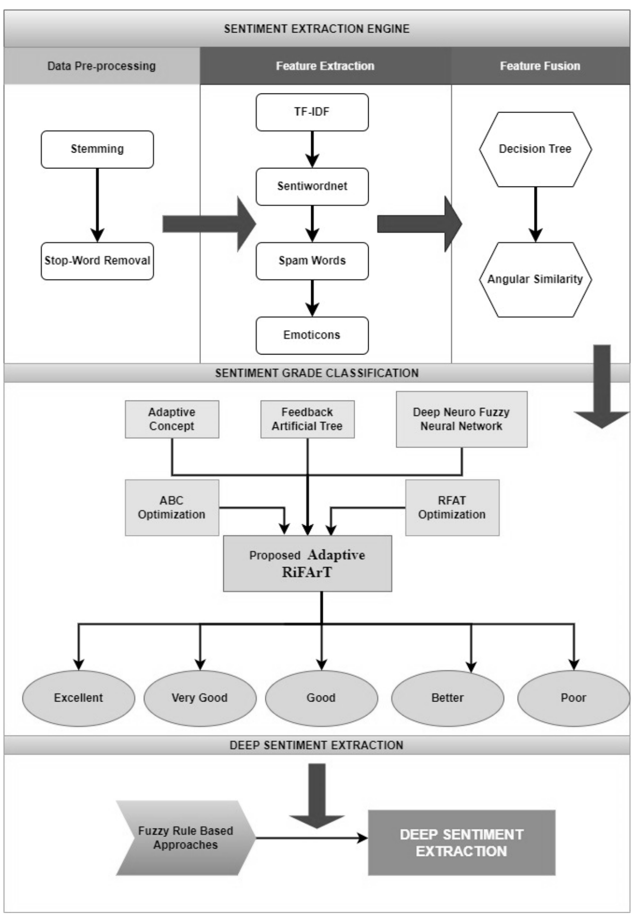

Fig. 1. Proposed Frameworks for Deep Sentiment Extraction.

A. Gated Recurrent Units (GRU) for Sentiment Analysis

By using the gating mechanism the gradient stays high for long time-steps and thus overcome the problem of vanishing gradient. Despite its effectiveness, the complexity found in the gating mechanism make the architecture expensive to execute and hard to analyze. As a result, many variations have been proposed but Gated Recurrent Unit (GRU) introduced by [32] is one of the strongest competitors to LSTM. Compared to LSTM, it is much simpler and faster and needs fewer gates to calculate. In sentiment classification task, RNN based layer could be used to generate sentence hidden representation and then, this representation could be used as input to the classification function to find a polarity of a given tweet [33, 34].

For the proposed models, GRU layer is used rather than LSTM as its GRU works better in this experiment as the validation loss on the dataset was lesser and it was faster also. Compared with standard RNN, GRU cell uses two gates to handle the flow of the data, z_t to represent update gate and r_t to represent reset gate. z_t determines the amount of previous information to keep while r_t determines the methodology of combining the old memory with the new input. The entire internal memory is output without an additional activation. At each step, each word-distributed representation is given as an input word x_i, the current cell state c_t and hidden state h_t can be updated with the previous cell state c_{t-1} and previous hidden state h_{t-1}. The equations (1) (2) (3) and (4) are used to compute the z_t, r_t, c_t and h_t respectively.

$$z_t = \sigma\left(x_t U_z + h_{t-1} W_z\right) \tag{1}$$

$$r_t = \sigma\left(x_t U_r + h_{t-1} W_r\right) \tag{2}$$

$$c_t = tanh\left(x_t U_z + Wc(h_t - 1 \odot r_t)\right) \tag{3}$$

$$h_t = (1 - z_t \odot c_t + z_t \odot h_t - 1) \tag{4}$$

The output of this layer is the hidden representation of the given tweet that is going to be fed into the attention layer to highlight the most informative words and to generate the final tweet representation.

