

Emotions in texts

Leeveshkumar Pokhun ¹, M Yasser Chuttur ^{2,*}

Department of Software and Information Systems, University of Mauritius, Moka, Mauritius

¹ leevesh.pokhun1@uom.ac.mu; ² y.chuttur@uom.ac.mu*

* corresponding author

ARTICLE INFO

Article history

Received July 21, 2020

Revised August 2, 2020

Accepted August 30, 2020

Keywords

Emotions Analysis

Emotion Models

Emotion Dataset

ABSTRACT

Several studies have used different techniques to detect and identify emotions expressed in various sets of texts corpora. In this paper, we review different emotion models, emotion datasets and the corresponding techniques used for emotion analysis in past studies. We observe that researchers have been using a wide variety of techniques to detect emotions in texts and that there is currently no gold standard on which dataset or which emotion model to use. Consequently, although the field of emotion analysis has gained much momentum in previous years, there seems to be little progress into relevant research with findings that may be useful in real world applications. From our analysis and findings, we urge researchers to consider the development of datasets, evaluation benchmarks and a common platform for sharing achievements in emotion analysis to see further development in the field.

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



1. Introduction

User-generated content on social media networks and the web is a common practice nowadays [1]. Users, who were once consumers of information, are now major information contributors [2]. This shift from information consumers to information producers has provided an abundance of data, both structured and unstructured, that can be collected, processed, and analysed to yield useful information. The observed increase in online user content is explained primarily by the latest technology advancement and wide adoption of Web 2.0, whereby expressing one's emotions or feelings over communication networks has become much simpler and easier [3]. For instance, using short text messages such as LOL for expressing '*laugh all loud*' or '*lots of love*', people are no longer bound to typing full texts when communicating. Also, with the introduction of emoticons and multimodal content (text with image, video, audio, animated images), users can quickly and explicitly convey their feelings without having to type long descriptive messages [4], [5]. Currently, most, if not all, online messages are written in short text format and contain some form of emoticons. Consequently, classical methods of text analysis, such as sentiment analysis, are not sufficient in obtaining a correct interpretation of the meaning conveyed in each message.

In general, sentiment analysis, also known as opinion mining, focuses on classifying text into three main categories namely *positive*, *negative*, and *neutral*. Those three categories are useful in various applications [6] but will, however, not provide accurate details regarding any emotion expressed in a message. For instance, sentiment analysis cannot perceive emotions such as *joy*, *fear*, *surprise*, *guilt* etc., which are essential to understanding the actual feeling expressed by a user [7]. For a more effective interpretation of a message, it is not only essential to identify the contents and context of the message, but it is also crucial to detect the emotion expressed within the message [8]. Emotion analysis, in this sense, seeks to capture the emotion expressed in a message [9]. Emotions by nature, however, are very complex to identify, and several techniques have been proposed for emotion analysis.

In this paper, we present the different achievements and challenges in emotion analysis. The rest of the paper is organized as follows: Section 2 introduces the concept of emotion along with considerations to take when detecting emotion. Different applications of emotion analysis are presented in section 3. Section 4 describes some popular emotion models used to classify emotions, followed by section 5, which gives an overview of datasets available for emotion analysis. Section 6 presents some of the techniques used for emotion analysis as found in past studies, and section 7 highlights some noteworthy findings concerning the state-of-the-art techniques for emotion analysis. We conclude the paper with some future research directions.

2. Method

Emotion is a complex state of mind, which represents the feeling of a person and which can influence both the physical and psychological behaviour of that individual [10]. Happiness, sadness, love, and hatred are some examples of emotions expressed by human beings. Emotions can be involved in that emotion, rather than being singular, which can consist of a cluster of emotions. It is also common for emotions to vary according to a person's personality traits and the corresponding context and environment [11].

In emotion analysis from the text, the aim is to correctly identify the actual state of mind of an individual when the written message was sent. In general, however, it is very challenging to detect emotions from texts for several reasons. Firstly, text messages can be highly unstructured and may not follow strict grammatical or syntactic rules such that a general approach for text processing cannot be applied. For instance, messages consisting of texts written as "*carooooooooooooooooooooooooooooo im going to kiiiiiiii uuuuuuuuuuuuuuuuuuuuu... n u know why but i still love u (a little bit:P) don't worry :P mwahhh*" are very hard to process [12].

Secondly, as [13] explained, context plays a vital role in emotion detection as evidenced by the following sentence, "The performers were greeted with joyless cheer". Here, the emotion factor of the word "cheer" is influenced by the word "joyless" which transforms the emotion expressed by the phrase "joyless cheer" to sadness rather than joy. Thirdly, sarcasm/metaphors, when expressed in messages, add further complexity when detecting emotion in text. For example, the sentence "*Yes. I guess your amazing delivery service has not yet arrived*" is clearly not expressing the feeling of a delighted customer [14]. Similarly, some expressions of anger emotion class are metaphorical, that is, they cannot be evaluated by the literal meaning of the expression (e.g. 'She is boiling with anger' or 'Don't snarl at me') [15].

Fourthly, expressions may constitute of multiple emotions, which may vary according to culture. For instance, according to [16], emotions are organized differently for each culture. In the Chinese culture, we can identify seven basic emotions: *happiness, anger, anxiety, thoughtfulness, sadness, fear, surprise* whereas in the Indian culture, we find eight core emotions, namely, *erotic, mirth, sorrow, anger, energy, fear, disgust, astonishment*. And finally, there are some messages that do use words at all. Rather, emotion expression is expressed via emoticons. E.g. *I still have Christmas shopping to do 😊* [17].

