Phrase-Level Emotion Intensity Detection of Text using Lexicon-based Unit Circle and Pipelined Neural Networks Approaches

Bhuvaneshwari Ajit Patil[1], Vishnu Vardhan Vadlakunta[1],
Poojaa Pranathi Guruvu[1], Meenu Mathew[1], Jay Prakash[1]

[1]Computer Science and Engineering, National Institute of Technology, Calicut, India.

E-mail(s): bhuvaneshwari9351@gmail.com; vishnuvardhan24603@gmail.com; poojaapranathiguruvu@gmail.com; meenu_p190104cs@nitc.ac.in; jayprakash@nitc.ac.in;

Abstract

Emotions have a significant impact on how people make decisions. Due to its potential applications in various fields, emotion intensity detection has attracted a lot of attention recently. Several methods have been proposed in the past using natural language processing to recognize emotions from the text. These methods include the keyword-based approach, the lexicon-based approach and the machine learning approaches. Nevertheless, there were a few limitations with the lexicon-based unit circle approach, as it focuses on detecting emotion intensity at the word-level. This research proposes an integrated model by considering the advantages of the lexicon-based unit circle and pipelined neural networks approaches. The model is trained on a large corpus of text data. The proposed approach aims to determine emotion intensity, analyzing text at the phrase-level, thereby achieving higher performance in terms of accuracy and F1 score compared to other models.

Keywords: Natural language processing . Lexicon-based approach . Emotion intensity . Neural networks approach . Phrase-level

1 Introduction

Emotions play an important role in regulating human behavior towards society. Hence, there is a need to analyze emotions properly to interact with the world in fields like customer feedback and mental health checkups. Emotion intensity detection remains an underexplored area in research. Since emotions exist in a spectrum ranging from subtle to intense, the solution is to measure emotion intensity instead of detecting sentiment using a binary system i.e., positive or negative.

Existing emotion models can be categorized into two types, namely categorical and dimensional, based on various emotion theories [1]. The classification of emotions in categorical emotion models involves the establishment of distinct categories that are separate from one another. In contrast, dimensional emotion models establish a limited number of dimensions accompanied by certain parameters and specify emotions based on these dimensions. Most dimensional emotion models typically employ two or three dimensions. For example, Plutchick's wheel of emotions and Valance, Arousal, Dominance (VAD) model are of two and three dimensions respectively. Robert Plutchik, an American psychologist introduced the wheel of emotions in the 1980s as part of his psycho-evolutionary theory of emotions [2]. Plutchik's model organizes emotions into a circular diagram, depicting eight primary emotions arranged radially with varying intensities. This multidimensional approach allows for a more nuanced understanding of emotional expressions compared to simpler models. The wheel illustrates relationships between emotions through their placement in proximity to each other. Emotions that are adjacent in the wheel share common attributes and transform into one another. This framework aids in understanding how emotions are interconnected and influence each other.

The motivation behind developing the emotion intensity detection model for text using phrase-level analysis arises from the need to capture the subtle and multifaceted emotional expressions embedded within the text. The existing emotion detection models have not paid attention to the variations in emotions that exist at the phrase-level, which can be crucial for understanding the true sentiment and emotional context of the information being conveyed. Methods such as keyword-based approach, lexicon-based approach and machine learning approaches can be adopted for emotion intensity detection of text.

Lexicon is referred to as a set of meaningful units in a language i.e., a dictionary. The lexicon-based approach [3–6] relies on a curated list of words and phrases associated with emotions and their intensity scores. By matching words to entries in the lexicon, it quantifies emotions in text. A pipeline model [7] consists of two neural networks, one for determining the intensity scores of seed words in the sentence and another for determining the shift in intensity scores due to modifiers.

The lexicon-based unit circle model [3], has limitations since it fails to account for the impact of words preceding negative words (such as not, no, except, etc.) on changes in polarity and intensity scores. The influence caused by the negative words and modifiers is considered in the pipelined neural networks model [7]. However, combining the lexicon-based unit circle model and pipelined neural networks model leads to an integrated model for emotion intensity detection analyzing at the phrase-level. Further, this model improves the capacity of existing emotion intensity detection models and aids in fields like mental health support by offering accurate and context-aware emotional perspectives to benefit both individuals and applications in the digital domain.

Thus, the major contributions of the present research are summarized as follows:

- This research is the first to detect emotion intensity using a lexicon-based approach analyzing the text phrase-by-phrase at the sentence-level.
- The proposed unit circle model using phrase-level analysis extends the lexicon-based unit circle model [3] and addresses the limitations of the existing model.
- The proposed method deals with eight emotions from the second level of Plutchik's emotion wheel namely anger, anticipation, disgust, fear, joy, sadness, surprise and trust by portraying them in a unit circle.
- Modifications are proposed for the NRC Emotion Intensity Lexicon (NRC-EIL) [8] to address the issue of words labeled with multiple emotions.
- This research analyzes the proposed integrated model with various parameters and compares the results with the existing model.

The remaining sections of the paper are structured as follows. Section 2 presents a review of various methods that are used for emotion and emotion intensity detection. Section 3 explains the detailed methodology along with the algorithm adopted in the research. Section 4, provides the detailed analysis of the results. It also contains the details of the experimental studies with different parameters of the model and comparisons of the proposed integrated model with existing emotion and emotion intensity detection models. Section 5 concludes the paper by highlighting the contributions of the research and providing insight into future work.

2 Literature review

In the literature, very few works have been reported on emotion intensity detection of text. Recently, some models have been proposed with different datasets and results. Tables 1 and 2 summarise the salient aspects of the literature related to the present research. Researchers have adopted mainly two approaches namely, lexicon-based approach and machine learning approaches [9]. The Lexicon-based approach is computationally less complex than the machine learning approach (in terms of computational resources). Despite the fact that, machine learning approaches are more accurate compared to the lexicon-based approach.

2.1 Lexicon-based approach

A lexicon is an entity that contains semantic, grammatical information about words. Sentiment polarity is determined by assessing the semantic values of words in the sentence against the lexicon. This approach calculates the sentiment orientations of the set of sentences from the semantic orientation of lexicons. Semantic orientation can be positive, negative or neutral. The lexicon dictionary can be generated automatically or created manually. The binary classification of financial news text based on the determination of positive or negative sentiment provides an accuracy of 46% [5]. A similar approach applied to a dataset classifying 50,000 movie reviews resulted in an accuracy of 76.585% [6]. A method has been introduced to calculate teachers' performance by analyzing feedback provided by students, achieving an accuracy of 86% [4]. The emotion scores generated by this model are determined as either positive (+) or negative (-) based on the respective weights assigned to each type of term.

The alternate approach for emotion intensity recognition is to represent Plutchik's wheel of emotions in a unit circle having polarity on the x-axis and intensity on the y-axis, as reported in [3]. The intensity and polarity are calculated for each word with the help of the unit circle equation. After calculating the intensity and polarity for each word, measures of central tendencies like mean or median are computed. Then Euclidean distances from reference points are calculated to the coordinates (polarity, intensity) and the least distanced emotion is the predicted one. This model was trained across different datasets and it achieved overall higher accuracy, which is comparatively better than other lexicon-based models while performing multi-level classification of emotions. The present research extends the model [3] to phrase-level analysis.

