

ARTICLE • OPEN ACCESS

Text emotion mining on Twitter

To cite this article: Suboh M Alkushayni *et al* 2020 *IOPSciNotes* 1 035001

View the [article online](#) for updates and enhancements.

physicsworld

From collaboration comes innovation.

Siemens Healthineers is driving continuous improvement in the planning, delivery and management of proton therapy for cancer patients.

Click to discover more

SIEMENS
Healthineers





ARTICLE

Text emotion mining on Twitter

OPEN ACCESS

RECEIVED
19 July 2020

REVISED
5 October 2020

ACCEPTED FOR PUBLICATION
12 October 2020

PUBLISHED
21 October 2020

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Suboh M Alkhushayni¹ , Daniel C Zellmer^{1,2}, Ryan J DeBusk² and Du'a Alzaleq¹

¹ Minnesota State University, Mankato WI, United States of America

² University of Wisconsin, Platteville WI, United States of America

E-mail: suboh.alkhushayni@mnsu.edu

Keywords: emotion mining, Social Media Analysis, machine learning, mental health, ETDD

Abstract

Twitter has become a medium through which a substantial percentage of the global population communicates their feelings and reactions to current events. Emotion mining from text aims to capture these emotions by using a series of algorithms to evaluate the contents of each tweet. In this study, tweets that expressed at least one of seven basic emotions were collected. The resulting dataset was a corpus of 42,000 tweets with a balanced presence of each emotion. From this corpus a lexicon of roughly 40,000 words, each associated with a weighted vector corresponding to one of the emotions, was created. Next, different methods of identifying emotion in these 'cleaned' tweets were performed and evaluated. These methods included both lexically-based classification and supervised machine learning-based classification. Finally, an ensemble method involving several multi-class classifiers trained on unigram features of the lexicon was evaluated. This evaluation revealed that the ensemble method outperformed all other tested methods when tested on existing datasets as well as on the dataset created for this study.

Abbreviations

CBET	Cleaned Balanced Emotional Tweet Dataset
ETTD	Emotionally Tagged Twitter Dataset
URL	Universal Resource Locator
SVM	Support Vector Machine
TEC	Twitter Emotion Corpus

Introduction

Emotion mining, or emotion identification, generally describes the practice of determining and analyzing the feeling(s) expressed towards companies, products, events, people, etc [1]. Emotion mining can be used to label a variety of forms of human input, including spoken words, written words, and facial expressions with one or more classifications from a set of basic emotions [15]. According to Neel Burton of Green-Templeton College, these basic emotions are 'hardwired', where each basic emotion corresponds to a distinct and dedicated neurological circuit' [2]. However, the disagreements on a set of primitive or basic emotions, as well as how one would classify an emotion as basic, is debated to this day among emotional psychologists. Thus, choosing the 'proper' set of emotions typically incorporates subjectivity. This study addressed the written word and classification of human writing based on a unique set of seven basic emotions consistent with the model proposed by Ekman [3] with the addition of *guilt*.

Related work

Although the emotion model chosen for this study has several benefits in the context of this work, several research groups have proposed alternative sets of 'basic' emotions that range in size from two to over 18

categories. Ortony and Turner construct a table of several of these emotion sets developed in the late nineteenth and twentieth centuries [4]. From this table, we can generate a frequency distribution, from which we conclude that the six most common ‘basic’ emotions are: *fear, anger, disgust, sadness, joy, and surprise*. Note that these six most popular emotions are a direct representation of the model of basic emotions proposed in 1972 by P Ekman, W Friesen, and P Ellsworth [3]. Shahraki and Zäiane, creators of the CBET (cleaned balanced emotional tweet) dataset, also recognize that basic human emotions have been a controversial issue among scientific studies and admit that many models of the human emotional spectrum exist [5]. Like Shahraki and Zäiane’s work on the CBET dataset, we base our emotion set on Ekman’s basic emotion model (*anger, disgust, fear, joy, sadness, surprise*) in this work due to its clear distinction between emotions and relative simplicity of the model [5]. Like [5] *guilt* was included as a basic emotion since *guilt* is another commonly recognized basic emotion by many psychologists such as C E Izard [6]. Since the far-reaching goal of this research includes assisting mental health professionals with the detection of those who are suffering or may suffer from symptoms of depression, *guilt* was included to pinpoint tweets that might help alert psychologists detect such conditions.