2.1. Applications of Emotion Analysis

Backed by the capabilities of the social web, emotion analysis has led to the creation or improvement of various applications and services [18]. For instance, by considering the emotions expressed during the navigation of websites, web designers obtain useful insights that help them design websites for improved user experience [19].

In the marketing field, customer emotional responses to promotional video campaigns or ads are used as a critical factor to determine strategies that can drive sales figures [20], [21]. In E-learning, automatic emotion recognition has been used to discover the emotional state of the learner to adapt to the learner's ability and provide an optimized learning experience [22], [23].

Psychologists can further use emotion analysis to detect topic sensitivity from facial emotion recognition and subtle voice changes. Emotion analysis can also help identify key emotional point that may require more in-depth investigations [24]. The application of emotion analysis also extends to social media. For instance, cyber-bullying incidents can be classified by detecting the emotional

state of users [25] and in customer service, tweets can be ranked based on the emotion expressed, and reply to customers can be prioritized accordingly, hence increasing customer satisfaction [14].

1) Emotion Models

Emotion models set forth the different criteria to make various emotions expressed by an individual measurable and distinguishable [26]. Five emotion models, namely, *discrete*, *dimensional*, *componential*, *circuit* and *appraisal model* have been reported in previous studies [27]. In this section a brief overview on each model is provided.

2) Discrete Model

Discrete models identify a set of fundamental or core emotions that are expressed by every person, regardless of religion, culture, or ethnicity [28]. In Ekman's basic emotion theory, six basic emotions identified are: *anger*, *disgust*, *fear*, *happiness*, *sadness*, and *surprise*. Extended Ekman's basic list by taking into consideration neuropsychological aspects to propose ten core emotions: *interest*, *joy*, *surprise*, *sadness*, *anger*, *disgust*, *contempt*, *fear*, *shame*, and *guilt* [29]. James Russell [30], however, critiqued discrete models as they do not necessarily provide accurate representations of an individual's feeling. For instance, *fear* is a basic emotion, but "*fear of getting wet*" cannot be considered equal to "*fear of bear*". "*Fear of getting wet*" will most likely be representing anger and "*fear of bear*" will be more representative of actual fear.

3) Dimensional Model

In contrast to the discrete emotion model, dimensional models map emotions into a continuous one-dimensional or multi-dimensional space. A one-dimensional model consists of a single dimension, and multi-dimensional may have two or more dimensions [27]. Dimension models represent different affective states as the point in a dimensional space with coordinates ranging from -1 to 1 in some cases, and in other cases, it can extend to higher ranges such as -100 to 100 [31].

The Pleasure Arousal and Dominance (PAD) emotional state model is an implementation of the dimensional model, which was introduced by [30]. To measure emotional states, PAD makes use of the following three dimensions: *pleasure*, for a degree of valence, *arousal*, indicating the level of affective activation and *dominance* for the degree of power or control [31].

Russel, for instance, put forth the circumplex model of affect, which states that any emotion can fit within two continuous dimensions of *valence* and *arousal* [30]. The *valence* (pleasure) dimension contrasts between positive and negative emotions, which can be evaluated on different axes with values nearer to zero, indicating neutral emotional states. Arousal (activation) dimension contrasts between active and passive emotional states, with a value of zero indicating intermediate states.

However, the Dimensional model of effect has been criticized for three main reasons [27]. Firstly, it is not natural for the human representation of emotions as people do not think about emotions as points. Secondly, it is hard to represent ambivalent emotional states, and finally, some emotions such as fear and anger are indistinguishable, as these emotions both lie in the same quadrant of high arousal and negative valence

4) Componential Model

Componential models consider that emotions are manifested by cerebral assessment of events and the sequence of reactions in different physiological responses, facial expressions, gestures, stance and affect [30]. Ortony, Clore and Collin (OCC) in 1988 defined a hierarchy of twenty-two emotion types in their OCC model to represent all possible affective states, which might be experienced by an individual [32]. In the OCC model, each emotion is a result of an affective reaction, which occurs after evaluating the aspects of a situation as positive (beneficial) or negative (harmful). The reactions are consequences of events, actions of agents, and aspects of objects [31].

Psychologist Robert Plutchik recognized that there are eight core emotions and that all other emotions grow from those core emotions. In Plutchik's psych evolutionary theory of emotion, he suggests that there are eight primary emotions, namely *anger*, *fear*, *sadness*, *disgust*, *surprise*, *anticipation*, *trust*, and *joy* and that various combinations of the eight basic emotions will form other emotions. For instance, *joy* and *surprise* can be combined to result into *delight*. The model is also known as *Plutchik Wheel of Emotions* [33]

[31] argued that complex models like the one proposed by [34] are rarely used in practical applications and the OCC model is used for emotion state prediction. Furthermore, in some cases, emotion recognition and prediction capabilities are merged to obtain a more precise emotional state.

5) Circuit Model

The Circuit model, also known as anatomical model, was proposed by neuroscientist, Joseph E. Ledoux [35]. Ledoux theorised that individual emotion can be processed in dissimilar, distinct neural circuits and essential tasks or systems are linked to those circuits. Neuropsychologists posit that evolutionary neural circuits in the brain construct elemental emotions and their distinctions. They discovered numerous important primitive emotions such as *rage, fear, expectancy, and panic*. Ledoux further insisted that language may not be enough to distinguish among emotions as there exists diverse circuits to compute dissimilar emotions in our brain, and which may not be necessary measurable or translated in a language format.

