Harnessing Emotive Features for Emotion Recognition from Text

Rutal Mahajan[1], Mukesh Zaveri[2]

Department of Computer Engineering
Sardar Vallabhbhai National Institute of Technology, Surat, Gujarat, India

Abstract—With the prevalence of affective computing, emotion recognition becomes vital in any work related to natural language understanding. The inspiration for this work is provided by supplying machines with complete emotional intelligence and integrating them into routine life to satisfy complex human desires and needs. The text being a common communication medium on social media even now, it is important to analyze the emotions expressed in the text which is challenging due to the absence of audio-visual cues. Additionally, the conversational text conveys many emotions through communication contexts. Emoticon serves the purpose of self-annotation of writer's emotion in text. Therefore, a machine learning-based text emotion recognition model using emotive features proposed and evaluated it on the SemEval-2019 dataset. The proposed work involves exploitation of different emotion-based features with classical machine learning classifiers like SVM, Multilayer perceptron, REPTree, and decision tree classifiers. The proposed system performs competitively well in terms of f-score 65.31% and accuracy 87.55%.

Keywords—Emotion recognition; emotive features; natural language processing; affective computing

I. INTRODUCTION

A human newborn comes with primary settings for understanding and communicating basic feelings, as well as an immense ability to learn. A newborn baby can cry or stay calm, smell, and turn her head towards her mother. With the growing age, her neural network starts learning facial expression and gradually develops to read and express more emotions using different senses. The sense of recognizing and expressing emotion develops gradually from basic to complex linguistic if-then scenarios. To make computers understand and manage emotions as a human baby can do requires lots of work in that direction. How communication takes place is important in emotion recognition because it's not about what is being said, but about how it is being said. Expressions matter, as the sentiment behind each encounter and the emotions, are raised. Emotion is knotted with the literal meaning of words used. New research in artificial intelligence is giving machines, like software agents, computers, robotic pets, and any digital device, smart capabilities along with emotional intelligence Artificial intelligence is progressing towards emotional intelligence by implementing different tasks in real-time. Sentiment analysis is now considered the general task of natural language understanding. It is evolved as coarse-grained emotion recognition that is multiclass classification of sentiments.

Emotion recognition in the text is similar to many other problems in text classification and analysis. It is considered as a sub-task of sentiment analysis. Text is categorized into different 5 basic categories of emotions based on different emotion models that exist in psychology [1]. Currently, researchers have categorized the text into more than 20 complex emotion categories and applied to detect depression, joy, happiness index of the country, and many more application domains [2],[3]. Text Emotion recognizing has a variety of applications including identifying anxiety or depression of individuals and measuring the well-being or public mood of a community. Emotion recognition concerns extracting detail-level sentiment and associated emotion extraction from text. [4]. Sentiment analysis classifies the text into different polarities. Whereas, emotion classification can follow a detailed level of emotion categories belonging to particular polarity. Moreover, perceiving emotions from only text expressions is a difficult task due to the lack of audio and visual expressions. Thus gradually emoticons or emoji are evolved as substitutes for facial expressions during written commination over time. Emoticons are the pictorial simplified representation of facial expression conveying affective information in text communication. By using, emoticons, readers can understand the sender's emotional state without using many more words than simple text during textual communication. If the origin of any language is studied, they have evolved from ideographs, i.e. a graphical symbol representing an idea or concept regardless of specific language, words, and phrases[5]. Due to common communication features and evolving emoticon lexicons regardless of any language, emoticons become the universal language of communication [6],[7]. The deep learning models gain popularity with rise in the resources running these heavy models requiring large amount of training data [8], [9], [10], [11]. But these models are computationally costly and require resources affecting the environment [12], [13].

In this paper, to produce the human like prediction results for the emotion of the conversational text, simple machine learning based models is used with emoticon-based features. Emoticons are important entities conveying emotion expressed by a writer using tiny facial images. They evolved from a single smiley face to emoticons to images and complex emotion conveying stickers. Among all these, non-verbal emotion conveying parts of the text, images most widely use symbolic language during digital written communication [14]. Thus to capture the non-verbal emotional clues from the text, which are conveyed by the writer by use of emoji, sentiment information-related resources for emoji are utilized in our

work. Moreover, emotions conveyed in the conversational text require contextual information to identify it. Thus other lexical resources to capture emotions conveyed in the text are also used here.

The rest of this paper is as follows. Section 2 explores the notions related to emotion, emotion representation models, work done in the field of emotion recognition. The details about the proposed work on emotion recognition in conversational text messages, features, and experimental setups are given in Section 3. In Section 4, an analysis and discussion on the results of the proposed work are carried out. Finally, a conclusion with possible future directions of this work is described in Section 5 followed by references.

II. RELATED WORKS

Emotion recognition in the text is gaining interest in the research community aiming to extract detail level emotions from personal opinions, reviews, or any kind of feeling expressed in text over various media [4]. Due to various ways to express emotion in text such as using a direct expression, indirect expression, using graphical emoji and many ways, from text with polarity, it is evolved to an interesting field of research these days. Different factors affecting emotion recognition in the text are the domain and scope of data, the language used in data, choice of emotion model for deciding emotion labels in output, and classifiers, or recognition models for predicting the output.