2.2 Machine Learning approaches

In machine learning, there are two types of approaches - supervised and unsupervised. Both approaches can be used to determine the emotion intensity of text. In these approaches, the model is trained with a part of the dataset and tested with the remaining part. Supervised models work on labeled data, whereas unsupervised models work on data that is not labeled with classes.

The machine learning approach, exemplified by Bidirectional Encoder Representations from Transformers (BERT) [10, 11] is a neural network model. It is designed for various Natural Language Processing (NLP) tasks and can be fine-tuned for emotion intensity detection. The model developed by researchers at Google features connections between every output and input element, with dynamically calculable weights between them. The BERT model preserves the original context whereas the traditional word embedding models like Word2Vec fail to do so [12]. A dataset (DENS) has been introduced for multi-class emotion analysis. This dataset comprises approximately 9,710 manually annotated passages, which were trained on the BERT model, resulting in a micro-F1 score of 60.4% [10]. Furthermore, a multilingual emotion dataset has been introduced, consisting of 30,000 English and 25,000 Finnish sentences, with labeling based on Plutchik's core emotions. The dataset was trained using the BERT model, resulting in an accuracy of 54.4% [11].

The Convolutional Neural Network (CNN) represents a class of artificial neural networks that have risen to prominence in various computer vision tasks, and it also has prominent applications in a variety of domains like NLP. It is designed to automatically and adaptively learn different hierarchies of features through back-propagation by using multiple building blocks, such as convolution layers, pooling layers, and fully connected layers. It is a multi-layered artificial neural network that can detect complex features in data.

The lexicon-based approach cannot completely utilize contextual information due to its word-level performance. A context-dependent lexicon-based model utilizing CNN has been developed to accurately extract the strength of emotion from words or text [13]. This model generates the Sentiment Strength Specific lexicon (SSS-Lex), which includes word associations along with their intensity scores. It has demonstrated superior accuracy in predicting emotion intensity compared to baseline methods such as Yelp, TS-Lex, and Sentiment140.

A pipelined neural networks model for phrase-level sentiment intensity prediction has two neural networks arranged in a pipelined fashion where the first neural network will deal with the intensity of individual words by calculating the semantic similarities of unknown words with seed words [7]. The second neural network will calculate the change in intensity due to the presence of modifiers and intensifiers. If there are multiple modifiers in a single phrase, then the process is performed recursively starting from the modifier which is the closest to the word. In the second neural network instead of using the exact word for the modifier, the embedding vector is used as some of the modifiers may not always have an adequate amount of training samples. For predicting emotion intensity for both the uni and multi-word phrases, this model successfully achieved higher performances as Kendall's τ value of 0.743 and Spearman's ρ value of 0.913.

Table 1: Summary of the literature

Method	Year	Model proposed	Dataset used	Limitations
Sentiment Polarity Identification in Financial News: A Cohesion-based Approach [5]	2007	Lexicon-based model	Financial news text	The algorithm inclines towards the positive emotions over the negative ones.
Unsupervised Emotion Detection from Text Using Semantic and Syntactic Relations [14]	2012	Machine learning model	ALM, ISEAR	Semantic relatedness scores of words depend upon the corpus from which they have been derived.
Pipelined Neural Networks for Phrase-level Sentiment Intensity Prediction [7]	2016	Pipelined neural networks model	ANEW [15], SemEval [16], [17] and SST [18]	Prediction task can be extended to sentence-level or document-level.
Recognizing emotions in text using ensemble of classifiers [19]	2016	Hybrid model	ISEAR, Affective texts datasets	The system performance was impacted by non-robust classification technique.
Keyword Based Emotion Word Ontology Approach for Detecting Emotion Class from Text [20]	2016	Keyword-based model	Not reported	Ambiguity in keyword definitions and lack of linguistic information
Sentiment Analysis of Tweets Including Emoji Data [21]	2017	Machine learning model	Tweets of size $\leq$ 140 characters	An effective classification technique is necessary to enhance performance.
Emotion Intensity Detection for Social Media Data [22]	2017	Hybrid model	WASSA	The amount of data is relatively little.
A Survey on Emotion Detection [23]	2017	Lexicon-based model	Newslines, Textbooks and direct speeches	Few categories are considered
Identifying Emotions in Social Media: Comparison of Word-Emotion Lexicons [24]	2017	Lexicon-based model	Facebook and twitter messages	Limited vocabulary from the lexicon.
Emotion Detection on Twitter Data using Knowledge Base Approach [25]	2017	Lexicon-based model	Twitter messages	The proposed model has a low generalization ability due to the limited number of categories.
SEDAT: Sentiment and Emotion Detection in Arabic Text [26]	2018	Machine Learning model	Arabic tweets	The amount of data is relatively little.
DENS: A Dataset for Multi-class Emotion Analysis [10]	2019	BERT model	9,710 Manually Annotated Passages	Do not consider the semantic context of words in the emotion analysis.

Table 2: Summary of the literature (continued)

Method	Year	Model proposed	Dataset used	Limitations
Actionable Pattern Discovery for Tweet Emotions [27]	2019	Hybrid model	Tweets	Not good for generalization due to the restricted number of emotion classes.
Bilingual Lexicon Approach to English-Filipino Sentiment Analysis of Teaching Performance [4]	2020	Lexicon-based model	1,699 Sentences	Incorrect classification when words that belong to more than one parts of speech are encountered. Need a POS tagger.
A Multilingual Dataset for Sentiment Analysis and Emotion Detection [11]	2020	BERT model	30,000 English sentences and 25,000 Finnish sentences	The performance depends upon the initial manual effort annotating thousands of sentences in different languages.
Sentiment strength detection with a context-dependent lexicon-based convolutional neural network [13]	2020	CNN model	6 real world datasets	Limited dataset coverage, which may not fully represent sentiment analysis diversity.Uncertainty about the model's generalizability across different tasks, datasets, and languages.
Sentence-Level Emotion Detection from Text Based on Semantic Rules [28]	2020	Lexicon-based model	ISEAR	Disregard for the contextual meaning of words.
Analyzing sentiment system to specify polarity by lexicon-based [6]	2021	Lexicon-based model	50,000 movie reviews	Reliance on a single dataset from IMDb.The proposed model's scalability and generalizability to larger datasets or real-time applications are not thoroughly evaluated.
Lexicon-based Sentence Emotion Detection Utilizing Polarity Intensity Unit Circle Mapping and Scoring Algorithm [3]	2022	Lexicon-based unit circle model	9,921 scored words	Do not consider effect due to modifiers and intensifiers.
Text-Based Emotion Recognition Using Deep Learning Approach [29]	2022	Deep learning model	ISEAR, WASSA, Emotion stimulus	Need for addressing potential biases in the datasets, handling nuances in regional languages.
Emotion Detection Using a Bidirectional Long-Short Term Memory (BiLSTM) Neural Network[20]	2023	Bi-LSTM model	Twitter emotional Dataset	Inability to recognize sentences without emotional keywords, ambiguity in keyword definitions.

3 Methodology

The proposed method is an integrated approach consisting of the lexicon-based unit circle and pipelined neural networks models [3, 7] to predict the resultant emotion. This approach combines the strengths of neural networks for modifier weight learning and a lexicon-based approach for emotion mapping. The structured framework of Plutchik's wheel is used for emotion categorization and the unit circle algorithm serves as a mapping mechanism for final emotion prediction.