B. Attention Layer for Sentiment Orientation

The attention mechanism is used to amplify the most informative words in a given tweet with higher weights inspired by [31]. The attention layer is placed after GRU layer to point out the most influence words to the tweet meaning and then the specific word representations are combined to form a vector representation for tweets. To generate the final tweet representation, the attention layer uses the following mathematical formulas.

Using each hidden state h_t and dot product with its weights is producing the hidden representation u_t. The equation (5) is used to compute u_t.

$$u_t = \tanh\left(h_t w_t + b_t\right) \tag{5}$$

The un-normalized hidden representation of the current word u_t is passed to softmax function to get the normalized importance weightα_t. The 'α_t'is computed using equation (6).

$$\alpha_t = \frac{\exp(u_t)}{\sum_{i=1}^{T} \exp(u_i)} \tag{6}$$

Where, T is count of time-steps in the input, w_t and b_t are the weights of attention layer which are optimized during training to allocate more weights to the most important words of a sentence. Finally, the high level vector representation of a tweet S is produced based on the weights by using weighted sum of the word annotations. The S is computed using equation (7).

$$S = \sum_{i=1}^{T} h_{it}\alpha_{it} \tag{7}$$

To classify the tweets into the sentiment labels, the final vector representation of tweet is fed into fully connected Sigmoid logistic regression layer. It produces an output in the range between 0 and 1 by using the probability distributions of all sentiment classes.

This result into a semantic orientation of the word or the phrase using the Point_wise Mutual Information:

$$SO(S) = PMI(S, Excellent) - PMI(S, poor) \tag{8}$$

Here,

SO(S), denotes the semantic orientation of the phrase.

PMI (S, Excellent), denotes the PMI for the phrase with positive sentiment.

PMI (S, poor), denotes the PMI for the phrase with negative sentiment.

IV. Proposed Model for Sentiment Extration

This section elaborates on the proposed mathematical model for the deep sentiment extraction process. Firstly, the generic set of positive and negative words accumulated in two different classes. The words are collected based on the popular vocabulary of the most commonly used phrases for consumer product review systems.

$$[P] = \langle Good, Excellent, Amazing, Incredible, Great, ... \rangle \tag{9}$$

and

$$[N] = \langle Bad, Poor, Terrible, ... \rangle \tag{10}$$

Here, the Positive and Negative reference words are represented by the sets P and N.

$$SO(S) = PMI(S,[P]) - PMI(S,[N]) \tag{11}$$

Further, only the phrases or the words are considered with the highest correlation as $SO(S) > 1$

Finally, perform the iterative classification of the phrases in the review or feedback for identifying the sentiment score and calculate the total score based on the weighted calculation of rating and text feedback, the mathematical model is converted into the workable algorithm as discussed in Algorithm-1.

Algorithm-1: Fuzzy-Rule Based Deep Sentiment Extraction (FBDSE)

Step -1. Create collection of positive and negative word sets

Step -2. Accept the list of products and analyse

- A. For each product
- B. Break the sentences into phrases
  - i. For each zone
    - a. Calculate the correlation score
    - b. if the score is higher than 1
      - Then consider the phrase
    - c. else
      - Reject the phrase from the set
    - d. Repeat for minimal set
- C. Consider the Compound score- Deep Sentiment can be measured by Compound Score and it lies between -1 to +1
- D. Calculate the Deep Sentiment Score = rating *0.8 + text score * 0.2
- E. Deep Sentiment score >= 0.5= Positive (P)
  neutral if 0 <Deep Sentiment score < 0.5= Neutral(Nu)
  negative if -1 >= Deep Sentiment score<=-0 = Negative(N)

V. Dataset Characteristics and Evaluation Measures

In this work, the experiment conducted on the dataset provided by the organizers of SemEval competition task of twitter sentiment analysis at tweet level. This dataset is a combination of the twitter datasets and the characteristics of a dataset are specified in the Table I.