Wherever human communication involves, emotions come into the picture. Emotions appear in varieties of speakers depending on the formal communication style or informal communication style. Sarcasm and humor are such types of figures of speech that convey a variety of emotions very complex manner [15]. It can be observed in the text communications over various mediums like Twitter, emails, Facebook or any websites providing platform to share personal views of people with numerous topics of discussions. It can be observed in the text communications over various mediums like Twitter, emails, Facebook or any website providing platform to share personal views of people with numerous topics of discussion. Emotion can occur in different sizes of text segments ranging from words to paragraphs and long documents, and even collections of documents. This parameter is crucial to decide the scope of the recognition model. Emotion recognition at the word level can be used to build the emotional vocabulary resources by using its connected emotional words. Document or paragraph level emotion recognition can be useful to give an emotional abstract view of the topic [16]. Emotion recognition from a collection of documents, paragraphs, and such longer scopes can give insights about the view of people on a certain topic, event, and product. For example, extracting and analyzing YouTube comments for emotion detection can give insights into the reaction of people on a particular topic, product, or user. Understanding emotions in absence of audio-visual cue is a hard problem. For example, when you read, "Why are you ignoring me", it is difficult to decide whether it is conveying an angry or sad emotion? The context of this conversation should be known to recognize emotion. In this work, emotion recognition is carried out on such textual conversation data.

Most of the work on emotion recognition is done in the English language due to the easy availability of resources required for linguistic processing. Chinese [17] and Japanese [18] are other frequently studied languages in this field. Numerous multilingual systems are also present in the literature for the sentiment analysis task but left understudied for a detailed level categorization task, that is, emotion recognition. In this work, we focused on the English language conversation texts for the categorization of emotion beyond positive, negative, and neutral tags.

To recognize emotions from text, first text need to be represented formally using different emotion models. For emotion recognition from text, researchers used both dimensional and categorical models of emotions. 2-dimensional Valence-Arousal space [19], and 3-dimensional Pleasure-Arousal-Dominance model are most commonly chosen dimensional model in this area of research. The most commonly used categorical models are Ekman's [20] emotion model with the 6 basic emotions and Plutchik's [1] emotion model with 8 primary emotions. In this work, we use the categorical model with simply four basic categories of emotions: happy, sad, angry, and others.

Different emotion recognition approaches in the literature are studied in this work, which can broadly be categorized into three categories:

A. Non Machine Learning Methods

Many methods make use of keywords in a sentence and use their co-occurrence with other keywords with explicit emotional value [21]–[23]. For finding the emotional values of words, different lexical resources are used. Popular lexical resources developed for English language are WordNet-Affect [24] and SentiWordNet [25]. These methods are heavily dependent on handcrafted rules created by human experts and resources used.

B. Non-neural Machine Learning Methods

Unlike other text categorization tasks, emotion recognition, most machine learning methods work by extracting features such as the presence of frequent words, negation, punctuation, emoticons, and so on to create feature representation of the sentence [15]. This representation is used as input by various classifiers to predict the output [26]–[28]. These methods often require a good knowledge of feature engineering for better prediction.

C. Deep Learning Methods

Neural network-based approaches with deeper network architecture have become a popular choice in varied tasks in text, speech, and image domains due to ease of not worrying about feature engineering. Variants of Recurrent Neural Networks, such as LSTM [29] has been effective in modelling sequential information. Also, Convolutional Neural Networks [30] have been a popular choice. These methods require large quantities of data to be worked on properly. The summary of various works reported in literature in the field of text emotion recognition with recent consideration is given in Table I. The majority of the work is done using resource costly deep learning models with a popular choice of deep learning model being variants of LSTM.

TABLE I.   SUMMARY OF RECENT EMOTION RECOGNITION WORK IN TEXT

Ref.	No. of Emotions	Approach	Language	Dataset	Performance	Remarks
[15]	2 (Humor, Not Humor)	Machine Learning based (Support Vector Machine, Naïve Bayes, Multilayer Perceptron)	English	Yelp Review Data (6000)	Precision:60.2% Recall:90.1% Fscore:72.2% Accuracy: Not given	Humor or Not Humor is predicted based on affect-based and user's writing style specific features (2020)
[8]	4 (Happy, Sad, Angry, Others)	Deep Learning (Transfer learning based model)	English	SemEval 2019 Task 3 (Train: 30160 Test: 6032)	Precision: Recall: Fscore: Accuracy: Not given	LSTM and BERT based emsemble model (2020)
[9]	4 (Happy, Sad, Angry, Others)	Deep Learning (Simplest Long short term memory network with fully connected neural network at output layer)	English	SemEval 2019 Task 3 (Train: 30160 Test: 6032)	Precision:50.21% Recall:71.04% Fscore:58.51% Accuracy: Not given	100 dimensional GloVe embedding, 128 dimensional representation of sentence,4 dimensional output representation, Baseline system of SemEval 2019 Task-3 (2019)
[10]	4 (Happy, Sad, Angry, Others)	Deep Learning (Fine-tuned BERT with hierarchical Long Short Term Memory Network)	English	SemEval 2019 Task 3 (Train: 30160 Test: 6032)	Precision:56% Recall:81.66% Fscore:77% Accuracy: Not given	Semantic and emotional content of text encoded via GloVe, ELMo and DeepMoji (2019)
[11]	4 (Happy, Sad, Angry, Others)	Deep Learning (Sentiment and Semnatic LSTM)	English	Text Conversation Dataset (Train: 30160 Test: 6032)	Precision:80.87% Recall:64.08% Fscore:79.34% Accuracy: Not given	LSTM based approach proposed to predict Happy , Sad or angry in text conversations (2017)
[31]	2 (positive, Negative)	Support Vector Machine, Naive Bayes. Semantic relatedness between aspect word and student's opinion sentence is calculated.	English	Tweets on Educational aspects (1728)	Precision:96.85% Recall:84.6% Fscore:89.95% Accuracy: Not given	Aspect based sentiment analysis. SentiWordnet used to categorize words into different polarity. (2017)
[32]	8 (joy, hate, love, expect, surprise, anxity, sorrow, anger)	Hierarchical Bayesian inference methods	Chinese	Blog articles (train:917, test: 230)	Precision:84.00% (documentl level) 46.68% (word-level) Recall: 83.24% (document-level) 36.06(word-level) Fscore: 53.06% (document – level) 45.11(word-level) Accuracy: Not given	Contextual information used, predicts emotion for word and document(2017)
[33]	2 (Happiness, Sadness) or (Positive, Negative)	Lexicon based backtracking approach for sentiment analysis	Bengali	301 Bengali Text sentences	Precision: Not given Recall: Not given Fscore: Not given Accuracy: 77.16%	Used lexical resources , experiment done on very limited data (2017)
[34]	4 (Happy, Sad, Angry, Others)	Machine Learning based method (Light-GBM tree)	English	SemEval 2019 Task 3 (Train: 30160 Test: 6032)	Precision: 76.8% Recall: 78.76% Fscore: 77.65% Accuracy: Not given	combination of lexical features such as word and character grams, along with additional signals like emotional intensity, valence-arousal-dominance scores (2019)