6) Appraisal Models

Smith and Lazarus proposed the appraisal model, which suggests that emotions emerge from the constantly changing communication of appraisal and coping processes that rely on the agent's depiction of its relationship with the environment [36]. The model also states that each environment relationship is assessed with the help of a defined number of appraisal dimensions or variables, and connections among environmental and affect changes. Appraisal variables are based on "if-then" rules, i.e. on relevance, implications, coping potential, and normative significance. Appraisal model is the leading theory related to human emotion in computer science, specifically in symbolic Artificial Intelligence (AI) systems.

2.2 Datasets available for Emotion Analysis

Emotion analysis is a relatively new field in computational linguistic [37], [38]. Consequently, quality datasets expressing all or most measurable emotions do not exist. In this section, we discuss some datasets, which have been used for emotion analysis.

The International Survey on Emotion Antecedents and Reactions (ISEAR) was a project lead by [39]. The authors constructed a dataset with seven out of ten basic emotions, joy, fear, anger, sadness, disgust, shame, and guilt based on a study in 37 countries. Despite its size, ISEAR only comprises of a personal reaction toward an event, which triggered an emotional response. For instance, "A girl entered in the division where I work and greeted everybody but not me", is an extract from the dataset which expresses anger. However, this form of the dialog is not expected in a conversation. The ISEAR dataset was built with the intention of sentence-level emotion recognition and not for conversation level analysis.

Klimt and Yang created the Enron Email Dataset, which consists of emails collected from one hundred and fifty-eight users from Enron senior management office [40]. The original dataset only contained emails and was not annotated. It was up to the researchers to annotate the mails. In 2017, Charlie Oxborough released sentence-level annotations to classify the mail dataset into two groups, namely negative and positive mails.

SemEval 2007 is another dataset created by Strapparava and Mihalcea [41]. SemEval2007 consists of news headlines extracted from news web sites like Google News, CNN, and newspapers. Collected news is classified under Ekman six basic emotions, Anger, Disgust, Fear, Joy, Sadness, Surprise with a valence indicating its sentence polarity. However, headlines are quite straightforward and enclosed such as, "Trucks swallowed in subway collapse."

Twitter is a popular social network where people share or express their opinions. Mohammad and Bravo-Marquez collected tweets and manually annotated the data under four discrete emotions, anger, fear, joy, and sadness for the WASSA-2017 dataset [42]. In another attempt to create a Twitter dataset, Figure Eight collected tweets that were classified under thirteen emotion labels: anger, enthusiasm, fun, happiness, hate, neutral, sadness, surprise, worry, love, boredom, relief, and empty. The dataset contains forty thousand tweets. However, the emotions used in the dataset do not match any of the popular emotion models discussed previously.

It is generally observed that labeled emotion datasets have been designed to address a specific text emotion classification problem. The main problem with currently available datasets is that the

emotions tagged belong to discrete human emotions. Authors attempting to resolve multi class emotion issues need to build their own or enhance existing dataset to suit their needs.

2.3 Emotion Analysis Techniques

Anusha and Sandhya proposed a supervised learning system, which makes use of Natural Language Processing (NLP), *Naïve Bayes Multinomial (NBM)* and *Support Vector Machines (SVM)* algorithms to classify text data into emotions [37]. The ISEAR dataset was used for training on five emotions classes: *Anger, Disgust, Fear, Joy, and Sadness* as per the Ekman emotion model. A 10-fold cross-validation was performed, and it was found that SVM yielded an Average F1 of 63.1 with Kappa statistics of 0.49 whereas NBM yielded an Average F1 of 52.9 with Kappa statistics of 0.43.

Udochukwu and He proposed a rule-based approach toward implicit emotion detection using five emotion classes in the OCC model [43]. The approach was tested on three datasets: ISEAR, SemEval2007 and Alm. Two baseline models were crafted to test the rule-based approach. The first one is a lexicon matching technique, which uses NRC emotion lexicon for sentence-level emotion detection, and the second one is a supervised Naive Bayes (NB) classifier. The performance was evaluated using F-Measure in five-fold cross-validation. The rule-base system was unable to classify any “Fear” affect bearing sentences on the three datasets. However, it surpassed NB by 9.5% on Alm dataset and achieved almost equivalent results to NB for the ISEAR and SemEval2007 datasets.

Perikos and Hatzilygeroudis, attempted emotion recognition from text using an ensemble classifier combining *Naïve Bayes (NB)*, *Maximum entropy (ME)*, and *Knowledge-Based Tools (KBTool)* [44]. The ensemble classifier is based on a voting function to make classification verdict based on the output of each base classifier. The authors argued that in comparison to *NB*, *ME* provides additional features, such as unigrams and bigrams, which can be added without risk of overlapping. The advantages of *ME* also extend to better performance in multiple Natural Language Processing tasks. However, it necessitates more time to be trained. The *KBTool* performs deeper sentence analysis by storing affect bearing words and using WordNet Affect. The model was trained on the ISEAR dataset on seven discrete emotions while Russell’s two-dimensional model of affect was used to detect emotion polarity. When tested on a manually crafted dataset, which consists of Tweets, news headlines and articles, an accuracy of 83% to 89%, precision of 85% to 90%, sensitivity of 79% to 91% and specificity of 86% to 89% were obtained.