Due to the complexity of multi-class supervised machine learning classification, few works have been done with the focus of such classification. Problems can arise such as needing large amounts of training data, computing power, and time. Within these works, even fewer have been done in the specific context of Twitter. Most notable among these is the work of Shahraki and Zäiane on the CBET dataset [5]. Their work consisted of testing and evaluating several lexical and learning-based methods on the CBET dataset, which included classification using both Naïve Bayes and binary SVM classifiers. Classification problems such as this are often addressed using one of two primary approaches. The first of these is a *lexical* approach, which employs a vocabulary to tag a piece of text with a corresponding emotion [5]. Also available is a *learning* approach, which uses a (set of) trained machine learning algorithm(s) to predict the emotion of test data based on patterns extracted from training data [5]. The classification method employed on our dataset is a hybrid of the method used by Kinsler [7] in that the method is based on ensemble of multi-class classifiers. The ‘votes’ in this case are the tags outputted by several machine learning algorithms, from which the most popular is selected as the tag for the inputted tweet. We say that our system is primarily based on Kinsler’s, but there exists a key difference between the application methodologies. Kinsler uses his voting classifier for sentiment analysis, wherein binary (or n-class where $n = 2$) classification is used [7]. Given the set of emotions used to construct our dataset, we require a seven-class classification system, which significantly increases the problem’s complexity.

Jabreel, and Antonio [8], presented a new approach to the multi-label emotion classification task, by proposing a transformation method to transform the problem into a single binary classification, Then they developed a deep learning-based system to solve the transformed problem; achieving a Jaccard (i.e., multi-label accuracy) score of 0.59 on the challenging SemEval2018 Task 1:E-c multi-label emotion classification problem.

Zhao, Xiaolin, and Xuejun [9], applied convolution algorithm on Twitter sentiment analysis to train deep neural network, in order to improve the accuracy and analysis speed. They first learned global vectors for word representation by unsupervised learning on large Twitter corpora, which expresses the word sentiment information as the words embeddings. Afterwards, they concatenated these word representation with the prior polarity score feature and state-of-the-art features as sentiment feature set.

One of the reasons that traditional text classification techniques were used on this work is that we can get good results when data available is not too large and computational resources are scarce, and many of these techniques do not need much training data to start providing accurate results. While deep learning algorithms require much more training data than traditional machine learning algorithms, i.e. at least millions of tagged examples. On the other hand, traditional machine learning algorithms such as SVM and NB reach a certain threshold where adding more training data doesn’t improve their accuracy. In contrast, deep learning classifiers continue to get better the more data you feed them with, also many Deep learning algorithms such as Word2Vec or GloVe are used in order to obtain better vector representations for words and improve the accuracy of classifiers trained with traditional machine learning algorithms.

Methodology

Creating the emotionally tagged twitter dataset (ETTD)

Twitter’s streaming API provides access to the enormous flow of tweets from millions of Twitter users [17]. These tweets provide not only the thoughts and feelings of those who write them, but also, a fascinating challenge of determining the emotion of each tweet despite the generally smaller amount of text when compared to other social media resources such as Facebook. Another challenge in using this data is presented by the rarity of Twitter-based datasets designed with the intention of text emotion mining or emotion identification. At the time of this study, there were only three publicly available datasets containing tweets tagged by emotion. Each dataset, however, had certain characteristics that necessitated the creation of a new dataset for this study.

Table 1. Hashtags used to search for tweets.