3.1 Architecture of the unit circle model at phrase-level analysis

Initially, the dataset is preprocessed and words are grouped into phrases. After grouping the words, unknown words (words that are not present in the lexicon) are replaced with respective synonyms. The polarity and intensity are computed using the emotion scores from the modified NRC-EIL and the reference points of the emotions. The values thus obtained along with the word embeddings of the modifier are passed to the neural networks model to calculate the modifier polarity and modifier intensity. Measures of central tendencies like mean and median are computed for the calculated phrase-polarities and phrase-intensities to obtain the resultant sentence-polarity and sentence-intensity which are later mapped to the emotions in the Plutchik's wheel using the unit circle algorithm [3].

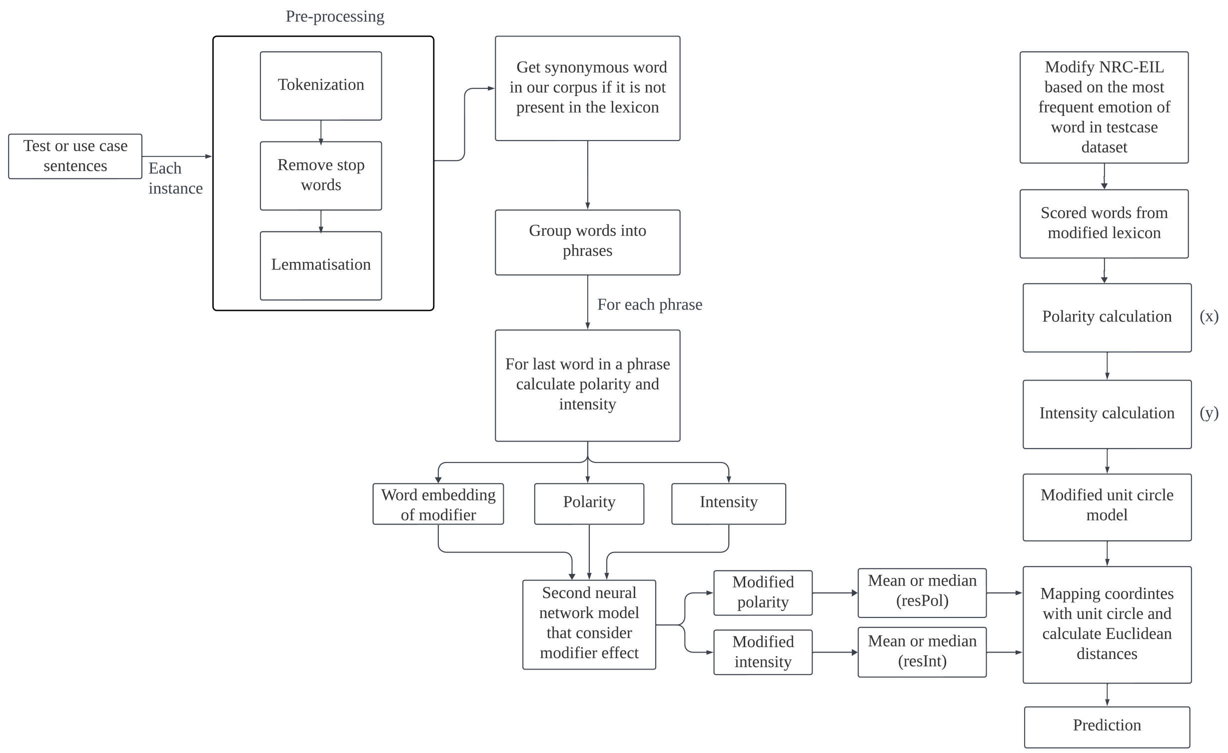


Fig. 1: Proposed methodology

3.1.1 Data preprocessing

To preprocess the data, tokenization is initially performed, which breaks down the text into smaller units. Then the stop words ('the', 'is', 'and', etc.) are removed as these words do not significantly affect emotion and are often excluded to reduce noise in the data. Stop word removal is followed by lemmatization, which converts the word to its root form (walking, walks, walked to walk). This increases the probability for the word to be present in the lexicon. The sentiment words that are not present in the lexicon are replaced with the respective synonyms that are present in the lexicon. Words are grouped into phrases such that the last word of the phrase is the sentiment word and the remaining all are modifiers (e.g. 'walking', 'very', 'good' to ['walking'], ['very good']).

The words in the lexicon which have an emotion score less than a threshold value (e.g. 0.15 or 0.3) are removed as those words do not have a significant impact on determining the resultant emotion. In this research, these threshold values are termed as Emotion Score Threshold (EST).

3.1.2 Polarity and intensity scoring

For the word having multiple emotions in the lexicon, NRC-EIL [8] is modified by taking the most frequently occurring emotion in the testcase dataset. The words present in modified NRC-EIL are also lemmatized to maintain consistency with preprocessed text data. Polarities and intensities are calculated for the sentiment words obtained after data preprocessing by multiplying the score of the word present in the modified NRC-EIL with x and y-coordinates of the reference points that are referred from UCM [3] (refer Equations 1 and 2). The x-coordinate represents the polarity of emotion, while the y-coordinate represents the intensity of the emotion associated with the word.

$$\text{polarity} = \text{emotion score} \times x\text{-coordinate of reference point} \tag{1}$$

$$\text{intensity} = \text{emotion score} \times y\text{-coordinate of reference point} \tag{2}$$

A noval approach is proposed for the calculation of polarity and intensity scores to avoid misclassification of the emotion of the word. For example, suppose there is the word 'elation' in the lexicon with a score of 0.944 mapped to the emotion 'joy'. According to the UCM [3], the polarity and intensity values obtained are 0.944 and 0.33 (polarity is calculated using emotion scores in the NRC-EIL and intensity is calculated using unit circle equation). The obtained coordinates of the word 'elation' (0.944, 0.33) fall into the trust region of the unit circle whereas it is mapped to 'joy' in the lexicon. However, according to the proposed approach obtained coordinates (0.472, 0.818) map the word 'elation' into the 'joy' region.

3.1.3 Modified weight learning

Modified polarity and intensity values are obtained using the second neural network of the pipelined neural networks method [7] that considers the effect due to modifiers. If more than one modifier is present in the phrase, then this neural network will calculate the modified intensity and polarity values by taking an average of word embeddings of the modifiers. Hence, all modifiers present in the phrase are considered as a single modifier (e.g. ['very', 'much', 'happy'] as ['very much', 'happy']).

To calculate the modified polarity and intensity values, word embedding of the modifier, polarity and intensity values are passed to the neural network.

3.1.4 Measures of central tendencies

For the modified values obtained as described in the previous subsection, measures of central tendencies (mean, median) are computed to obtain the resultant polarity and intensity of the sentence. The final x and y coordinates of the sentence are ($resPol$, $resInt$).

3.1.5 Emotion mapping

The Plutchik's wheel contains eight emotions and specific coordinates are assigned to each emotion on to the unit circle, which are referred to as reference points [3]. The Euclidean distances from those eight reference points to the resultant x and y coordinates ($resPol$, $resInt$) obtained previously are calculated. The emotion that has a minimum Euclidean distance from the resultant coordinates is considered as the predicted emotion of the sentence.

The algorithm adopted in this research is presented below (refer Algorithm 1).