Yasmina et al. used Point Wise Mutual Information (PMI) to compute text emotion similarity on YouTube comments [45]. The authors extended Agrawal and an, learning algorithm to take into consideration any convergence between sets of an emotion, which can result into distortion in classification results [13]. The classifier categorized comments for each category under Ekman six basic emotion model and obtained results accounting to 90% in precision, 72% for recall and 67-70% in accuracy.

Razek and Frasson, used Dominant Meaning Classifier (DMC) to recognize emotion from text [46]. A dominant tree is trained on the ISEAR dataset to form seven emotion classes, *joy, fear, anger, sadness, disgust, shame, and guilt*, which is in line with Carroll [29] discrete emotions. For each emotional class, a sub class is associated, for instance, under *anger* class, the sub classes *mislead, punishment, argument* and other emotions are added. The dominant tree is then utilized to classify text retrieved from users chat sessions. Instead of using keywords-based techniques, the authors adopted dominant meaning techniques to enhance the accuracy and refine the emotion classes. The following metrics were used to evaluate the accuracy of the model, *average precision, recall, and F-measure*. A ten-fold cross validation comparison was made between Support Vector Machine (SVM) and DMC. It was observed that DMC yielded superior results across all emotion classes.

To tackle the problem of emotion analysis from a psychological and linguistic perspective, [38] developed a framework designed to capture emotions from multilingual text using Ekman six basic emotions model. Two additional emotion classes were added: “Mixed Emotion” for sentences with multiple affects and “No Emotion” for sentences bearing no affective words. The dataset used was constructed from Twitter in three different areas, political election, healthcare, and sports. Latent Dirichlet Allocation (DLA) was applied for the extraction of repeating topics and keywords. Each tweet was manually labeled with an emotion class by four human annotators. The task of emotion classification was done in two segments. In the first segment, the dataset is split into two categories, emotion, and non-emotion, which is classified using SVM. In the second segment, fine tuning is

performed using SVM and NB. To cater for automatic emotion classification, three publicly available lexical resources, WorldNet-Affect (WNA), Hindi WordNet-Affect (HWNA) and Senti-WorldNet, were used to create three features sets to distinguish between affective and non-affective words. Evaluation was conducted using the following metrics: Precision, Recall, Accuracy, and F-measure. Comparison between SVM and NB showed that NB outperformed SVM with an accuracy of 72.81%.

For messages with text sparsity, proposed two supervised intensive topic model namely: Weighted Labelled Topic Model (WLTM) and Intensive Emotion Topic Model (IETM) for emotion detection over short texts [47]. WLTM performs biterm (pair of words) extraction, which matches the topic from a document label set. Gibbs sampling algorithm was used to estimate the required parameters for WLTM. Then, Support Vector Regression (SVR) is used to forecast emotion distributions. IETM also performs biterm extraction. The two models were trained on the SemEval and ISEAR datasets. Averaged Pearson's correlation coefficient, APdocument and APemotion were used to evaluate the models. WLTM result on SemEval did not perform well in terms of APdocument (0.24); however, the same model achieved the highest score on APemotion (0.45).

In another emotion classification task, [48], employed Prediction by Partial Matching (PPM) technique to recognize Ekman's six basic emotions in character-based text. The PPM technique was experimented on three datasets, LiveJournal dataset, Alm's dataset and Aman's dataset to categorize emotions. The model obtained Accuracy within the range of 88% to 96%, Precision between 71% to 90%, Recall between 70% to 88% and F-Measure between 67% to 88%.

Hasan et al. further proposed an approach to automatic emotion detection from tweets by developing two systems called Emotex and EmotexStream [49]. The authors extended the circumplex model with WordNet's synsets (synonym) to capture a broader spectrum of affect bearing words. NB, SVM and Decision tree as classifiers were used in Emotex. EmotexStream as an extension to Emotex was developed as a real-time tweet classifying system whose aim is to discover temporal distributions of aggregate emotion and detect emotional burst during major events. An unsupervised method (Binary classification using Linguistic Inquiry and Word Count (LIWC) and Affective Norms for English Words (ANEW) was developed to classify tweets in two groups, emotion-present and emotion-absent, which were fed in Emotex for emotion classification. Results for precision between 78% to 93%, recall between 77% to 95% and F-Measure between 77.8% to 85.6% were observed.

A method for unlabelled text emotion classification, called Universal Affective Model (UAM) was proposed by [50]. Their objectives were to detect social emotions from the point of view of social media users and to classify unlabelled text with limited features. Three steps are involved, keywords identification, biterm extraction and emotion prediction of unlabelled text with limited features. Their model was tested on three datasets, SemEval, Six and Sinanews. Six is a collection of small texts from BBC Forum posts, Digg.com, Myspace, Twitter, YouTube, and Runners World. Sinanews is a collection of news articles and contains eight emotion class, touching, empathy, boredom, anger, amusement, sadness, surprise, and warmth. The three datasets were evaluated using three metrics, namely, AP, APemotion, and Accu@1. The authors reported 35.7, 24.1, 36.7 in SemEval for each metric respectively, 54, 44.4, 77.2 in Six, and 54.5, 41.3, 54.7 in Sinanews.