Emotion	Hashtags used
Joy	#joy, #joyful, #happy, #happiness, #bliss
Fear	#fear, #anxiety, #dread
Surprise	#surprise, #surprised, #revelation
Sadness	#sadness, #sad
Disgust	#disgust, #disgusted, #disgusting
Anger	#anger, #angry, #livid, #mad, #enraged, #furious, #outraged, #rage
Guilt	#guilt, #guilty, #sorry

Wang *et al* [10] created an unbalanced dataset, with the most popular emotional tag, *joy*, occurring in roughly 700,000 tweets while the least common tag, *surprise*, occurred in only 24,000 tweets. Also, the words used for filtering tweets included some words that were interpreted as having a weaker correspondence to their associated emotion (*annoyed* for *anger*, for instance). Mohammad's *Twitter Emotion Corpus (TEC)* [11], like that of Wang *et al* was unbalanced, in addition to being restricted in size due to the limited number of keywords used in their data collection. In contrast to the first two, Hasan's dataset [12] was excluded not for the previously stated issues, but instead for the use of an emotion model (*Happy-active*, *Happy-inactive*, *Unhappy-active*, and *Unhappy-inactive*) which differed greatly from the model used in this study (*Joy*, *Fear*, *Surprise*, *Sadness*, *Disgust*, *Anger*, and *Guilt*). The CBET dataset, created by Shahraki and Zäiane [5], was distinct as it was designed with a balanced presence of each emotion in mind. The CBET dataset also showed promise due to its construction being based on a similar set of emotions and similar data collection methodology. However, due to the inclusion of the emotions (*love* and *thankfulness*) outside the set of emotions used in this study, the CBET dataset was not used as the study's primary dataset. Thus, the creation of the ETTD was centered on equal representation of all emotions for the model proposed for this study. In a method consistent with [5, 10–12], hashtags served as filtering criteria for tweets during data collection.

Table 1 displays the list of hashtags used to retrieve tweets of the corresponding emotion. In each case, more than one hashtag was used for retrieval. Multiple keywords were used to filter for each emotion to add variability within each emotional class and to better represent the entirety of each emotional class. The goal of the data collection period was to ensure a roughly equivalent proportion of each emotion's tweets in the entire dataset after preprocessing. The hashtags #joy, #fear, #surprise, #sadness, #disgust, #anger, and #guilt match the name of their corresponding emotions. However, hashtags such as #happy with #joy were included for *joy* due to the more common use of the term *happy* when describing *joy* on Twitter. This same principle applied for #sad with #sadness for *sadness* and #rage with #anger for *anger*.

Tweets were collected in a time frame of one week from August 13th, 2018 to August 20th, 2018. These tweets were filtered, using the keywords from table 1, from Twitter's general stream via Twitter's streaming API and labeled with the emotion corresponding to the keywords for which the tweet was selected. All collected tweets were composed between the dates of August 13th, 2018 and August 20th, 2018. From this dataset, all retweet prefixes ('RE') were removed. Next, any duplicated tweets were detected and removed. All non-alphanumeric characters were then removed from each tweet, as well as all user mentions (i.e. '@username'). Tweets that were shorter than four words were removed next. For tweets with a Dice similarity greater than 0.4, only one tweet was kept. Any contractions and acronyms in the remaining tweets were expanded to their component words. Prior to tokenization, all URL's, stop words, numbers, useless punctuation, and redundant white spaces were removed. Finally, 3,500 samples of each emotion were selected and tokenized to create a preliminary dataset of 24,500 samples.

Emotion classification

Several methods for predicting the emotion of tweets were performed and evaluated. We compare: a lexical-based approach, individual multi-class classifiers, and an ensemble of the top-performing multi-class classifiers as similar to the classifier created by Kinsler [7].