In the training dataset, the number of tweets in neutral class is more compared to other two classes. The dataset is not balanced which means the three classes contain varying number of tweets. The researchers of sentiment analysis used various measures such as F1-Score, recall, precision and accuracy for evaluating the machine learning techniques. In this work, accuracy is used to test the efficiency of the machine learning techniques. Accuracy is the ratio of the number of test tweets are correctly classified their sentiment and the number of test tweets considered in the experiment. Precision and recall metrics are highly recommended especially if the dataset is unbalance. Precision measures the probability that a positive prediction is really positive. Recall measures the efficiency of a model to find all positive reviews in a dataset. Here, the detailed classifications of sentiment types based on given user ratings are shown in Table II.

TABLE I. THE DATASET CHARACTERISTICS

Dataset	Total	Positive	Negative	Neutral
SemEval train	47831	18377	7442	22012
SemEval dev	5653	2412	1056	2185
SemEval test	12284	2375	3972	5937

TABLE II. USER RATINGS AND SENTIMENT TYPE MAPPING

User Rating (Given)	Sentiment Type
-2	Poor
-1	Better
0	Good
1	Very Good
2	Excellent

The following Table III consists of sample of tweets classified into different feature types rated in between -2 and 2.

TABLE III. EXTRACTION OF SENTIMENT FEATURES AND TYPES

Company Name	Total tweets	Poor (-2)	Better (-1)	Good (0)	Very Good (1)	Excellent (2)
Microsoft	100	12	34	20	31	3
Amazon	100	6	12	12	55	15
Bentley	100	1	2	32	62	3
David Cameron	100	7	31	37	24	1
Donald trump	100	2	42	40	15	1
Google	100	0	8	59	33	0
Harry Potter	100	2	0	18	64	16
i-phone	100	0	5	30	56	9
Jurassic World	100	0	5	11	62	22
Madonna	100	0	2	21	73	4

User sentiment score have been calculated as three categories Negative(N), Positive(P) and Neutral (Nu) following table (Table IV) consists of sample of tweets classified into different feature types rated in between -2 and 2.

Next, Deep sentiment score discussed in the proposed approach for the above mentioned three sentiment types is calculated and shown in Table V.

Next, the comparison between the extracted user sentiment score and deep sentiment score have been composed and the observations are shown in Table VI.

TABLE IV. USER SENTIMENT SCORE

Company Name	Negative(N)	Neutral(Nu)	Positive(P)
Microsoft	0.46	0.2	0.34
Amazon	0.18	0.12	0.7
Bentley	0.03	0.32	0.65
David Cameron	0.38	0.37	0.25
Donald Trump	0.44	0.4	0.16
Google	0.08	0.59	0.33
Harry Potter	0.02	0.18	0.8
i-phone	0.05	0.3	0.65
Jurassic World	0.05	0.11	0.84
Madonna	0.02	0.21	0.77

TABLE V. DEEP SENTIMENT SCORE

Company Name	Negative	Neutral	Positive
Microsoft	0.092	0.04	0.068
Amazon	0.036	0.024	0.14
Bentley	0.006	0.064	0.13
DavidCameron	0.076	0.074	0.05
Donald Trump	0.088	0.08	0.032
Google	0.016	0.118	0.066
Harry Potter	0.004	0.036	0.16
i-phone	0.01	0.06	0.13
Jurassic World	0.01	0.022	0.168
Madonna	0.004	0.042	0.154

TABLE VI. COMPARISON OF EXTRACTED USER AND DEEP SENTIMENT SCORE

Company Name	Negative(N)		Neutral(Nu)		Positive(P)	
	USER	DEEP	USER	DEEP	USER	DEEP
Microsoft	0.46	0.092	0.2	0.04	0.34	0.068
Amazon	0.18	0.036	0.12	0.024	0.7	0.14
Bentley	0.03	0.006	0.32	0.064	0.65	0.13
David Cameron	0.38	0.076	0.37	0.074	0.25	0.05
Donald Trump	0.44	0.088	0.4	0.08	0.16	0.032
Google	0.08	0.016	0.59	0.118	0.33	0.066
Harry Potter	0.02	0.004	0.18	0.036	0.8	0.16
i-phone	0.05	0.01	0.3	0.06	0.65	0.13
Jurassic World	0.05	0.01	0.11	0.022	0.84	0.168
Madonna	0.02	0.004	0.21	0.042	0.77	0.154

The graphical illustration of the above calculations is made visible using the following graph as shown in Fig. 2.