III. PROPOSED EMOTION RECOGNITION MODEL

In natural language understanding, emotion recognition is a hierarchical task carried out based on the indirect expression of feelings and emotions. Text communications are very common these days due to the prevalence of various digital communication media, and they may contain figurative language. So the novelty of this work is to implement the emotion recognition model for textual conversation data using classical machine learning algorithms along with minimalistic features. The set of various emotion features based on emoticon (emoji) sentiments and lexical emotion resources are proposed in this work for text emotion recognition.

Emotion recognition in text conversation is implemented as a classification task. Let $U = \{u_1, u_2\}$ be the users involved in 3-turn text conversation. $C = \{(c_1u_1, c_1u_2, c_1u_1), (c_2u_1, c_2u_2, c_2u_1) \dots (c_nu_1, c_nu_2, c_nu_1)\}$ be the set of n 3-turn conversations between user u_1 and u_2. In this input conversation text, 3 conversation texts are provided. First is by user 1, second is its reply from user 2 and last is again the turn of user 1. $E=\{e_1,e_2,e_3,\dots e_n\}$ be the corresponding set of emotion labels given to each conversation triplets where $e \in$ {happy, sad, angry, other} indicates the corresponding output emotion conveyed in 3-turn conversation. The objective of the proposed model is to predict conditional label distribution $P(e|c)$ from the text conversation dataset of SemEval-2019 [11] to assess the competitiveness of the proposed work globally. The proposed emotion recognition model works as shown in Fig. 1. The file containing 3-turn conversation text and emotion labels are given as input to the feature extractor where 20 different features related to emotions and emoticons are extracted. The classifier is trained using extracted features to classify the text conversation triplets into any of four emotion labels, namely, happy, sad, angry, and others. The detailed process is explained below:

A. Phase 1: Pre-processing

Before using the dataset on the proposed emotion recognition model, it is pre-processed to resolve inconsistencies.

1) Emoticons/emoji from text conversations are replaced with their corresponding description of conveyed emotion. The words in the emoticon description are used to extract the sentiment and emotion-related information using lexical resources. For replacing emoticon sentiments and related description details, the sentiment of emoji is used as a lexical resource [35].

2) Tokenization: short text conversation in each turn is divided into distinct words, punctuations across the white spaces. Emotion description is considered separately for processing by lexical resources as per emotive word presence in it.

B. Phase 2: Feature Extraction

To identify different emotions from the given text fragment, suitable features are extracted from the given text. These features act as determinants of different output emotion classes. Phase 1 replaces emoticons appearing in text with their short descriptions. Emotions in the text can be conveyed majorly by emoticons, if present, and by various linguistic units like specific emotive words and contexts used by writers. The Emotion-based and emoticon sentiment-based features are used as an emotion feature vector in the emotion recognition model. These features use various lexical resources like EmoLex [36], EmoSenticNet [37], and Emoji Sentiment Ranking [35]. 10 features are extracted using 2 EmoLex sentiments (positive and negative) and 8 different emotion labels, namely, Surprise, joy, anger, trust, anticipation, fear, disgust. Sadness. Similarly, from EmoSenticnet lexicon 6 features using different emotion labels, namely, anger, disgust, fear, joy, sadness, and surprise are extracted. Another emotive resource used in this work is related to the sentiment of emoticons, which is related to calculating the negative, positive, neutral, and overall sentiment score of the emoticon.
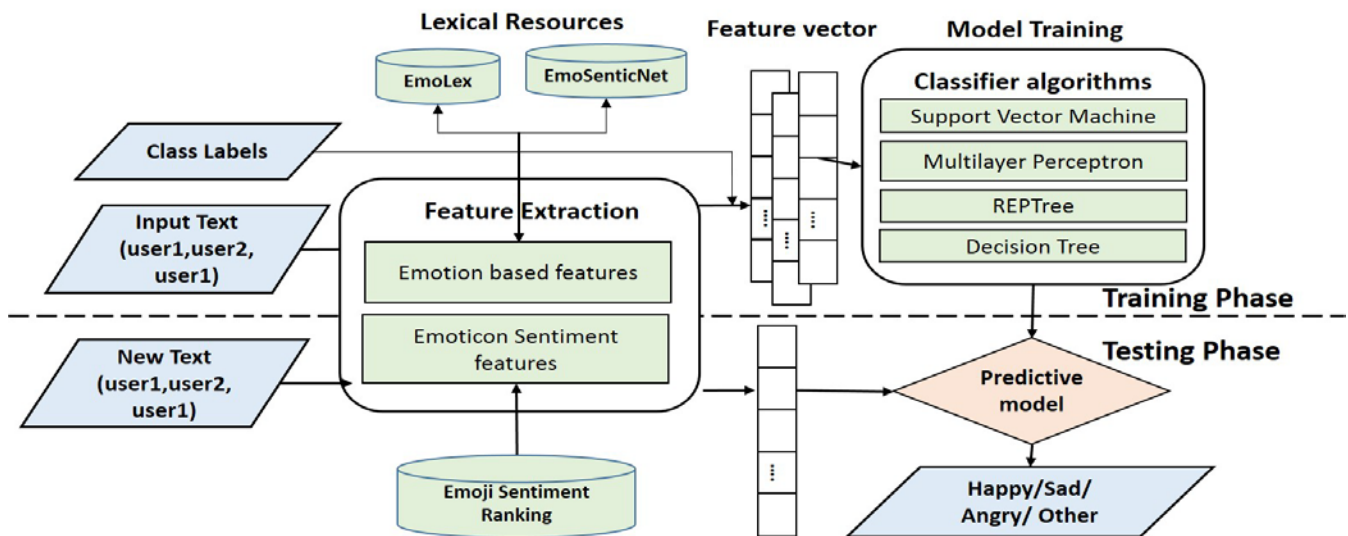

Fig. 1. Proposed Emotion Recognition Model using Emotion and Emoticon Sentiment-based Features.