Lexical approach

Adopting the methodology used by [5], a lexical approach was used for initial analysis. It must be noted that this method was conducted under two naïve assumptions: that each word reflects the emotion of the tweet(s) in which it appears, and that the meaning of each word in the tweet was independent of all other words in the tweet. Incorporating the latter assumption in this approach allows for the use of unigrams as features. With these assumptions in place the vocabulary, which shall be called V , was built from all words that appear in the

Table 2. Results of lexical-based classification.

Emotion	Accuracy	Precision	Recall	F1
Angry	89.44	78.73	76.55	77.51
Disgust	84.83	70.95	76.99	73.18
Fear	90.07	80.58	75.99	77.96
Guilt	87.48	75.05	79.89	77.05
Joy	89.70	78.26	82.47	80.12
Sad	88.91	77.71	76.90	77.21
Surprise	84.66	65.66	58.17	59.83
Multi-class	57.78	57.91	57.80	56.81

Table 3. Confusion matrix for lexical-based classification.

Emotion	Anger	Disgust	Fear	Guilt	Joy	Sadness	Surprise
Anger	444	109	25	63	44	29	34
Disgust	51	491	34	68	28	23	47
Fear	31	108	440	56	55	31	41
Guilt	19	75	18	502	42	30	38
Joy	24	55	23	44	577	43	30
Sadness	26	100	29	76	44	444	38
Surprise	68	133	55	136	104	50	175

ETTD. This vocabulary was chosen over the set of English words due to the frequent appearances of unique, non-English words in online communications. The lexicon L was built as a $V \times E$ matrix where the value $L_{i,j}$ is the weighted correspondence of unigram v_i to emotion e_j . That is, each unigram was represented by a seven-element weight vector that consisted of weighted association values for each of the seven basic emotions. These weight values, as described in the equation below, were sums of the quotients of the frequency of the specific unigram in each tweet and the length of that tweet. For example, for the sentences '*I was lost, but now I am found*' with label *joy* and '*I feel so lost in this world*' with label *sadness*, then the weight vector for the word *lost* would be $\{0, 0, 0, 0, 0.125, 0.143, 0\}$ where indices five and six are corresponding to *joy* and *sadness* respectively.

$$L_{i,j} = \sum_{k=1}^{T_{e_j}} \frac{\# \text{ of appearances of } V}{N_k}$$

Equation (1). Correlation score of unigram V to emotion e_j , where T_{e_j} is the number of tweets labeled with emotion e_j and N_k is the length of tweet k .

Two methods were used to classify tweets in the testing set: one-vs-all and multi-class classification. For each method the output was a single emotion label per inputted tweet. These outputs were determined by calculating the overall correlation scores of an inputted tweet to each of the seven emotions. That is, for each tweet, the scored correlations of each appearing word were summed to produce the overall correlation score for each emotion. For testing one-vs-all classifications, the output was positive if the emotion being tested for had the strongest correlation to the current tweet, and negative if the emotion being tested for did not have the strongest correlation. Alternatively, the output of a multi-class classification trial was the emotion that had the strongest correlation to the tweet being tested. The generated label for each tweet was then compared to the label given to the tweet during data collection and marked as correct if the labels were equivalent. Tables 2 and 3 indicate the precision, recall, and F1 measure percentage values for each emotion (for which a binary classification was used), and the emotion set (for multi-class classification), as well as the confusion matrix in the case of multi-class classification. All values were results of five independent trials. For each trial, the ETTD was shuffled to preserve generality and avoid overfitting on a single set of training data.

While the lexical approach is accurate and inexpensive, there are still drawbacks with such an approach. The first of these drawbacks is the availability of external lexicons due to the model of basic emotions used in this study. To the best knowledge, the ETTD is the only existing dataset for this specific set of emotions. The ETTD was also constructed specifically for this work, which may hinder its effectiveness when used for other works. The other major issue with this approach is the semantics of the English language, through which many words can take a different meaning based on their context. Inflection, tone, tempo, and stress placement within a statement can alter the interpreted meaning of the statement. As an example, the statement '*I would love to have your parents over for dinner.*' can be read in two ways. The literal meaning can be interpreted as joy. Alternatively,

Table 4. Confusion matrix for voting classifier model.