Chen et al. used Agglomerative Hierarchical Clustering and the Valence-Arousal (V-A) emotion (based on Plutchik wheels of emotions) dimensional space to monitor and analyze users' emotions when chatting online [51]. Emoticons were manually mapped into the V-A space. PointWise Mutual Information criterion was used to calculate the correlations between chat messages and emoticons in terms of scores. The authors conducted data clustering to automatically detect emotions in a conversation and reported an accuracy of 88%.

In an attempt to detect emotion from Twitter data, [52] made use of C-GRU (Context-aware Gated Recurrent Units) for context extraction when determining user feelings. Emotions were classified using the twelve discrete emotions: Anger, Anticipation, Disgust, Fear, Joy, Love, Optimism, Pessimism, Sadness, Surprise, Trust, and Neutral. The authors reported an accuracy of 0.532 and an F1 score of 0.64.

Kratzwald et al. attempted emotion recognition using long-short term memory (LSTM) as per [53]. The proposed approach was tested on SemEval-2015 (a set election tweets datasets) and SemEval-2018 (a set of general tweets datasets). Four discrete emotions categories, *Anger*, *Fear*, *Joy*, *Sadness* were targeted. Overall, the authors reported performance F1-score 58.4% for election tweets and

58.6% for general tweets. Another method consisting of the use of LSTM on text emotion recognition was further proposed by Su et al. (2018). The method was tested against Natural Language Processing and Chinese Computing (NLPCC) database which contains seven emotion categories: anger, boredom, disgust, anxiety, happiness, sadness, surprise and results obtained indicated an accuracy of 70.66%.

In an attempt to address Multiclass emotion classification, [55] put forth an approach using emotion distribution learning and a Multi-task Convolutional Neural Network for text emotion analysis. The proposed approach was evaluated on SemEval 2007 dataset on six Distribution Prediction: Euclidean, Sørensen, SquaredX2, KL divergence, Cosine, and Intersection and four classification performance: Precision, Recall, F-score, and Accuracy metrics. Distribution Prediction for each metric are as follows: 44.38, 41.96, 55.19, 73.06, 72.91, 58.04 and classification performance for each metric are as follows: 48.33, 42.23, 41.41, 51.60. Single class emotion classification is evaluated on the ISEAR, Fairy Tales, TEC and CBET datasets using the same performance metrics as multiclass emotion classification. Precision between 61.20 and 67.11, Recall between 52.57 and 79.2, F-score between 56.94 and 78.72 and Accuracy between 61.52 and 79.21 were obtained.

Chatterjee et al. attempted the detection of four emotion labels, Happy, Sad, Angry, Others, using a Deep learning approach called Sentiment and Semantic-Based Emotion Detector (SS-BED) which is based on the combination of semantic and sentiment representations of user text and makes use of two LSTM layers [14]. The authors created a dataset based on Twitter conversations. Those conversations are pre-processed and were assigned to five judges who in turn classified the conversation to an emotion class. SS-BED obtained precision, recall and F1 score of 69.51%, 52.29%, 59.68, respectively for Happy, 85.42%, 76.63%, 80.79% for Sad and 87.69%, 63.33%, 73.55%, for Angry

3. Results and Discussion

Table 1 provides a summary of the different techniques, datasets, and evaluation metrics commonly used in emotion analysis research since 2015. It is generally observed that researchers have adopted hybrid implementation, that is, a combination of two or more algorithms to handle multiclass emotion classification with SVM and NB as the most used algorithms.

The most used datasets are ISEAR and SemEval, both of which contain texts that express discrete emotions. We also note that the highest performing algorithm is PPM with an accuracy of up to 96% and the least performing algorithm is C-GRU with an accuracy of 53.2%. We also observe an inconsistency in the way performances are reported. Authors do not use the same metrics, datasets, and emotion models such that it is not possible to compare the different emotion analysis techniques. Moreover, it is found that earlier studies adopted a general statistical approach for classifying emotions, but recent studies are seen to adopt machine learning techniques, notably deep learning to address the problem of emotion analysis.

Furthermore, as summarized in Table 1, authors either limit themselves to a single dataset or they may combine different datasets with a varying choice of different algorithms for emotion analysis. Given that each dataset adopts different emotion models (basic level emotions to different emotion levels), results reported would be significant only to the actual emotions present within the datasets. Eventually, there is concern regarding the relevance of reported performance into real-world applications. As is, most studies tend to focus mostly on the datasets overlooking any implications to real-world settings. Moreover, current datasets suffer from emotion imbalances, that is, they do not contain equal amount of all emotions classes, which makes it difficult for any classification system to calculate performance correctly.