C. Case 3: Classification

For classification of text conversations according to emotion, it exhibits, four different classifiers are tested on the dataset, namely support vector machine, multilayer perceptron, REP (reduced error pruning) Tree, and decision tree classifiers. These classifiers are chosen based on their performances in the literature on tasks related to emotion recognition [38], [39],[2]. We used the implementation of these algorithms as given in the experimental environment [40]. Support Vector Machine: SVM (Support Vector Machine) is a supervised algorithm working well for both classification and regression. We implemented sequential minimal optimization-based SVM here. Multilayer Perceptron is a simple yet effective neural classifier algorithm. REPTree is a fast decision tree algorithm for classification.

D. Dataset

The difficulty of emotion recognition increases when we need to recognize the emotion conveyed in the conversational text. Because as humans, on reading the text "You missed our 5th anniversary!!" It can be either interpreted as a sad or angry emotion and the same ambiguity exists for machines. So the performance of this work is evaluated on a 3-turn text conversation dataset released by SemEval 2019 [11]. One sample of the dataset contains conversations from Twitter in three turns, i.e. User 1's tweet; User 2's response to the tweet, user 1's response to User 2's response [41]. A total of 30160 samples are used in the training of the classifier, each sample labeled with Happy or Angry or sad or other. 2755 samples are used for validation and 6032 samples are used for testing of the model [11]. The details of emotion class label distribution for samples in the dataset are as given in Table II. As the distribution of data in various class labels in the training set impacts the classifier performance, Fig. 2 represents the data distribution in various emotion classes in the training set. 50% of samples are having other labels and 50% of samples are having emotion class labels from happy, sad, and angry. These emotion labels are distributed as 18% of samples in the dataset are having angry, 18% samples are having sad and 14% samples are having happy emotion labels.

The challenge of working with this dataset is the size of data in each conversation sample, which is relatively small, approximately 4 words in each user conversation. During pre-processing and feature extraction it becomes of utmost importance to ensure minimal loss of useful information with noise removal from each conversation samples.

TABLE II. DETAILS OF SEMEVAL 2019 TEXT CONVERSATION DATASET

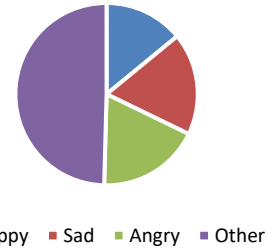|	Train	Validation	Test
Happy	4243	142	803
Sad	5463	125	1114
Angry	5506	150	1121
Others	14948	2338	2994
Total Samples	30160	2755	6032

Fig. 2. Emotion Class Label Distribution in the Training Dataset.

IV. EXPERIMENTAL SETUP

A. Experiment Configuration

The experimental environment used for training of the proposed emotion recognition model is a CPU-based system having Intel 64-bit core i5 2.5 GHz processor with Windows 7 operating system and 4 GB RAM. For evaluation, we used different metrics as described in Section 4.2. In the experiment, we have evaluated our hypothesis for the effectiveness of emotion-based and emoticon-based features which we defined to explore its power for recognizing emotions in text.

Hypothesis: Does the proposed emotion recognition model benefit from emotion and emoticon-based features?

Experiment: To check the effectiveness of this hypothesis ,the emotion recognition model with implementations of simple state-of-the art classifiers from Weka such as support vector machine, different variants of multilayer perceptron, decision tree classifiers, and REP Tree classifier is trained with three variations: 1) individual set of emotion-based features, 2) emoticon-based features, and 3) emotion-based and emoticon-based features altogether. Comparative evaluation is done on the results of all classifiers with these three cases. We have not taken results on only emoticon-based features as most of the time any text in the dataset contains approximately 3 to 4 words and optional presence of emoticons.

B. Evaluation Measures

Considering emotion recognition as a task of classification, different measures used to evaluate the performance of classification model are precision, recall, f-score, accuracy, and Mathew's Correlation Coefficient for evaluating the performance of most natural language understanding systems.

Precision is calculated by Eq. (1) as follows. The higher values of precision indicate not only more occurrences of correctly classified values but mainly fewer occurrences of False Positive (FP) values.

$$precision = \frac{TP}{TP + FP} \tag{1}$$

Recall is calculated by Eq. (2) as follows. The higher values of recall indicate not only more occurrences of correctly classified values but mainly fewer occurrences of False Negative (FN) values.

$$recall = \frac{TP}{TP + FN} \tag{2}$$

Precision and recall cannot be used independently to evaluate the performance of any classifier. System recall can be easily increased by labelling class values at the cost of precision and vice versa. Thus F-score is used which considers both the precision and recall to balance. It can be given by Eq. (3) as follows.

$$f-score = 2 * \frac{precision * recall}{precision + recall}$$ (3)

Accuracy is the proximity of measurement results to the true value. It can be given by Eq. (4) as follows.

$$Accuracy = \frac{TP + TN}{P + N}$$ (4)

Matthew's correlation coefficient is not much popular evaluation measure in classification tasks but it is promising than f-score and accuracy while evaluating qualitatively the performance of the classifier. It can be given by Eq. (5) as follows.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TN + FN)(FP + TP)(TN + FP)(FN + TP)}}$$ (5)

F-score ignores the count of true negatives (TN), whereas MCC considers all the entries of the confusion matrix for evaluating the performance of the classification model. This measure does well only when the classifier is doing well on both negative and positive elements. So F-score gets affected more when the minority class is labeled as negative. When the majority class is labeled as negative f-score can be considered as a good measure because in such cases rare items are of interesting samples for classification.