Algorithm 1 Algorithm for unit circle model using phrase-level analysis

Data: Testcase sentences, modified NRC-EIL
Result: Predict emotion of the sentence

```
Preprocess the test case sentences
Lemmatize the words in the lexicon
for each word in test case sentences do
    if word is not in lexicon then
        Replace the word with its respective synonym in the lexicon
    end if
end for
group words into phrases
for each word in lexicon do
    Calculate polarity, intensity and store values in the lexicon
end for
for each sentence do
    for each phrase in sentence do
        W ← word embedding of the modifier
        P ← Polarity of the sentiment word (last word) in the phrase
        I ← Intensity of the sentiment word (last word) in the phrase
        Pass W, P, I into second neural network to obtain modified polarity and intensity
    end for
    Compute mean or median for values of all phrases in a sentence to obtain resultant polarity and intensity
    Map resultant polarity and intensity with the unit circle to predict the emotion
end for
```

4 Results and analysis

Experiments have been conducted to compare and analyze the results of the existing Lexicon-based Unit Circle Model (UCM)[3], modified Lexicon-based Unit Circle Model (Mod-UCM) and proposed Unit Circle Model using Phrase-Level Analysis (UCM-PLA).

In the UCM approach, NRC-EIL [8] is used as the standard lexicon, whereas for Mod-UCM and UCM-PLA the modified NRC-EIL is used. A combination of the two datasets (XED dataset and Emotion dataset for NLP) is used to analyze the models. It consists of nearly 28,500 annotated sentences. The lexicon used for models, synonyms and Emotion Score Threshold (EST) are the parameters considered for comparison of UCM and Mod-UCM. In addition, word embeddings and the number of epochs are considered for UCM-PLA. Measures of central tendency (mean, median) and performance of the models with balanced and imbalanced datasets are also analyzed.

4.1 Data sources

There are two different types of datasets used for training the UCM-PLA. The dataset modified NRC-EIL is used for the lexicon-based unit circle model which is the modification of NRC-Emotion-Intensity-Lexicon [8] dataset. The second neural network of pipelined neural networks model learned the modifier weights using the set of phrases which are collected from SemEval datasets [16, 17]. It consists of 9,829 words mapped with emotions from Plutchik's wheel and the respective score of the emotion. The test-case sentences are combination of XED dataset [11] and Emotions dataset for NLP classification tasks (https://www.kaggle.com/datasets/praveengovi/emotions-dataset-for-nlp).

In the testcase dataset, it is observed that the emotion 'joy' has a higher number of sentences compared to 'disgust' and 'trust' which have relatively fewer occurrences. This result is evident from fig 2 as shown below.

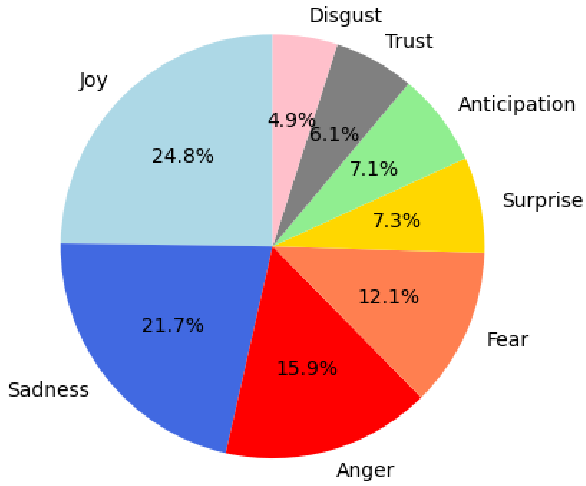


Fig. 2: Emotion distribution

Tables 3 and 4 reveal the distribution of the number of scored words for each emotion class from NRC-EIL and modified NRC-EIL respectively. The scored words that are associated with surprise and anticipation classes are low in number in comparison with the other classes.

Table 3: NRC Emotion Intensity Lexicon (NRC-EIL) [8]

Emotion	Number of Words	Mean score
Anger	1481	0.499623
Anticipation	862	0.500283
Disgust	1092	0.500697
Fear	1763	0.499044
Joy	1264	0.501782
Sadness	1294	0.501991
Surprise	583	0.500422
Trust	1490	0.510451

Table 4: Modified NRC-EIL

Emotion	Number of Words	Mean score
Anger	658	0.517386
Anticipation	329	0.505304
Disgust	394	0.537797
Fear	738	0.490331
Joy	885	0.526272
Sadness	599	0.500227
Surprise	184	0.558266
Trust	934	0.515723

4.2 Evaluation metrics

For the purpose of analysis, evaluation metrics considered are accuracy, weighted precision, weighted recall and weighted F1-score (refer Equations 3, 4 and 5).

Weighted Precision:

$$\text{Precision}_i = \frac{\text{True Positives}_i}{\text{True Positives}_i + \text{False Positives}_i}$$

$$\text{Weighted Precision} = \frac{\sum_{i=1}^{N} \text{Precision}_i \times \text{Support}_i}{\sum_{i=1}^{N} \text{Support}_i} \tag{3}$$

Weighted Recall:

$$\text{Recall}_i = \frac{\text{True Positives}_i}{\text{True Positives}_i + \text{False Negatives}_i}$$

$$\text{Weighted Recall} = \frac{\sum_{i=1}^{N} \text{Recall}_i \times \text{Support}_i}{\sum_{i=1}^{N} \text{Support}_i} \tag{4}$$

Weighted F1 Score:

$$\text{F1 Score}_i = \frac{2 \times \text{Precision}_i \times \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}$$

$$\text{Weighted F1 Score} = \frac{\sum_{i=1}^{N} \text{F1 Score}_i \times \text{Support}_i}{\sum_{i=1}^{N} \text{Support}_i} \tag{5}$$

Note: $Support_i$ is the number of true labels in the $class_i$.

4.3 Comparison of unit circle and modified unit circle model

Various parameters are considered in Tables 5, 6, 7 and 8 for comparison of UCM and Mod-UCM. The detailed analysis of each parameter and the value of the standard parameter recommended for Mod-UCM is presented in the subsections provided below. It is noted that irrespective of any parameter, the Mod-UCM achieves better metrics than that of the UCM in all the cases.

4.3.1 Analyzing the lexicon used for models

The NRC-EIL [8] consists of 9,829 words mapped with emotions from Plutchik's wheel. Some of the words are mapped with more than one emotion. For example, the word 'fury' is mapped with three emotions (anger, fear and sadness) having different scores. From Tables 5, 6, 7 and 8, it can be inferred that the modified NRC-EIL achieves better results compared to the NRC-EIL in every case. Using the modified NRC-EIL in UCM [3] leads to an increase in accuracy by approximately 2% on average. In the Mod-UCM approach, modified NRC-EIL is used as the standard lexicon. Using NRC-EIL in the Mod-UCM approach leads to a reduction in accuracy by approximately 3% due to the reason that, contains words mapped with multiple emotions having different scores in NRC-EIL.

It can be concluded that modification performed in the NRC-EIL improves the performance of the model and reduces the mispredictions of the emotion.