Table 1. Techniques used for text emotion analysis over past five years

Technique(s)	Dataset(s)	Emotion Model	Performance (%)	Source
NLP, NBM, SVM	ISEAR	5 discrete emotions	F1- Score: 63.1	Anusha and Sandhya (2015)
Rule-based approach: NRC emotion lexicon, NB NB, ME	ISEAR, SemEval-2007, Alm	OCC model	F1- Score: 51.3 - 65.5	Udochukwu and He (2015)
PMI	Twitter	7 discrete emotions, Circumplex model of affect	Accuracy: 83 - 89 Precision: 85 - 90 Sensitivity: 79 - 91 Specificity: 86 - 89	Perikos and Hatzilygeroudis (2016)
DMC	YouTube comments	Ekman basic emotions	Precision: 92.75 Recall: 72 Accuracy: 67 - 70	Yasmina et al. (2016)
SVM, NB	ISEAR	7 discrete emotions	Precision: 20.2 - 27.2 Recall.: 46.9 - 60.2	Razek and Frasson (2017)
WLTM, IETM	Twitter	Ekman Basic Emotions	Accuracy: 72.81	Jain et al. (2017)
PPM	SemEval, ISEAR	7 discrete emotions	AP _{document} : 0.24 AP _{emotion} : 0.45	Rao et al. (2017)
UAM	LiveJournal, Alm, Aman	Ekman basic emotions	Accuracy: 88 - 96 Precision: 71 - 90 Recall: 70 - 88 F1-Score: 67 - 88	Almahdawi and Teahan (2017)
NB, SVM, and Decision tree	SemEval, Six, Sinanews	Discrete + Dimensional	AP: 35.7 - 54.5 AP _{emotion} : 24.1 - 44.4 Accu@1: 36.7 - 77.2	Liang et al. (2018)
Agglomerative hierarchical clustering, PMI C-GRU	Twitter	Circumplex model of the effect	Precision: 78 - 93 Recall: 77 - 95 F1-Score: 77.8 - 85.6	Hasan et al. (2018)
LSTM,	Real-time chat data	Plutchik wheels of emotions	Accuracy: 88	Chen et al. (2018)
LSTM	Twitter	12 discrete emotions	Accuracy: 53.2 F1-score: 64	Samy et al. (2018)
Multi-task: CNN, LDL,	SemEval-2015, SemEval-2018	4 discrete emotions	F1-score: 58.4 - 58.6	Kratzwald et al. (2018)
LSTM	NLPCC-MHMC-TE	7 Emotions categories	Accuracy: 70.66	Su et al. (2018)
LSTM	SemEval-2007, ISEAR, Fairy Tales, TEC, CBET	Mix emotions	Precision: 61.20 - 67.11 Recall: 52.57 - 79.2 F1 score: 56.94 - 78.72 Accuracy: 61.52 - 79.21	Wang et al. (2018)
LSTM	Twitter	4 discrete emotions	Precision: 69.51 Recall: 52.29 F1 Score: 59.68	Chatterjee et al. (2019)

4. Conclusion

This paper has provided an overview of the emotion analysis techniques used in past studies. It is observed that various algorithms have been adopted to classify emotions, but results reported are not comparable and sometimes deceiving since there is little scope for practical applications. Despite the existence of various emotion models, none can be considered sufficient to cover the range of emotions that are usually expressed by an individual. The same limitation is observed within the current datasets available for emotion analysis. Datasets adopt a single emotion model and are imbalanced. With the increasing amount of data and improved techniques for data analysis, there is a good scope for the field of emotion analysis to move beyond academic research and find its way into real-world applications. However, prior to that, researchers must consider the development of robust and reliable emotion models, which can be adopted in emotion analysis studies. Moreover, researchers are

encouraged to consider the development of benchmarks to allow for comparison of reported results across studies

References

- [1] T. Kamboj and M. Dayal, "Social media: Collaborating Web 2.0 and user-generated content (UGC)," *Int. Multidiscip. J. Appl. Res.*, vol. 1, no. 3, pp. 35–41, 2014.
- [2] N. A. Ghani, S. Hamid, I. A. T. Hashem, and E. Ahmed, "Social media big data analytics: A survey," *Comput. Human Behav.*, vol. 101, no. December, pp. 417–428, 2019, doi: <https://doi.org/10.1016/j.chb.2018.08.039>.
- [3] A. B. Youssef and H. B. Youssef, "Social networking on Web 2.0: From emotional intelligence to cyber emotional intelligence," *Manag. Inf. Syst.*, vol. 6, no. 2, pp. 21–28, 2011.
- [4] M. Fernández-Gavilanes, J. Juncal-Martínez, S. García-Méndez, E. Costa-Montenegro, and F. J. González-Castaño, "Creating emoji lexica from unsupervised sentiment analysis of their descriptions," *Expert Syst. Appl.*, vol. 103, no. August, pp. 74–91, 2018, doi: <https://doi.org/10.1016/j.jocs.2017.01.010>.
- [5] Y. Zhang, D. Song, P. Zhang, P. Wang, X. Li, J. Li, and B. Wang, "A quantum-inspired multimodal sentiment analysis framework," *Theor. Comput. Sci.*, vol. 752, no. December, pp. 21–40, 2018, doi: <https://doi.org/10.1016/j.tcs.2018.04.029>.
- [6] J. Serrano-Guerrero, J. A. Olivas, F. P. Romero, and E. Herrera-Viedma, "Sentiment analysis: A review and comparative analysis of web services," *Inf. Sci. (Ny)*, vol. 311, no. August, pp. 18–38, 2015, doi: <https://doi.org/10.1016/j.ins.2015.03.040>.
- [7] M. Munezero, C. S. Montero, E. Sutinen, and J. Pajunen, "Are they different? Affect, feeling, emotion, sentiment, and opinion detection in text," *IEEE Trans. Affect. Comput.*, vol. 5, no. 2, pp. 101–111, 2014, doi: <https://doi.org/10.1109/TAFFC.2014.2317187>.
- [8] A. A. Lazarus and C. N. Lazarus, "The 60-second shrink," *Atascadero, CA*. 1997.
- [9] C. Y. Yam, "Emotion detection and recognition from text using deep learning," *Microsoft Developer Blog*. 2015.
- [10] D. Seal, U. K. Roy, and B. Basak, "Sentence-level emotion detection from text based on semantic rules," in *Information and Communication Technology for Sustainable Development*, Springer, 2020, pp. 423–430.
- [11] Ö. F. Şimşek, "Happiness revisited: Ontological well-being as a theory-based construct of subjective well-being," *J. Happiness Stud.*, vol. 10, no. 5, pp. 505–522, 2009, doi: <https://doi.org/10.1007/s10902-008-9105-6>.
- [12] M. Yassine and H. Hajj, "A framework for emotion mining from text in online social networks," in *2010 IEEE International Conference on Data Mining Workshops*, 2010, pp. 1136–1142, doi: <https://doi.org/10.1109/ICDMW.2010.75>.
- [13] A. Agrawal and A. An, "Unsupervised emotion detection from text using semantic and syntactic relations," in *Proceedings - 2012 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2012*, 2012, pp. 346–353, doi: [10.1109/WI-IAT.2012.170](https://doi.org/10.1109/WI-IAT.2012.170).
- [14] A. Chatterjee, U. Gupta, M. K. Chinnakotla, R. Srikanth, M. Galley, and P. Agrawal, "Understanding emotions in text using deep learning and big data," *Comput. Human Behav.*, vol. 93, no. April 2018, pp. 309–317, 2019, doi: [10.1016/j.chb.2018.12.029](https://doi.org/10.1016/j.chb.2018.12.029).
- [15] M. Aksan, "Metaphors of anger: An outline of a cultural model," *Dil ve Edeb. Derg.*, vol. 3, no. 1, pp. 31–67, 2011.
- [16] Z. Shao, R. Chandramouli, K. P. Subbalakshmi, and C. T. Boyadjiev, "An analytical system for user emotion extraction, mental state modeling, and rating," *Expert Syst. Appl.*, vol. 124, no. 15, pp. 82–96, 2019, doi: [10.1016/j.eswa.2019.01.004](https://doi.org/10.1016/j.eswa.2019.01.004).
- [17] P. S. Dandannavar, S. R. Mangalwede, and S. B. Deshpande, "Emoticons and their effects on sentiment analysis of Twitter data," in *EAI International Conference on Big Data Innovation for Sustainable Cognitive Computing*, 2020, pp. 191–201.