V. RESULT AND DISCUSSION

Higher average accuracy is considered as a good score, where the task is to predict the label of the emotion considering all the emotion class labels have an equal number of the sample distribution. Here as mentioned in the dataset description given in Table I, happy, sad, and angry emotion classes have almost equal proportion of samples but another 50% of samples belong to other emotion categories. So we evaluate the performance of our proposed model using f-score values for each classifier. Table III shows a summary of results with different classifiers on a different set of features.

From the results of the experiment, it is evident that the accuracy and f-score of REP Tree are higher in the cases of individual 16 emotion-based features and with all 20 emotion and emoticon-based features. Tree-based classifiers have performed better than other classifiers in the emotion recognition task. Moreover, from Fig. 2 and Fig. 3, the effect of the use of emotion-based features on the performance of classifiers can be observed. It is marked that after including emoticon-based features in the feature set for classification of emotion, a significant improvement is observed in the f-score as well as in accuracy.

In Fig. 3, f-score (20) represents all 20 features used for classification, and f-score (16) represents all emotion-based features used for classification. Similarly in Fig. 4, accuracy

(20) represents all 20 features used for classification, and accuracy (16) represents all emoticon-based features used for classification. With the inclusion of 4 emoticon-based features in classification, SVM performance is improved by 1.91% accuracy and 0.89% f-score. In REPTree, it is improved by 2.08% accuracy and 0.7% f-score. With the inclusion of emoticon based features proposed emotion recognition model achieves the highest accuracy of 87.55% and f-score of 65.31% with the use of the REPTree classifier.

TABLE III. PERFORMANCE OF PROPOSED EMOTION RECOGNITION MODEL WITH VARIOUS FEATURE SETS ON SEMEVAL 2019 DATASET

|	Classifier	SemEval 2019 Text Conversation Dataset			
		Precision	*Recall*	*F-Score*	*Accuracy*
Emotion Features (16)	SVM	56.15%	38.06%	45.37%	83.07%
	MLP	61.35%	48.74%	54.32%	84.86%
	MLP-3	51.34%	59.02%	54.92%	84.43%
	REPTree	62.29%	54.12%	57.92%	85.47%
	Decision Tree	65.46%	43.89%	52.55%	85.36%
Emotion (16) + Emoticon Sentiment Features(4)	SVM	61.95%	48.38%	54.33%	84.98%
	MLP	67.30%	57.09%	61.77%	86.95%
	MLP-3	61.09%	61.04%	61.06%	85.62%
	REPTree	**67.26%**	**63.46%**	**65.31%**	**87.55%**
	Decision Tree	69.52%	47.30%	56.30%	86.43%
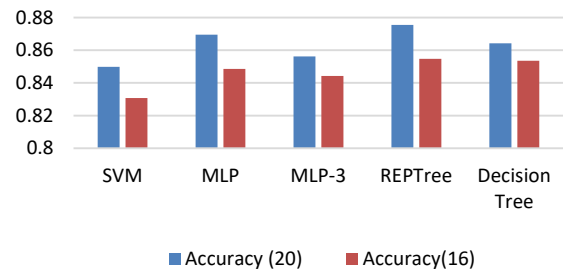
Fig. 3. Comparison of the Effectiveness of Features of Classifier Performance in Terms of Average f-score of all Emotion Classes.

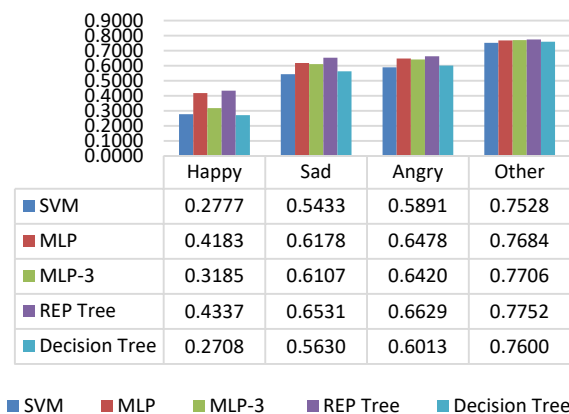|	Happy	Sad	Angry	Other
SVM	0.2777	0.5433	0.5891	0.7528
MLP	0.4183	0.6178	0.6478	0.7684
MLP-3	0.3185	0.6107	0.6420	0.7706
REP Tree	0.4337	0.6531	0.6629	0.7752
Decision Tree	0.2708	0.5630	0.6013	0.7600

Fig. 4. Comparison of the Effectiveness of Features on Classifier Performance in Term of Average Accuracy of all Emotion Classes.

The performance comparison of our model with different machine learning classifiers with other similar systems on SemEval 2019 text conversation dataset is summarized in Table IV. This global evaluation was carried out in the SemEval 2019 in task 3. Majority participant systems has implemented their emotion recognition module with LSTM / BiLSTMs with well-known word embedding like GloVe is used for input representation and BERT is used for transfer learning by most systems using deep learning. We proposed machine learning based solution to give results close to human performance on text based emotion prediction, which is less resource expensive in comparison with deep learning approaches. The best performing classifier in our model repots the f-score higher than the mean f-score of all participating teams in this evaluation, which are implemented using deep learning; the top few among them are represented here in summary Table IV. Another interesting fact to notice from the analysis is that our model with all chosen classifiers except SVM performed well in terms of higher f-score than the baseline provided for the global evaluation. The baseline system was also implemented using deep learning model with 100 dimensional GloVe embedding. Our model with 20 different emotion-based features performs well.