4.3.2 Analyzing the synonym replacement

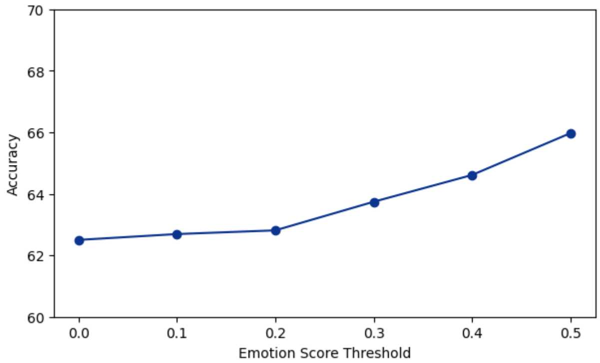
The NLTK Wordnet is used to obtain synonyms of the words in sentences that are not present in the lexicon. From Tables 5, 6, 7 and 8, it can be noted that better results are obtained without synonym replacement of the words. The synonym replacement leads to the reduction in the accuracy of both models by approximately 5% on average. This reduction is due to the reason that, for many words that are not present in the lexicon, there may not exist a synonym conveying the same meaning and a word can have more than one meaning depending on the context. For example, the word 'do' is not present in the lexicon and the word that is synonymous with it and present in the lexicon is 'practice'. However, these words do not convey similar meanings in a particular context, which leads to the misclassification of the emotion of the sentence.

It can be concluded that synonym replacement of the words leads to a reduction in the accuracy of the model. Hence, for the Mod-UCM without performing synonym replacement is recommended as the standard approach.

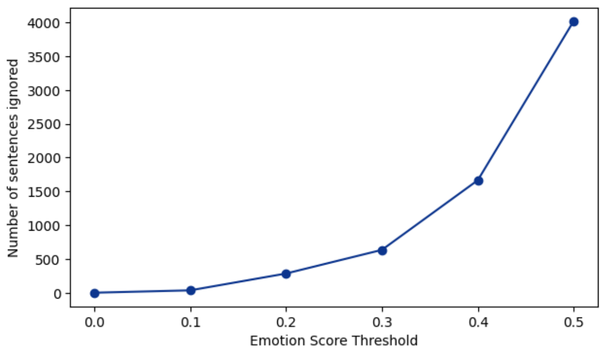
4.3.3 Analyzing the Emotion Score Threshold

The Emotion score Thresholds (ESTs) considered for the analysis are 0, 0.15 and 0.3. From Tables 5, 6, 7 and 8, it can be noted that better results are obtained for the EST value 0.3 than 0 or 0.15. Using 0.15 and 0.3 for the EST has demonstrated enhanced performance compared to performance achieved without EST (EST value of 0). It can also be inferred that setting the EST as 0.3 achieves comparatively higher accuracies in almost all the cases. This inference is due to the reason that, words having low emotion scores do not contribute much in predicting resultant emotion and these low scores affect the resultant scores obtained after applying measures of central tendencies (mean, median).

From Fig. 3a, it can be observed that the increase in the EST value leads to a subsequent increase in accuracy. However, Figure 3b demonstrates that numerous pivotal words in the lexicon are eliminated, causing the model to disregard multiple sentences. A sharp escalation in the number of ignored sentences is noticeable when the EST value is elevated from 0.3 to 0.4. Consequently, 0.3 emerges as the optimal value for EST.



(a) Comparison of accuracies for different ESTs



(b) Sentences ignored for different ESTs

Fig. 3: Results of the analysis for EST

It can be concluded that the EST value of 0.3 improves the overall performance in both models. For the Mod-UCM, the EST value of 0.3 is recommended as the standard value.

Table 5: Lexicon-based Unit Circle Model (UCM) with mean approach

			Balanced dataset				Imbalanced dataset			
Lexicon	Synonyms	EST[1]	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
NRC-EIL[2]	with	0	48.1926%	0.5319	0.4819	0.4614	50.6859%	0.5325	0.5069	0.4809
		0.15	48.0117%	0.5319	0.4801	0.4582	50.6923%	0.5347	0.5069	0.4793
		0.3	48.0714%	0.5362	0.4807	0.4576	50.7717%	0.5367	0.5077	0.4794
	without	0	53.1105%	0.5609	0.5311	0.5249	55.2356%	0.5652	0.5524	0.5425
		0.15	53.1481%	0.5623	0.5315	0.525	55.5101%	0.5693	0.5551	0.5448
		0.3	53.6363%	0.586	0.5364	0.5266	56.0910%	0.5851	0.5609	0.5469
Modified NRC-EIL	with	0	50.0896%	0.5559	0.5009	0.4872	52.1774%	0.5485	0.5218	0.4976
		0.15	49.9871%	0.5582	0.4999	0.4850	52.2118%	0.5510	0.5221	0.4964
		0.3	49.8641%	0.5556	0.4986	0.4820	52.3045%	0.5520	0.5230	0.4963
	without	0	56.0315%	0.5918	0.5603	0.5567	57.8316%	0.5903	0.5783	0.5692
		0.15	56.2076%	0.5950	0.5621	0.5581	58.2337%	0.5959	0.5823	0.5728
		0.3	**56.9347%**	**0.6174**	**0.5693**	**0.5632**	**59.2063%**	**0.6146**	**0.5921**	**0.5805**

Table 6: Lexicon-based Unit Circle Model (UCM) with median approach

			Balanced dataset				Imbalanced dataset			
Lexicon	Synonyms	EST[1]	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
NRC-EIL[2]	with	0	45.6507%	0.6263	0.4565	0.4326	48.7675%	0.5790	0.4877	0.4442
		0.15	45.3778%	0.6232	0.4538	0.4284	48.6171%	0.5777	0.4862	0.4411
		0.3	45.5258%	0.6193	0.4553	0.427	48.8325%	0.5749	0.4883	0.4427
	without	0	49.5515%	0.6265	0.4955	0.4899	53.4801%	0.5971	0.5348	0.5128
		0.15	49.5419%	0.6258	0.4954	0.4895	53.6239%	0.5986	0.5362	0.5142
		0.3	50.0151%	0.6293	0.5002	0.4886	54.0187%	0.6047	0.5402	0.5154
Modified NRC-EIL	with	0	46.3883%	0.6578	0.4639	0.4520	49.6039%	0.6198	0.4960	0.4582
		0.15	46.1390%	0.6575	0.4614	0.4482	49.5116%	0.6207	0.4951	0.4557
		0.3	46.1508%	0.6488	0.4615	0.4454	49.6976%	0.6154	0.4970	0.4566
	without	0	52.3352%	0.6662	0.5234	0.5263	55.8493%	0.6365	0.5585	0.5412
		0.15	52.3287%	0.6655	0.5233	0.5255	56.0238%	0.6382	0.5602	0.5426
		0.3	**53.0955%**	**0.6706**	**0.5310**	**0.5291**	**56.8679%**	**0.6453**	**0.5687**	**0.5498**

Table 7: Modified Unit Circle Model (Mod-UCM) with mean approach

			Balanced dataset				Imbalanced dataset			
Lexicon	Synonyms	EST[1]	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
NRC-EIL[2]	with	0	51.0665%	0.5522	0.5107	0.4932	53.4081%	0.5739	0.5341	0.5190
		0.15	51.2912%	0.5559	0.5129	0.4954	53.4984%	0.5760	0.5350	0.5196
		0.3	51.2085%	0.5545	0.5121	0.4948	53.4629%	0.5737	0.5346	0.5191
	without	0	56.3223%	0.5990	0.5632	0.5582	58.2093%	0.6096	0.5821	0.5777
		0.15	56.6381%	0.6028	0.5664	0.5612	58.4613%	0.6136	0.5846	0.5801
		0.3	56.8333%	0.6153	0.5683	0.5640	58.8087%	0.6250	0.5881	0.5842
Modified NRC-EIL	with	0	54.2905%	0.5948	0.5429	0.5287	55.6507%	0.5965	0.5565	0.5405
		0.15	54.3484%	0.5957	0.5435	0.5293	55.7415%	0.5981	0.5574	0.5414
		0.3	54.5349%	0.5978	0.5453	0.5319	55.8710%	0.5989	0.5587	0.5430
	without	0	61.0888%	0.6509	0.6109	0.6082	62.1903%	0.6433	0.6219	0.6151
		0.15	61.2257%	0.6530	0.6123	0.6093	62.4570%	0.6470	0.6246	0.6177
		0.3	**61.7602%**	**0.6669**	**0.6176**	**0.6158**	**63.2003%**	**0.6607**	**0.6320**	**0.6262**