-
- [18] R. Newman, V. Chang, R. J. Walters, and G. B. Wills, "Web 2.0—The past and the future," *Int. J. Inf. Manage.*, vol. 36, no. 4, pp. 591–598, 2016, doi: <https://doi.org/10.1016/j.ijinfomgt.2016.03.010>.
- [19] S. F. Pengnate and R. Sarathy, "An experimental investigation of the influence of website emotional design features on trust in unfamiliar online vendors," *Comput. Human Behav.*, vol. 67, no. February, pp. 49–60, 2017, doi: <https://doi.org/10.1016/j.chb.2016.10.018>.
- [20] R. P. Bagozzi, M. Gopinath, and P. U. Nyer, "The role of emotions in economic," *J. Acad. Mark. Sci.*, vol. 27, no. 2, pp. 184–206, 1999, doi: <https://doi.org/10.1001/jama.1953.03690150021006>.
- [21] I. O. Pappas, P. E. Kourouthanassis, M. N. Giannakos, and V. Chrissikopoulos, "Sense and sensibility in personalized e-commerce: How emotions rebalance the purchase intentions of persuaded customers," *Psychol. Mark.*, vol. 34, no. 10, pp. 972–986, 2017, doi: <https://doi.org/10.1002/mar.21036>.
- [22] A. Kołakowska, A. Landowska, M. Szwoch, W. Szwoch, and M. R. Wrobel, "Emotion recognition and its applications," in *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Springer, 2014, pp. 51–62.
- [23] P. Sharma, M. Esengönül, S. R. Khanal, T. T. Khanal, V. Filipe, and M. J. Reis, "Student concentration evaluation index in an e-learning context using facial emotion analysis," in *International Conference on Technology and Innovation in Learning, Teaching and Education*, 2018, pp. 529–538, doi: <https://doi.org/10.1007/978-3-030-20954-4>.
- [24] M. F. Wagner, J. S. Milner, R. J. McCarthy, J. L. Crouch, T. R. McCanne, and J. J. Skowronski, "Facial emotion recognition accuracy and child physical abuse: An experiment and a meta-analysis," *Psychol. Violence*, vol. 5, no. 2, p. 154, 2015.
- [25] S. Salawu, Y. He, and J. Lumsden, "Approaches to automated detection of cyberbullying: A survey," *IEEE Trans. Affect. Comput.*, 2017.
- [26] K. R. Scherer and H. G. Wallbott, "Evidence for universality and cultural variation of differential emotion response patterning," *J. Pers. Soc. Psychol.*, vol. 66, no. 2, p. 310, 1994.
- [27] K. Kim and R. W. Picard, "Emotion modeling and machine learning in affective computing," *Unpublished manuscript*. 2014.
- [28] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *J. Pers. Soc. Psychol.*, vol. 17, no. 2, p. 124, 1971.
- [29] C. E. Izard, D. Z. Libero, P. Putnam, and O. M. Haynes, "Stability of emotion experiences and their relations to traits of personality," *J. Pers. Soc. Psychol.*, vol. 64, no. 5, p. 847, 1993.
- [30] A. Mehrabian and J. A. Russell, *An approach to environmental psychology*. Cambridge, MA: The MIT Press, 1974.
- [31] A. Kołakowska, A. Landowska, M. Szwoch, W. Szwoch, and M. R. Wróbel, *Modeling emotions for affect-aware applications*, vol. January. 2015.
- [32] A. Ortony, G. L. Clore, and A. Collins, *The cognitive structure of emotions*. Cambridge university press, 1990.
- [33] R. Plutchik, "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," *Am. Sci.*, vol. 89, no. 4, pp. 344–350, 2001.
- [34] R. Plutchik, "The nature of emotions," *Am. Sci.*, vol. 86, no. 4, p. 344, 2001, doi: <https://doi.org/10.1511/2001.4.344>.
- [35] J. M. Fellous, J. L. Armony, and J. E. LeDoux, *Emotional circuits and computational neuroscience*, vol. 2. MIT Press Cambridge, MA, 2002.
- [36] C. A. Smith and R. S. Lazarus, *Emotion and adaptation*. New York, 1990.
- [37] E. S. M. El-Alfy, S. M. Thampi, H. Takagi, S. Piramuthu, and T. Hanne, *Advances in intelligent informatics*, vol. 320. Springer International Publishing, 2015.
-