Class label distribution has a great effect on classifier performance, which can be observed very clearly in the performance of each classifier. Fig. 5 describes the emotion-wise performance of each classifier in terms of an f-score. The happy emotion class has not performed well with any of the classifiers used in the proposed model. The highest f-score value achieved on the happy emotion class is 43.47% by using the REPTree classifier. The evident reason found for the bad performance of all classifiers on the happy emotion class is that the happy class has the least sample distribution in the dataset. Another reason we observed that the statements in conversations conveying the happy emotion need context to understand them. Detailed analysis of each emotion class is carried out with best performing classifier REPTree in terms of correctly classified and misclassified samples in each emotion class. The confusion matrix is prepared as shown in Fig. 6 for each emotion class, namely, happy, sad, angry, and others. For each emotion class, Matthew's correlation coefficient (MCC) is calculated, which is useful to evaluate the classifier in terms of quality. This coefficient takes into account true positives and true negatives as well as false positives and false negatives. It is considered as a balanced measure that can be used even if the classes are of very different sizes and hence suitable for evaluating the performance of the proposed model. The REPTree classifier in the proposed model has achieved the highest average accuracy of 85.47% and 87.55% by considering all 16 emotion features and 20 emotion and emoticon-based features altogether respectively. This information does not exhibit the weaknesses of the classification process where the performance can be further

improved. But from Fig. 6(a)-(e), it is evident that MCC gives information regarding the quality of classification for each emotion class. MCC of 0.4161 for the happy class depicts the number of misclassified samples are more in it, compared to other emotion classes. Total correctly classified samples using the REPTree classifier are 4225 samples (70.043%) and 1807 (29.95%) are incorrectly classified from a total of 6032 samples. From Fig. 6(a), it is clear that emotion class anger and others have major contributions is this average correctly classified samples by this classifier. In the case of the SVM classifier, the happy class has more misclassified samples. The misclassified happy class samples reversely affect the performance of classifiers which has comparatively performed better at classifying samples at other emotion classes. Among the total 6032 samples, SVM has classified 3929 (65.14%) samples correctly and 2103 (34.86%) samples incorrectly. In this total correct prediction of all emotion classes, the major share of correct samples is from emotion class other and angry.

TABLE IV. PERFORMANCE COMPARISON OF VARIOUS DEEP LEARNING AND MACHINE LEARNING MODELS ON SEMEVAL 2019 DATASET

	Precision	Recall	F-Score	Accuracy
Proposed Model (REPTRee)	67.26%	63.46%	65.31%	87.55%
Proposed Model (SVM)	61.95%	48.38%	54.33%	84.98%
Proposed Model (MLP)	67.30%	57.09%	61.77%	86.95%
Proposed Model (MLP-3)	61.09%	61.04%	61.06%	85.62%
Proposed Model (Decision Tree)	69.52%	47.30%	56.30%	86.43%
Baseline*	50.21%	71.04%	58.51%	-
SS-LSTM [11]	80.87%	64.08%	79.34%	-
BERT Based[8]	86.2%	86.4%	86.3%	-
Light-GBM Tree[34]	76.8%	78.76%	77.65%	-
BERT based Hierarchical LSTM[10]	56%	81.66%	77%	-

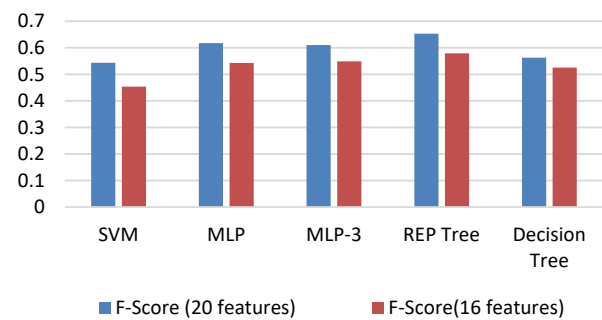*It is Deep Learning based baseline –LSTM with 100-D GloVe embeddings

Fig. 5. Individual Emotion Class Wise Classifier Performance Comparison based on f-score Values.

SVM — Predicted happy

Actual happy		
TP 3	FN 800	
FP 20	TN 5209	

MCC = -0.00048964

SVM — Predicted sad

Actual sad		
TP 424	FN 690	
FP 331	TN 4587	

MCC = 0.3674

SVM — Predicted angry

Actual angry		
TP 607	FN 514	
FP 324	TN 4587	

MCC = 0.5119

SVM — Predicted other

Actual other		
TP 2679	FN 315	
FP 1644	TN 1394	

MCC = 0.3924

(a) Confusion Matrix for Individual Emotion-class (Happy, Sad, Angry, other) with SVM Classifier.

REPTree — Predicted happy

Actual happy		
TP 255	FN 548	
FP 118	TN 5111	

MCC = -0.4161

REPTree — Predicted sad

Actual sad		
TP 707	FN 407	
FP 344	TN 4574	

MCC = 0.5777

REPTree — Predicted angry

Actual angry		
TP 697	FN 424	
FP 285	TN 4626	

MCC = 0.594

REPTree — Predicted other

Actual other		
TP 2566	FN 428	
FP 1060	TN 1978	

MCC = 0.5188

(d) Confusion Matrix for Individual Emotion-class (Happy, Sad, Angry, other) with REP Tree Classifier.

MLP — Predicted happy

Actual happy		
TP 128	FN 675	
FP 138	TN 5091	

MCC = -0.2200

MLP — Predicted sad

Actual sad		
TP 543	FN 571	
FP 342	TN 4576	

MCC = 0.4583

MLP — Predicted angry

Actual angry		
TP 632	FN 489	
FP 251	TN 4660	

MCC = 0.5641

MLP — Predicted other

Actual other		
TP 2635	FN 359	
FP 1363	TN 1675	

MCC = 0.4562

(b) Confusion Matrix for Individual Emotion-Class (Happy, Sad, Angry, other) with MLP Classifier.