Table 8: Modified Unit Circle Model (Mod-UCM) with median approach

Lexicon	Synonyms	EST[1]	Balanced dataset				Imbalanced dataset			
			Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
NRC-EIL[2]	with	0	47.5029%	0.5680	0.4750	0.4476	50.7109%	0.5896	0.5071	0.4883
		0.15	47.6093%	0.5703	0.4761	0.4479	50.7333%	0.5926	0.5073	0.4882
		0.3	47.6086%	0.5772	0.4761	0.4503	51.0128%	0.6034	0.5101	0.4958
	without	0	50.6221%	0.5922	0.5062	0.4958	54.5384%	0.6055	0.5454	0.5360
		0.15	50.8652%	0.5953	0.5087	0.4986	54.8002%	0.6101	0.5480	0.5392
		0.3	51.3788%	0.6214	0.5138	0.5065	55.4672%	0.6336	0.5547	0.5494
Modified NRC-EIL	with	0	49.2316%	0.5989	0.4923	0.4649	51.4104%	0.5874	0.5141	0.4824
		0.15	48.8840%	0.5975	0.4888	0.4604	51.3680%	0.5900	0.5137	0.4814
		0.3	49.0749%	0.6064	0.4907	0.4631	51.7035%	0.5992	0.5170	0.4880
	without	0	53.9685%	0.6339	0.5397	0.5311	56.8730%	0.6205	0.5687	0.5515
		0.15	53.9870%	0.6350	0.5399	0.5309	57.0756%	0.6238	0.5708	0.5534
		0.3	**54.5675%**	**0.6548**	**0.5457**	**0.5386**	**57.9304%**	**0.6431**	**0.5793**	**0.5642**

Note: Weighted precision, recall and F1-score are considered.
The values are given in bold to highlight that particular configuration of the model achieving higher performance than all other configurations.

[1]Emotion Score Threshold

[2]NRC Emotion Intensity Lexicon [8]

Figure 4 provides the comparison of UCM and Mod-UCM. It is evident that Mod-UCM performs better than UCM.

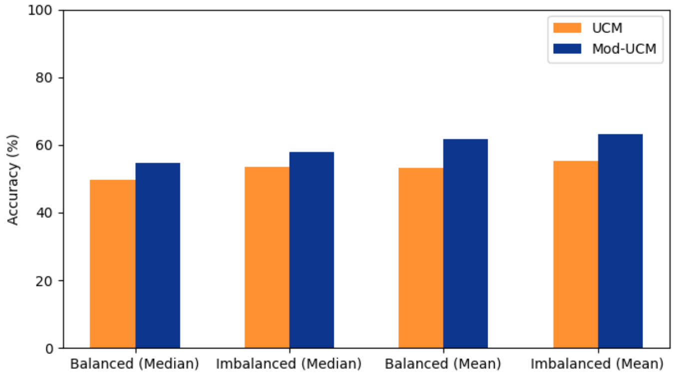


Fig. 4: Comparison of unit circle and modified unit circle model

4.4 Comparison of UCM, modified UCM and UCM-PLA

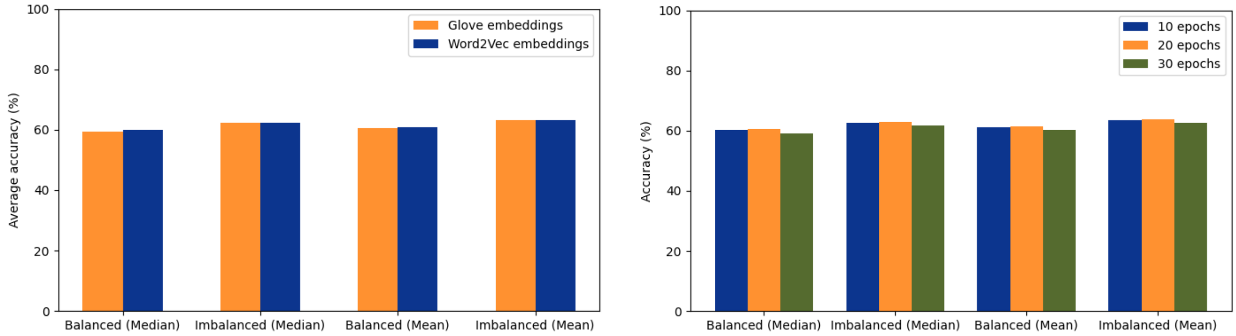
Various parameters are considered for experimenting with UCM-PLA. Further subsections contain a detailed analysis of each parameter along with the standard parameters recommended for UCM-PLA.

4.4.1 Analyzing the word embeddings

Word2Vec and glove word embeddings are considered for the analysis. Pipelined neural networks model [7] adopts the pre-trained Google News Word2Vec embeddings with vectors of dimension 300. Vectors of dimension 50 are used for glove embeddings. From Figure 5a it is observed that Word2Vec is slightly more accurate than glove embeddings. This might be because of the higher dimension word embedding vectors of word2vec. However, there is no notable difference in accuracies for both word embeddings. Word2Vec is recommended as the standard word embedding for UCM-PLA.

4.4.2 Analyzing the number of epochs in neural network

Epochs considered for the analysis are 10, 20 and 30. From Figure 5b, the model having 20 epochs is slightly more accurate than that of 10 and 30. However, there are not much differences in accuracies between 10, 20 and 30 epochs. As number of epochs exceeds 20, the model starts overfitting. For the UCM-PLA, 20 epochs are recommended as the standard parameter.



(a) Analysis of Glove and Word2Vec embeddings

(b) Analysis of the number of epochs for neural networks

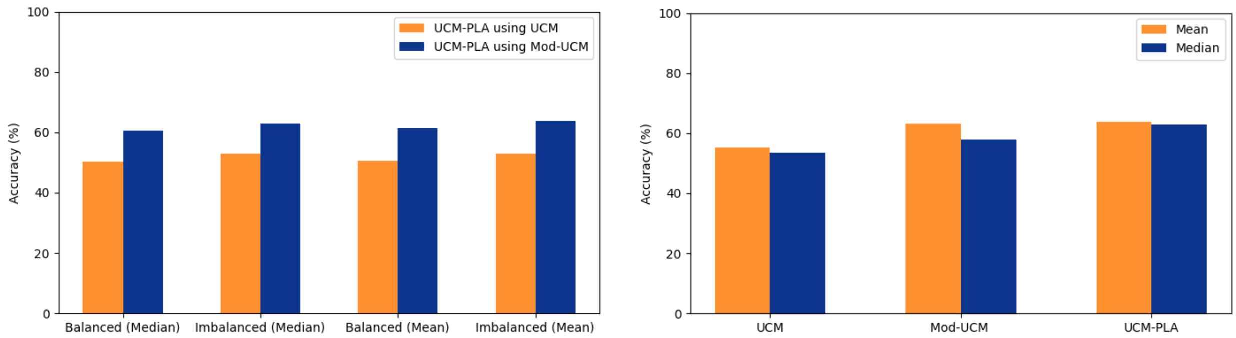
Fig. 5: Results of the analysis for word embeddings and number of epochs

4.4.3 Analyzing both the unit circle models for UCM-PLA

Unit circle models namely UCM and Mod-UCM are considered in the UCM-PLA for the analysis. From Fig 6a, it is observed that higher accuracies are obtained for Mod-UCM with its standard parameters (mentioned in the above subsection) for UCM-PLA than those of UCM by an average of 10.5%. This is due to the fact that the UCM-PLA performs better than UCM. Hence, the UCM-PLA with Mod-UCM is recommended for obtaining better results.