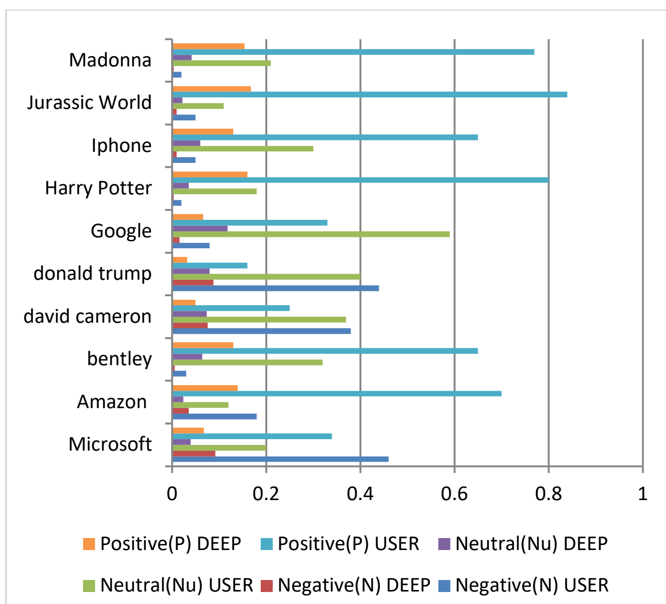

Fig. 2. Comparison between User and Deep Sentiment Score.

TABLE VII. COMPOUND SCORE

Company Name	Compound Score	Final Opinion (Towards)
Microsoft	0.552	Positive
Amazon	0.84	Positive
Bentley	0.78	Positive
David Cameron	0.456	Neutral
Donald Trump	0.528	Neutral
Google	0.708	Positive
Harry Potter	0.96	Positive
i-phone	0.78	Positive
Jurassic World	1	Positive
Madonna	0.924	Positive

Finally, as discussed in the proposed Fuzzy-Rule Based Deep Sentiment Extraction (FBDSE) Algorithm; the compound score of the sentiments have been calculated as shown in Table VII to understand the sentiment orientation which helps in identification of the polarity of the opinion.

Next, as we discussed in the proposed algorithm extracts the opinion based on the Compound Score (CS). If the CS >= 0.5, neutral if 0 < CS < 0.5 and negative if -1 >= CS<=-0.The graph represents the final compound score is shown in Fig. 3.

The Precision, Recall and F1-Score of the proposed Fuzzy-Rule Based Deep Sentiment Extraction (FBDSE) are compared with the earlier proposed algorithms as shown in Table VIII.

Finally, observed that there is a remarkable enhancement in the performance as shown in Fig. 4.

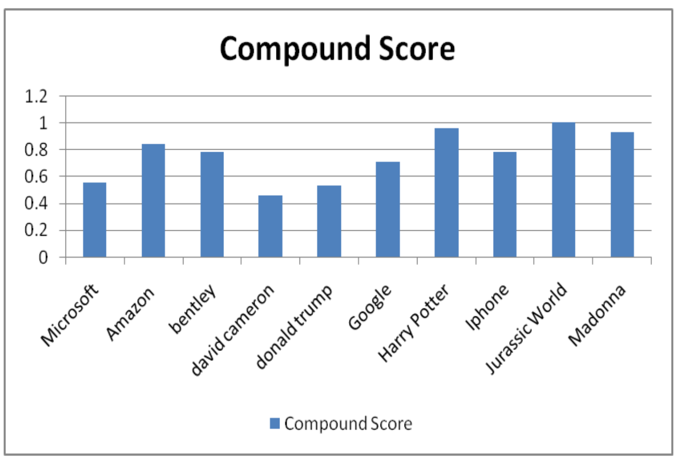

Fig. 3. Compound Score.