Decision Tree — Predicted happy

Actual happy		
TP 84	FN 719	
FP 158	TN 5071	

MCC = -0.1287

Decision Tree — Predicted sad

Actual sad		
TP 489	FN 625	
FP 258	TN 4660	

MCC = 0.4553

Decision Tree — Predicted angry

Actual angry		
TP 561	FN 560	
FP 271	TN 4640	

MCC = 0.5022

Decision Tree — Predicted other

Actual other		
TP 2656	FN 338	
FP 1555	TN 1483	

MCC = 0.4087

Fig. 6. (e) Confusion Matrix for Individual Emotion-class (Happy, Sad, Angry, other) with Decision Tree Classifier.

MLP-3 — Predicted happy

Actual happy		
TP 67	FN 736	
FP 56	TN 5173	

MCC = -0.1748

MLP-3 — Predicted sad

Actual sad		
TP 572	FN 542	
FP 397	TN 4521	

MCC = 0.4572

MLP-3 — Predicted angry

Actual angry		
TP 601	FN 520	
FP 254	TN 4657	

MCC = 0.5402

MLP-3 — Predicted other

Actual other		
TP 2650	FN 344	
FP 1435	TN 1603	

MCC = 0.4413

(c) Confusion Matrix for Individual Emotion-class (Happy, Sad, Angry, other) with MLP-3 Classifier.

VI. CONCLUSION

In this work, we addressed the issue of recognizing emotions conveyed in text conversations using traditional machine learning approaches. This had explored the directions of research for the proposed emotion recognition model for text conversations. The emotion recognition model based on a set of emotion and emoticon sentiment related features are proposed and implemented, which has achieved competitive results with modern resource costly deep learning approaches used in the literature. On evaluating our contributions in this work by employing emoticon-based features with simplistic state-of-the-art machine learning classifiers competitive performance is observed. This finding may help apply emotion recognition with low cost resources with human like prediction results. Some cases are left by the classifier to identify correctly, which can be incorporated by exploiting rules in the future. The possible direction of research are identified based on this work are as 1) exploring and experimenting with hybrid machine learning based and rule-

based approach to recognize the emotions from textual conversations. 2) Applicability of emotion recognition model to further understand the figurative language components in textual conversation. 3) Findings from the results of the multilayer perceptron say that there might be a significant improvement in performance if the sample size in the emotion class increases while using neural classifiers.

REFERENCES

[1] R. Plutchik, "integration, differentiation, and derivatives of emotion," Evolution and cognition, vol. 7, no. 2, pp. 114–125, 2001.

[2] N. Alswaidan and M. E. B. Menai, "A survey of state-of-the-art approaches for emotion recognition in text," Knowledge and Information Systems, 2020, doi: 10.1007/s10115-020-01449-0.

[3] S. Baccianella, A. Esuli, and F. Sebastiani, "SENTIWORDNET 3.0: An enhanced lexical resource for sentiment analysis and opinion mining," 2010.

[4] B. Liu, "Sentiment analysis and opinion mining," Synthesis lectures on human language technologies, vol. 5, no. 1, pp. 1–167, 2012.

[5] G. Santamaría-Bonfil and O. Grabiel Toledano López, "Emoji as a Proxy of Emotional Communication," in Becoming Human with Humanoid - From Physical Interaction to Social Intelligence, 2020.

[6] F. Barbieri, G. Kruszewski, F. Ronzano, and H. Saggion, "How cosmopolitan are Emojis? Exploring Emojis usage and meaning over different languages with distributional semantics," in MM 2016 - Proceedings of the 2016 ACM Multimedia Conference, 2016, pp. 531–535, doi: 10.1145/2964284.2967278.

[7] B. Eisner, T. Rocktäschel, I. Augenstein, M. Bosnjak, and S. Riedel, "emoji2vec: Learning Emoji Representations from their Description," in Proceedings of The Fourth International Workshop on Natural Language Processing for Social Media, 2016, pp. 48–54, doi: 10.18653/v1/w16-6208.

[8] Z. Ahmad, R. Jindal, A. Ekbal, and P. Bhattachharyya, "Borrow from rich cousin: transfer learning for emotion detection using cross lingual embedding," Expert Systems with Applications, vol. 139, 2020, doi: 10.1016/j.eswa.2019.112851.

[9] A. Chatterjee, K. N. Narahari, M. Joshi, and P. Agrawal, "SemEval-2019 Task 3: EmoContext Contextual Emotion Detection in Text," 2019.

[10] A. Basile, M. Franco-Salvador, N. Pawar, S. Štajner, M. C. Rios, and Y. Benajiba, "SymantoResearch at SemEval-2019 Task 3: Combined Neural Models for Emotion Classification in Human-Chatbot Conversations," in Proceedings of the 13th International Workshop on Semantic Evaluation, 2019, pp. 330–334.

[11] U. Gupta, A. Chatterjee, R. Srikanth, and P. Agrawal, "A sentiment-and-semantics-based approach for emotion detection in textual conversations," arXiv preprint arXiv:1707.06996, 2017.

[12] L. Amarpuri, N. Yadav, G. Kumar, and S. Agrawal, "Prediction of CO2 emissions using deep learning hybrid approach: A Case Study in Indian Context," 2019, doi: 10.1109/IC3.2019.8844902.

[13] L. Lu and B. Lyu, "Reducing energy consumption of Neural Architecture Search: An inference latency prediction framework," Sustainable Cities and Society, vol. 67, 2021, doi: 10.1016/j.scs.2021.102747.