Emotion	Anger	Disgust	Fear	Sadness	Guilt	Joy	Surprise
Anger	340	95	40	19	32	202	19
Disgust	86	452	24	21	12	139	15
Fear	54	64	476	15	22	119	14
Sadness	64	41	27	460	18	116	9
Guilt	54	25	18	11	486	111	14
Joy	132	77	65	29	54	389	17
Surprise	58	44	28	13	20	102	508

added sarcasm creates a different tone, one which creates a disgusted opposition to the idea. The absence of such inflection in the reading of these tweets hinders the ability to determine the correct emotional tone of a tweet.

Learning-based approach

As a comparison, several supervised machine learning algorithms were employed. These algorithms were used to automatically extract patterns from training data and make educated guesses according to such patterns for a set of test data. Such algorithms have a proven to be extremely useful in data science and sentiment analysis applications. The challenge of this approach was modifying the use of such algorithms, which are used primarily for binary classification in sentiment analysis, to create an effective tool for multi-class classification. Given the capabilities of existing multi-class classifiers, a set of classifiers was trained based on the following methods: Bernoulli Naïve Bayes, Logistic Regression, Linear Support Vector, Multilayer Preceptron, Linear Discriminant Analysis, Nearest Centroid, Quadratic Discriminant Analysis, Radius Neighbors, Random Forest, Decision Tree, Extra Tree, and Ridge Regression [16]. These classifiers were chosen strictly for the task of multi-class classification.

The trained classifiers were then collected into an ensemble like that of [7], in which each classifier casted a single ‘vote’ which was collected and the most popular ‘vote’ across the set of classifiers was selected as the tag of a given tweet. This tag was accompanied by a confidence value, calculated as the count of votes in agreement with the final tag divided by the total number of votes. As with the previous lexical approach, five independent trials were conducted, for which the data was randomly shuffled to prevent overfitting on a single training set. The results, coupled with the confusion matrix in table 4, are an average of the independent trials ($A = 59.26\%$, $P = 63.53$, $R = 59.29$, $F1 = 60.71$).

While this voting classifier outperformed the previous lexical classification method, table 4 provides evidence that improvements may be made. For instance, the supervised-learning method favored the *joy* tag, and generally mislabeled tweets of other emotions as *joy* tweets.

Classification on other datasets

Although the previous methods showed promising results on the ETTD, the generality and performance of these methods compared to existing solutions came into question. These factors were evaluated in an experiment using the dataset created by Wang *et al* [10] and the CBET dataset [5].

The dataset created by Wang *et al* [10] contained an estimated 2.5 million tweets labeled with seven different emotion categories (joy, sadness, anger, love, fear, thankfulness, surprise) taken between November and December of 2011. Due to the unavailability of many tweets in this dataset, the analysis of this dataset was limited. The classification methods Wang used on this dataset were LIBLINEAR and Multinomial Naïve Bayes, the second of which was included in the set of classifiers from this study. With a unigram-based feature set, these classifiers achieved accuracy scores of 60.31% and 57.75% respectively, with a maximum precision of 62% and F1-score of over 64%.

The methods described earlier were applied to the new dataset in the exact same fashion as they were used previously. Over five independent trials on the lexical classifier, average values presented in table 5 ($A = 52.79\%$, $P = 32.24$, $R = 33.48$, $F1 = 32.47$) were attained. Conducting five trials with the ensemble classifier produced average values as shown in the right-hand side of table 5 ($A = 55.43\%$, $P = 40.58$, $R = 36.36$, $F1 = 36.15$).