TABLE VIII. COMPOUND SCORE

Model Performance	ABC Bi-Gram Model	Proposed RFATO Enabled Deep-RNN	Adaptive RFATO Based Deep Neuro Fuzzy Network	Proposed FBDSE Approach
Recall	0.823	0.887	0.889	0.937
Precision	0.797	0.8	0.801	0.897
F1-Score	0.81	0.841	0.843	0.917

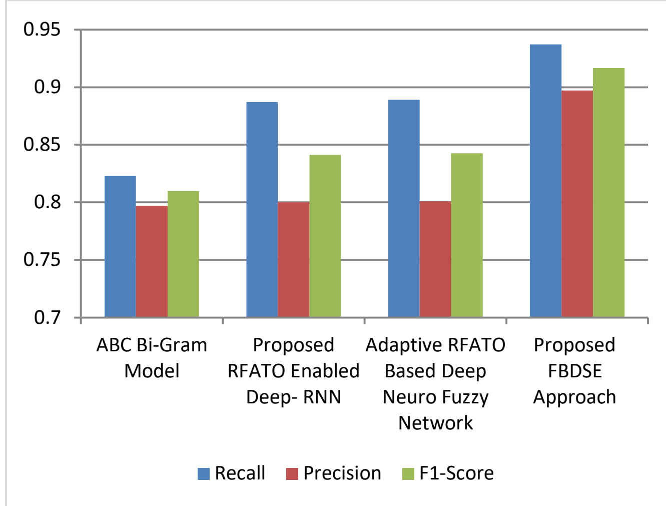

Fig. 4. Comparison of Performance.

VI. CONCLUSION

This research work proposed two approaches for sentiment classification task of SemEval competition. The first approach used machine learning techniques like NB, RF and SVM by using a set of stylistic, syntactic and semantic features. Among these techniques, the SVM obtained good accuracy for sentiment classification. The second approach used deep neural architectures like RNN, LSTM and GRU for sentiment classification. Two word embedding models are used for generating word vectors. The deep learning based approach performance is good compared with machine learning based approach.

A new methodology is proposed to deep sentiment extraction and grading on N-Scale by developing a fuzzy rule based system for sentiment analysis, which can offer more refined outputs through the use of fuzzy membership degrees. The experimental results indicate that our fuzzy-based approach performs marginally better than the other algorithms. In addition, the fuzzy approach allows the definition of different degrees of sentiment without the need to use a larger number of classes and almost 91% of accurate deep sentiments extracted on the given products. Future scope of this work is focused on Multi-Lingual Deep Sentiment Extraction using semantic annotations and AI-based reasons.

REFERENCES

[1] Pang, Bo, Lillian Lee, and ShivakumarVaithyanathan. "Thumbs up?: sentiment classification using machine learning techniques." Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics, 2002.

[2] Abbasi, Ahmed, Hsinchun Chen, and Arab Salem. "Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums." ACM Transactions on Information Systems (TOIS) 26.3 (2008): 12.

[3] Pang, Bo, and Lillian Lee. "Opinion mining and sentiment analysis." Foundations and Trends® in Information Retrieval2.1–2 (2008): 1-135.

[4] Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, “Sentiment Analysis of Twitter Data", In Proceedings of the ACL 2011Workshop on Languages in Social Media,2011 , pp. 30-38.

[5] Go, R. Bhayani, L.Huang. "Twitter Sentiment ClassificationUsing Distant Supervision". Stanford University, Technical Paper,2009.

[6] EdilsonAnselmoCorrˆeaJ´unior, Vanessa Marinho, and Leandro Santos. 2017. NILC-USP at SemEval-2017 Task 4: A Multi-view Ensemble for Twitter Sentiment Analysis. In Proceedings of the 11th International Workshop on Semantic Evaluation. Vancouver, Canada, SemEval '17, pages 610–614.