[14] Q. Bai, Q. Dan, Z. Mu, and M. Yang, "A Systematic Review of Emoji: Current Research and Future Perspectives," Frontiers in Psychology. 2019, doi: 10.3389/fpsyg.2019.02221.

[15] R. Mahajan and M. Zaveri, "Humor identification using affect based content in target text," Journal of Intelligent and Fuzzy Systems, vol. 39, no. 1, pp. 697–708, 2020, doi: 10.3233/JIFS-191648.

[16] V. Sintsova, "Advancing Fine-Grained Emotion Recognition in Short Text," 2016.

[17] C. Quan and F. Ren, "A blog emotion corpus for emotional expression analysis in Chinese," Computer Speech & Language, vol. 24, no. 4, pp. 726–749, 2010.

[18] K. Matsumoto, Y. Konishi, H. Sayama, and F. Ren, "Analysis of Wakamono Kotoba emotion corpus and its application in emotion estimation," International Journal of Advanced Intelligence, vol. 3, no. 1, pp. 1–24, 2011.

[19] J. A. Russell and A. Mehrabian, "Evidence for a three-factor theory of emotions," Journal of research in Personality, vol. 11, no. 3, pp. 273–294, 1977.

[20] P. Ekman, "An argument for basic emotions," Cognition and Emotion, vol. 6, no. 3–4, pp. 169–200, 1992, doi: 10.1080/02699939208411068.

[21] A. Balahur, J. M. Hermida, and A. Montoyo, "Detecting implicit expressions of sentiment in text based on commonsense knowledge," in Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, 2011, pp. 53–60.

[22] F.-R. Chaumartin, "UPAR7: A knowledge-based system for headline sentiment tagging," in Proceedings of the 4th International Workshop on Semantic Evaluations, 2007, pp. 422–425.

[23] Z. Kozareva, B. Navarro, S. Vázquez, and A. Montoyo, "UA-ZBSA: a headline emotion classification through web information," in Proceedings of the 4th international workshop on semantic evaluations, 2007, pp. 334–337.

[24] C. Strapparava, A. Valitutti, and others, "Wordnet affect: an affective extension of wordnet.," in Lrec, 2004, vol. 4, pp. 1083–1086.

[25] S. Baccianella, A. Esuli, and F. Sebastiani, "Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining.," in LREC, 2010, vol. 10, no. 2010, pp. 2200–2204.

[26] R. C. Balabantaray, M. Mohammad, and N. Sharma, "Multi-class twitter emotion classification: A new approach," International Journal of Applied Information Systems, vol. 4, no. 1, pp. 48–53, 2012.

[27] D. Davidov, O. Tsur, and A. Rappoport, "Enhanced sentiment learning using twitter hashtags and smileys," in Proceedings of the 23rd international conference on computational linguistics: posters, 2010, pp. 241–249.

[28] C. O. Alm, D. Roth, and R. Sproat, "Emotions from text: machine learning for text-based emotion prediction," in Proceedings of the conference on human language technology and empirical methods in natural language processing, 2005, pp. 579–586.

[29] C. Zhou, C. Sun, Z. Liu, and F. Lau, "A C-LSTM neural network for text classification," arXiv preprint arXiv:1511.08630, 2015.

[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances In Neural Information Processing Systems, 2012, pp. 1–9, doi: http://dx.doi.org/10.1016/j.protcy.2014.09.007.

[31] M. Sivakumar and U. S. Reddy, "Aspect based sentiment analysis of students opinion using machine learning techniques," 2018, doi: 10.1109/ICICI.2017.8365231.

[32] X. Kang, F. Ren, and Y. Wu, "Exploring latent semantic information for textual emotion recognition in blog articles," IEEE/CAA Journal of Automatica Sinica, 2018, doi: 10.1109/JAS.2017.7510421.

[33] T. Rabeya, S. Ferdous, H. S. Ali, and N. R. Chakraborty, "A survey on emotion detection: A lexicon based backtracking approach for detecting emotion from Bengali text," in 20th International Conference of Computer and Information Technology, ICCIT 2017, Feb. 2018, vol. 2018-January, pp. 1–7, doi: 10.1109/ICCITECHN.2017.8281855.

[34] P. Agrawal and A. Suri, "NELEC at SemEval-2019 Task 3: Think Twice Before Going Deep," 2019.

[35] P. Kralj Novak, J. Smailović, B. Sluban, and I. Mozetič, "Sentiment of emojis," PLoS ONE, vol. 10, no. 12, p. e0144296, 2015, [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0144296.

[36] S. M. Mohammad and P. D. Turney, "Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon," in Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text, 2010, pp. 26–34.

[37] S. Poria, A. Gelbukh, A. Hussain, N. Howard, D. Das, and S. Bandyopadhyay, "Enhanced SenticNet with affective labels for concept-based opinion mining," IEEE Intelligent Systems, vol. 28, no. 2, pp. 31–38, 2013.

[38] R. Mahajan and M. Zaveri, "SVNIT $@$ SemEval 2017 Task-6: Learning a Sense of Humor Using Supervised Approach," in Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), Aug. 2017, pp. 412–416, [Online]. Available: http://www.aclweb.org/anthology/S17-2069.

[39] A. P. Gopi, R. N. S. Jyothi, V. L. Narayana, and K. S. Sandeep, "Classification of tweets data based on polarity using improved RBF kernel of SVM," International Journal of Information Technology, 2020, doi: 10.1007/s41870-019-00409-4.

[40] M. Hall et al., "The WEKA Data Mining Software : An Update," ACM SIGKDD explorations newsletter, 2009.

[41] A. Chatterjee, U. Gupta, M. K. Chinnakotla, R. Srikanth, M. Galley, and P. Agrawal, "Understanding Emotions in Text Using Deep Learning and Big Data," Computers in Human Behavior, vol. 93, pp. 309–317, Apr. 2019, doi: 10.1016/j.chb.2018.12.029.