4.4.4 Analyzing the measures of central tendencies

Mean and median are the measures of central tendencies considered for the analysis. From Figure 6b, it can be observed that models using mean as a measure of central tendency perform better than those using median with the accuracy of 3% on average. Hence, the mean is recommended as the standard measure of central tendency for all three models.



(a) Comparison of UCM-PLA using UCM and Mod-UCM

(b) Comparison of mean and median approaches for all three models

Fig. 6: Results of the comparison of all three models and measures of central tendencies

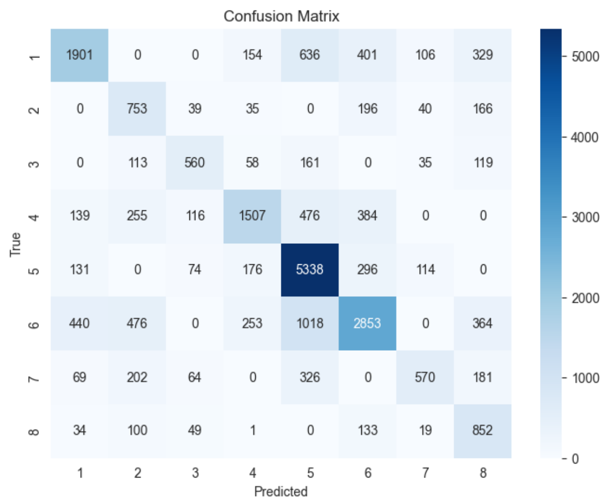
4.4.5 Analyzing the balanced and the imbalanced datasets

The imbalanced dataset contains 28,500 sentences, in which sentences labeled with joy are maximum in number and the ones labeled with disgust are minimum in number. The balanced dataset contains 8,000 sentences of which 1,000 sentences are present for each emotion label

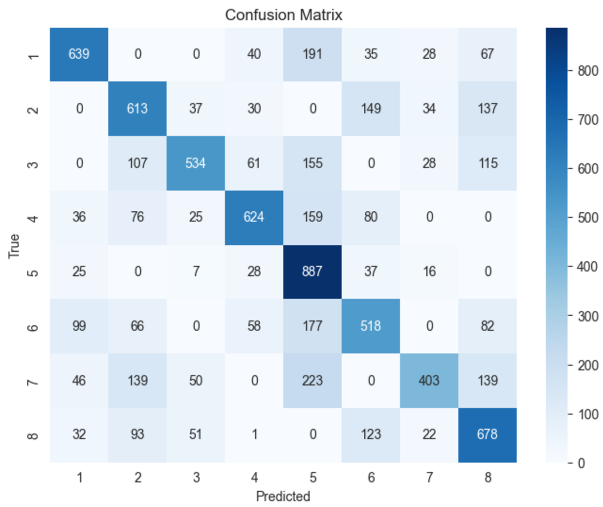
From the above analysis, it is noted that better results are obtained for the imbalanced dataset than for the balanced one. For UCM-PLA, the imbalanced dataset achieved an accuracy of 2.5% more than the balanced dataset. The sentences labeled with joy have been predicted more accurately than those of others and the number of sentences is also high, which leads to better performance for the imbalanced dataset over the balanced one.

Figures 7a and 7b outline the confusion matrices obtained for the UCM-PLA with standard parameters for imbalanced and balanced datasets respectively. From Figure 7b, it is observed that sentences labeled with the emotion '5' (joy) are predicted more accurately compared to others, while Figure 2 indicates that sentences labeled with 'joy' are more frequent. Consequently, the balanced dataset achieves lower accuracy compared to the imbalanced dataset.

Furthermore, it is observed from Figure 7b that sentences labeled with the emotion '7' (surprise) are predicted less accurately than others. This discrepancy can be attributed to the modified lexicon, where words labeled with the emotion 'surprise' are fewer in number (refer to Table 4). Hence, the lexicon-based models struggle to accurately predict sentences containing words associated with the emotion 'surprise'.



(a) Confusion matrix for imbalanced dataset



(b) Confusion matrix for balanced dataset

Fig. 7: Confusion matrices for both imbalanced and balanced datasets

Table 9 provides a summary of the comparison of the models. It can be inferred that, by considering standard parameters, the proposed unit circle model using phrase-level analysis with the modified unit circle model performs better than the existing lexicon-based unit circle model [3] in all the cases.

Table 9: Comparison of UCM-PLA with UCM and Mod-UCM

Model	Dataset	Mean/Median	Accuracy	Precision	Recall	F1-Score
UCM [3]	Balanced	Median	49.5515%	0.6265	0.4955	0.4899
		Mean	53.1105%	0.5609	0.5311	0.5249
	Imbalanced	Median	53.4801%	0.5971	0.5348	0.5128
		Mean	55.2356%	0.5652	0.5524	0.5425
UCM[1]	Balanced	Median	53.0955%	0.6706	0.5310	0.5291
		Mean	56.9347%	0.6174	0.5693	0.5632
	Imbalanced	Median	56.8679%	0.6453	0.5687	0.5498
		Mean	59.2063%	0.6146	0.5921	0.5805
Mod-UCM	Balanced	Median	54.5675%	0.6548	0.5457	0.5386
		Mean	61.7602%	0.6669	0.6176	0.6158
	Imbalanced	Median	57.9304%	0.6431	0.5793	0.5642
		Mean	63.2003%	0.6607	0.6320	0.6262
UCM-PLA	Balanced	Median	60.4000%	0.6413	0.6040	0.6027
		Mean	61.4500%	0.6453	0.6145	0.6123
	Imbalanced	Median	62.8921%	0.6476	0.6289	0.6243
		Mean	**63.7445%**	**0.6551**	**0.6374**	**0.6337**

Note: Weighted precision, recall and F1-score are considered.
The values are given in bold to highlight that UCM-PLA achieves higher performance than other models.

[1]As modified NRC-EIL is considered for all three models, the metrics obtained for UCM are different than those obtained for UCM [3].

4.5 Comparison of UCM-PLA with existing emotion detection models

The Table 10 illustrates the outcomes of the Unit Circle Model using Phrase-Level Analysis (UCM-PLA), in comparison to other existing lexicon-based approaches [5], [6], [4], and a machine learning approach [14]. The proposed integrated model is also evaluated using binary classification and six-class classification to compare UCM-PLA with other models that categorize the same number of classes. The 64% overall accuracy of this model, capable of categorizing eight classes, exceeds that of models handling fewer classifications.