[7] Tzu-Hsuan Yang, Tzu-Hsuan Tseng, and Chia-Ping Chen. 2017. deepSA at SemEval-2017 Task 4: Interpolated deep neural networks for sentiment analysis in Twitter. In Proceedings of the 11th International Workshop on Semantic Evaluation. Vancouver, Canada, SemEval '17, pages 615–619.

[8] Yichun Yin, Yangqiu Song, and Ming Zhang. 2017. NNEMBs at SemEval-2017 Task 4: Neural Twitter sentiment classification: a simple ensemble method with different embeddings. In Proceedings of the 11th International Workshop on Semantic Evaluation. Vancouver, Canada, SemEval '17, pages 620–624.

[9] Tao Lei, Hrishikesh Joshi, Regina Barzilay, Tommi S. Jaakkola, KaterynaTymoshenko, Alessandro Moschitti, and Llu´ısM`arquez. 2016. Semi-supervised question retrieval with gated convolutions. In NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016. pages 1279–1289.

[10] Saif Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu. 2013. Nrc-canada: Building the state-of-theart in sentiment analysis of tweets. In Proceedings of the 7th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2013, Atlanta, Georgia, USA, June 14-15, 2013. pages 321–327.

[11] Duyu Tang, Furu Wei, Nan Yang, Ming Zhou, Ting Liu, and Bing Qin. 2014. Learning sentimentspecific word embedding for twitter sentiment classification. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Baltimore, Maryland, pages 1555–1565.12.

[12] AbeedSarker and Graciela Gonzalez. 2017. HLP@UPenn at SemEval-2017 Task 4A: A simple, self-optimizing text classification system combining dense and sparse vectors. In Proceedings of the 11th International Workshop on Semantic Evaluation. Vancouver, Canada, SemEval '17, pages 639–642.

[13] Raphael Troncy, Enrico Palumbo, EfstratiosSygkounas, and Giuseppe Rizzo. 2017. SentiME++ at SemEval-2017 Task 4: Stacking state-of-theart classifiers to enhance sentiment classification. In Proceedings of the 11th International Workshop on Semantic Evaluation. Vancouver, Canada, SemEval '17, pages 647–651.

[14] Thomas Proisl, Paul Greiner, Stefan Evert, and BesimKabashi. 2013. KLUE: Simple and robust methods for polarity classification. In 7th International Workshop on Semantic Evaluation (SemEval).

[15] Yasuhide Miura, Shigeyuki Sakaki, Keigo Hattori, and Tomoko Ohkuma. 2014. TeamX: A Sentiment Analyzer with Enhanced Lexicon Mapping and Weighting Scheme for Unbalanced Data. In 8th International Workshop on Semantic Evaluation (SemEval). T8 Philip Stone, Dexter C Dunphy, Marshall S Smith, and DM Ogilvie. 1968. The general inquirer: A computer approach to content analysis. Journal of Regional Science 8(1):113–116.

[16] ChukwuyemOnyibe and NizarHabash. 2017. OMAM at SemEval-2017 Task 4: English sentiment analysis with conditional random fields. In Proceedings of the 11th InternationalWorkshop on Semantic Evaluation. Vancouver, Canada, SemEval '17, pages 669–673.

[17] Mohammed Jabreel and Antonio Moreno. 2017. SiTAKA at SemEval-2017 Task 4: Sentiment analysis in Twitter based on a rich set of features. In Proceedings of the 11th International Workshop on Semantic Evaluation. Vancouver, Canada, SemEval '17, pages 693–698.

[18] Duyu Tang, Furu Wei, Bing Qin, Ting Liu, and Ming Zhou. 2014a. Coooolll: A Deep Learning System for Twitter Sentiment Classification. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Association for Computational Linguistics and Dublin City University, Dublin, Ireland, pages 208–212.

[19] Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pages 168–177.

[20] Erik Cambria, Daniel Olsher, and DheerajRajagopal. 2014. SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. In Twenty-eighth AAAI conference on artificial intelligence.