-
- [38] V. K. Jain, S. Kumar, and S. L. Fernandes, "Extraction of emotions from multilingual text using intelligent text processing and computational linguistics," *J. Comput. Sci.*, vol. 21, no. July, pp. 316–326, 2017, doi: 10.1016/j.jocs.2017.01.010.
- [39] K. R. Scherer, "What are emotions? And how can they be measured?," *Soc. Sci. Inf.*, vol. 44, no. 4, pp. 695–729, 2005.
- [40] B. Klimt and Y. Yang, "Introducing the Enron corpus," *CEAS*. 2004.
- [41] C. Strapparava and R. Mihalcea, "Semeval-2007 task 14: Affective text," in *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, 2007, pp. 70–74.
- [42] S. M. Mohammad and F. Bravo-Marquez, "WASSA-2017 shared task on emotion intensity," *arXiv preprint arXiv:1708.03700*. 2017.
- [43] O. Udochukwu and Y. He, "A rule-based approach to implicit emotion detection in text," in *International Conference on Applications of Natural Language to Information Systems*, 2015, pp. 197–203.
- [44] I. Perikos and I. Hatzilygeroudis, "Recognizing emotions in text using ensemble of classifiers," *Eng. Appl. Artif. Intell.*, vol. 51, no. May, pp. 191–201, 2016, doi: 10.1016/j.engappai.2016.01.012.
- [45] D. Yasmina, M. Hajar, and A. M. Hassan, "Using YouTube comments for text-based emotion recognition," *Procedia Comput. Sci.*, vol. 83, no. Ant, pp. 292–299, 2016, doi: 10.1016/j.procs.2016.04.128.
- [46] M. A. Razek and C. Frasson, "Text-based intelligent learning emotion system," *J. Intell. Learn. Syst. Appl.*, vol. 9, no. 1, pp. 17–20, 2017, doi: 10.4236/jilsa.2017.91002.
- [47] Y. Rao *et al.*, *Supervised intensive topic models for emotion detection over short text*, vol. 1. 2017.
- [48] A. Almahdawi and W. J. Teahan, "Emotion recognition in text using ppm," in *In International Conference on Innovative Techniques and Applications of Artificial Intelligence*, 2017, pp. 149–155, doi: 10.1007/978-3-319-71078-5_13.
- [49] M. Hasan, E. Rundensteiner, and E. Agu, "Automatic emotion detection in text streams by analyzing twitter data," *Int. J. Data Sci. Anal.*, vol. 7, no. 1, pp. 35–51, 2019, doi: <https://doi.org/10.1007/s41060-018-0096-z>.
- [50] W. Liang, H. Xie, Y. Rao, R. Y. Lau, and F. L. Wang, "Universal affective model for Readers' emotion classification over short texts," *Expert Syst. Appl.*, vol. 114, no. December, pp. 322–333, 2018, doi: 10.1016/j.eswa.2018.07.027.
- [51] C. H. Chen, W. P. Lee, and J. Y. Huang, "Tracking and recognizing emotions in short text messages from online chatting services," *Inf. Process. Manag.*, vol. 54, no. 6, pp. 1325–1344, 2018, doi: 10.1016/j.ipm.2018.05.008.
- [52] A. E. Samy, S. R. El-Beltagy, and E. Hassanien, "A context integrated model for multi-label emotion detection," *Procedia Comput. Sci.*, vol. 142, pp. 61–71, 2018, doi: 10.1016/j.procs.2018.10.461.
- [53] B. Kratzwald, S. Ilić, M. Kraus, S. Feuerriegel, and H. Prendinger, "Deep learning for affective computing: Text-based emotion recognition in decision support," *Decis. Support Syst.*, vol. 115, no. December, pp. 24–35, 2018, doi: 10.1016/j.dss.2018.09.002.
- [54] M. H. Su, C. H. Wu, K. Y. Huang, and Q. B. Hong, "LSTM-based text emotion recognition using semantic and emotional word vectors," in *2018 1st Asian Conference on Affective Computing and Intelligent Interaction, ACII Asia 2018*, 2018, pp. 1–6, doi: 10.1109/ACIIAsia.2018.8470378.
- [55] Y. Zhang, J. Fu, D. She, Y. Zhang, S. Wang, and J. Yang, "Text emotion distribution learning via multi-task convolutional neural network," in *IJCAI International Joint Conference on Artificial Intelligence*, 2018, vol. 2018-July, pp. 4595–4601.
-