Table 10: Performance Benchmarking

Models	Number of classes	Accuracy
Sentiment Polarity Identification in Financial News: A Cohesion-based Approach [5]	2	46%
Analyzing sentiment system to specify polarity by lexicon-based [6]	2	76%
Bilingual Lexicon Approach to English-Filipino Sentiment Analysis of Teaching Performance [4]	2	86%
Unit Circle Model using Phrase-Level Analysis (UCM-PLA) for Binary Classification	**2**	**89%**
Unsupervised Emotion Detection from Text using Semantic and Syntactic Relations [14]	6	57%
Unit Circle Model using Phrase-Level Analysis (UCM-PLA) for Classification with six classes	**6**	**67%**
Unit Circle Model using Phrase-Level Analysis(UCM-PLA)	**8**	**64%**

Note: The values are given in bold to highlight that the proposed integrated model achieves higher performance than the existing models with different number of classes.

5 Conclusion

This research extends the lexicon-based unit circle model [3]. The proposed models such as Mod-UCM and UCM-PLA outperform the existing models due to the noval approach for computing polarity and intensity scores. Further, it also introduces the modification of NRC-EIL [8], handling the words that are mapped with multiple emotions having different emotion scores. In addition, this work increases the capacity of the existing emotion intensity detection model from word-by-word at the sentence-level to phrase-by-phrase at the sentence-level. Integrating the modified lexicon-based unit circle model with the neural network handles the effect due to modifiers, intensifiers and negators like 'much', 'very', 'not' etc. The study also analyzes various parameters such as lexicon used for models, emotion score threshold, synonym replacement, word embedding, measures of central tendencies and number of epochs by using different evaluation metrics. The UCM-PLA outperforms the UCM [3] obtaining an accuracy of 64% by considering the standard parameters, whereas UCM obtained an accuracy of 51.4% on average.

There has been limited research undertaken on emotion intensity detection, with the lack of datasets mapping sentences to the eight emotions outlined in the Plutchik's wheel. In the future, this work can be extended by predicting emotions for sentences in multiple languages and creating new datasets for the experiments.

Declarations

Conflict of interest- The authors declare that they have no conflicts of interest.

References

[1] Calvo, R., Kim, S.: Emotions in text: Dimensional and categorical models. Computational Intelligence (2012)

[2] Plutchik, R.: Chapter 1 - a general psychoevolutionary theory of emotion. Theories of Emotion, 3–33 (1980)

[3] Timothy Walter G. Cuizon, H.S.A.: Lexicon-based sentence emotion detection utilizing polarity intensity unit circle mapping and scoring algorithm. 11th International Young Scientist Conference on Computational Science (2022)

[4] Pacol, C., Palaoag, T.: Bilingual lexicon approach to english-filipino sentiment analysis of teaching performance. IOP Conference Series: Materials Science and Engineering **vol. 1077**, 1–012044 (Feb. 2021)

[5] Devitt, A., Ahmad, K.: Sentiment polarity identification in financial news: A cohesionbased approach. Conference: ACL 2007, Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics **vol. 45**, 1 (June 2007)

[6] D. H. Abd, A.R.A., Sadiq, A.T.: Analyzing sentiment system to specify polarity by lexicon-based. Bulletin of Electrical Engineering and Informatics **vol. 10**, 1–283289 (Feb. 2021)

[7] Liang-Chih Yu, I.J.W.K.R.L. Member, Zhang, X.: Pipelined neuralnetworks for phrase-level sentiment intensity prediction. IEEE Transcation on Affective Computing **vol. xx**(no. x) (Dec. 2016)

[8] Mohammad, S.M.: Word affect intensities. In Proceedings of the 11th Edition of the Language Resources and Evaluation Conference (LREC-2018), Miyazaki, Japan (May 2018)

[9] W. Medhat, A.H., Korashy, H.: Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal **vol. 5**, 4–10931113 (Dec. 2014)

[10] C. Liu, M.O., Andrade, A.D.: Dens: A dataset for multi-class emotion analysis (2019)

[11] E. Ohman, K.K. M. P'amies, Tiedemann, J.: Xed: A multilingual dataset for sentiment analysis and emotion detection (2020)

[12] Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding (2019)

[13] Minghui Huang, Y.R.J.F.F.L.W. Haoran Xie: Sentiment strength detection with a context-dependent lexicon based convolutional neural network (2020)

[14] Agrawal, A., An, A.: Unsupervised emotion detection from text using semantic and syntactic relations. 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology **vol. 1**, 346–353 (2012)

[15] Bradley, M.M., Lang, P.J.: Affective norms for english words (anew): Instruction manual and affective ratings. Technical Report C-1, The Center for Research in Psychophysiology,University of Florida (1999)

[16] Kiritchenko, S., Mohammad, S.M., Salameh, M.: Semeval2016 task 7: Determining sentiment intensity of english and arabic phrases, in proc. of semeval, 42–51 (2016)

[17] S. Rosenthal, S.K.S.M.M.A.R. P. Nakov, Stoyanov, V.: Semeval 2015 task 10: Sentiment analysis in twitter, in proc. of semeval, 451–463 (2015)

[18] R. Socher, J.Y.W.C.C.D.M.A.N. A. Perelygin, Potts, C.: Recursive deep models for semantic compositionality over a sentiment treebank, in proc. of emnlp, 1631–1642 (2013)

[19] Isidoros Perikos, I.H.: Recognizing emotions in text using ensemble of classifiers. Eng. Appl. Artif. Intell **vol. 51**, 191–201 (2016)

[20] Kumar, D.S.B.R.: Keyword based emotion word ontology approach for detecting emotion class from

text (2016)

[21] Mashal Sonia Xylina, A.K.: Emotion intensity detection for social media data. 2017 International Conference on Computing Methodologies and Communication (ICCMC), 155–158 (2017)

[22] LeCompte Travis, C.J.: Sentiment analysis of tweets including emoji data. Sentiment Analysis of Tweets Including Emoji Data, 793–798 (Dec. 2017)

[23] Rabeya Tapasy, A.H.S.C.N.R. Ferdous Sanjida: A survey on emotion detection: A lexicon based backtracking approach for detecting emotion from bengali text. 2017 20th International Conference of Computer and Information Technology (ICCIT), 1–7 (2017)

[24] Kušen Ema, F.K.C.M.S.M. Cascavilla Giuseppe: Identifying emotions in social media: Comparison of word-emotion lexicons (Aug 2017)

[25] Badugu Srinivasu, S.M.: Emotion detection on twitter data using knowledge base approach. International Journal of Computer Applications **vol. 162**, 28–33 (March 2017)

[26] Abdullah Malak, S.S. Hadzikadicy Mirsad: Sedat: Sentiment and emotion detection in arabic text using cnn-lstm deep learning. 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 835–840 (2018)

[27] Tzacheva Angelina, M.S.Y. Ranganathan Jaishree: Actionable pattern discovery for tweet emotions, 46–57 (June 2019)

[28] Seal Dibyendu, B.R. Roy Uttam K.: Sentence-level emotion detection from text based on semantic rules. Information and Communication Technology for Sustainable Development, 423–430 (2020)

[29] Bharti Santosh Kumar, G.R.K.S.P.K.B.M.H.S.K.M.A. Varadhaganapathy S.: Text-based emotion recognition using deep learning approach. Computational Intelligence and Neuroscience (Aug 2022)