[21] SymeonSymeonidis, DimitriosEffrosynidis, John Kordonis, and AviArampatzis. 2017. DUTH at SemEval-2017 Task 4: A voting classification approach for Twitter sentiment analysis. In Proceedings of the 11th International Workshop on Semantic Evaluation. Vancouver, Canada, SemEval '17, pages 703–707.

[22] Ming Wang, Biao Chu, Qingxun Liu, and Xiaobing Zhou. 2017. YNUDLG at SemEval-2017 Task 4: A GRU-SVM model for sentiment classification and quantification in Twitter. In Proceedings of the 11th International Workshop on Semantic Evaluation. Vancouver, Canada, SemEval '17, pages 712– 716.

[23] Joosung Yoon, Hyeoncheol Kim, and KigonLyu. 2017. Adullam at SemEval-2017 Task 4: Sentiment analyzer using lexicon integrated convolutional neural networks with attention. In Proceedings of the 11th International Workshop on Semantic Evaluation. Vancouver, Canada, SemEval '17, pages 731– 735.

[24] MickaelRouvier. 2017. LIA at SemEval-2017 Task 4: An ensemble of neural networks for sentiment classification. In Proceedings of the 11th International Workshop on Semantic Evaluation. Vancouver, Canada, SemEval '17, pages 759–764.

[25] Haowei Zhang, Jin Wang, Jixian Zhang, and Xuejie Zhang. 2017. YNU-HPCC at SemEval 2017 Task 4:Using a multi-channel CNN-LSTM model for sentiment classification. In Proceedings of the 11th InternationalWorkshop on Semantic Evaluation. Vancouver, Canada, SemEval '17, pages 795–800.

[26] AmitAjitDeshmane and Jasper Friedrics. 2017. TSA-INF at SemEval-2017 Task 4: An ensemble of deep learning architectures including lexicon features for Twitter sentiment analysis. In Proceedings of the 11th International Workshop on Semantic Evaluation. Vancouver, Canada, SemEval '17, pages 801–805.

[27] Iv´an Castro, Sebasti´anOliva, Jos´e Abreu, Claudia Mart´ınez, and YoanGuti´errez. 2017. UCSC-NLP at SemEval-2017 Task 4: Sense n-grams for sentiment analysis in Twitter. In Proceedings of the 11th International Workshop on Semantic Evaluation. Vancouver, Canada, SemEval '17, pages 806–810.

[28] Yunxiao Zhou, Man Lan, and Yuanbin Wu. 2017. ECNU at SemEval-2017 Task 4: Evaluating effective features on machine learning methods for Twitter message polarity classification. In Proceedings of the 11th International Workshop on Semantic Evaluation. Vancouver, Canada, Semeval 2017, pages 811–815.

[29] Peter F Brown, Peter V Desouza, Robert L Mercer, Vincent J Della Pietra, and Jenifer C Lai. 1992. Class-based ngram models of natural language. Computational linguistics 18(4):467–479.

[30] OlutobiOwoputi, Brendan O'Connor, Chris Dyer, Kevin Gimpel, Nathan Schneider, and Noah A Smith. 2013. Improved part-of-speech tagging for online conversational text with word clusters. Association for Computational Linguistics.

[31] Bengio, Yoshua, Aaron Courville, and Pascal Vincent. "Representation learning: A review and new perspectives." Pattern Analysis and Machine Intelligence, IEEE Transactions on 35.8 (2013a): 1798-1828.

[32] Chung, Junyoung, CaglarGulcehre, KyungHyun Cho, and YoshuaBengio. "Empirical evaluation of gated recurrent neural networks on sequence modeling." arXiv preprint arXiv:1412.3555 (2014).

[33] Lin, Zhouhan, MinweiFeng, Cicero Nogueira dos Santos, Mo Yu, Bing Xiang, Bowen Zhou, and YoshuaBengio. "A structured self-attentive sentence embedding." arXiv preprint arXiv:1703.03130 (2017).

[34] Wang, Yequan, Minlie Huang, and Li Zhao. "Attention-based lstm for aspect-level sentiment classification." Proceedings of the 2016 conference on empirical methods in natural language processing. 2016.