At the time of this study, the most modern publicly available dataset related to emotion classification was the CBET dataset [5]. This dataset, containing 76,860 tagged tweets, was created with an emotion set differing from the set used in this study only with the inclusion of *love* and *thankfulness* as emotion classes. The results of binary SVM classification on the CBET dataset after five independent trials were a precision score of 45.01%, recall score of 42.26%, and F1 score of 43.59%. Again, five independent trials of the lexical-based method described in this work were conducted on the common emotion classes between CBET and ETTD (*Anger*, *Fear*, *Guilt*, *Sadness*, *Disgust*, *Joy*, and *Surprise*). From this, the values ($A = 35.57\%$, $P = 42.5\%$, $R = 34.65\%$, $F1 = 35.1\%$)

Table 5. Statistical results for testing on external datasets.

	Lexical classification			Supervised learning classification	
	CBET	Wang		CBET	Wang
Accuracy	35.57	52.79	Accuracy	45.67	55.43
Precision	42.5	32.24	Precision	48.82	40.58
Recall	34.65	33.48	Recall	45.09	36.36
F1	35.1	32.47	F1	46.03	36.15

Table 6. Confusion matrix for lexical classification on the CBET dataset.

Emotion	Anger	Disgust	Fear	Guilt	Joy	Sadness	Surprise
Anger	286	173	81	762	183	105	131
Disgust	90	598	57	839	64	115	80
Fear	51	114	603	502	127	89	92
Guilt	55	147	90	1420	255	103	199
Joy	30	41	40	515	629	40	202
Sadness	77	138	46	960	138	260	117
Surprise	60	56	57	565	177	65	614

Table 7. Confusion matrix for supervised-learning classification on the CBET dataset.

Emotion	Anger	Disgust	Fear	Guilt	Joy	Sadness	Surprise
Anger	345	147	45	253	54	79	61
Disgust	66	546	27	227	23	96	48
Fear	44	79	488	155	28	51	42
Guilt	84	129	48	758	118	109	95
Joy	32	38	28	275	327	43	110
Sadness	72	154	42	321	44	292	64
Surprise	56	54	32	219	58	53	441

were attained. Next, an ensemble classifier was trained and tested on the same ratio of the dataset as used in this work (75% for training to 25% for testing). Again, five independent trials were conducted, from which the values ($A = 45.67\%$, $P = 48.82$, $R = 45.09$, $F1 = 46.03$) were attained. Due to the similar construction of the CBET dataset and ETDD, the similar results were as was expected from these trials. Also, due to the similarity of these datasets, confusion matrices were generated for this experiment, as shown in tables 6 and 7.

Note that the confusion matrices in tables 6 and 7 show similar behavior, made evident by the amount of tweets tagged with *guilt*. Note that tweets from every other emotional class were mistaken for a *guilt* tweet, with the largest overlap being with *anger*, *joy*, and *sadness*. Considering the ongoing debate over the true set of basic emotions, such behaviors as this may assist in the discovery of the distinctions among basic and complex emotions. Also, such behaviors can also provide insight on the composition of potentially complex emotions (e.g. *guilt*).

Discussion

Limitations and future work

Despite promising results from the trials in this work, any emotion mining solution is subject to limitations. Within an emotional piece of text, there exists segments that were hard to distinguish from the others using lexical and learning-based analysis. As an example, sentence structures that contain sarcasm or question were difficult to distinguish due to inability of text-focused systems to analyze inflection in a piece of text. Also, as mentioned in the lexical analysis section of this work, specific assumptions were made to reduce the complexity of the task. Such assumptions, despite improving results, may limit the real-world usability of the solution described in this work. Furthermore, these algorithms rely heavily on the conventions of the English language for lexical analysis, a limitation that restricts the use of text in any other language in this study. Future intents for this research address such limitations with intent of internationalization.

In future iterations, the challenge of recognizing alternate interpretations of a piece of text, whether due to sarcasm or some other cause shall be addressed. Adjustments corresponding to the methods referenced by S Dhawan *et al* [13] will be implemented to complete such a task. Besides these process improvements, the first task moving forward is to expand and improve the ETTD. In a plot created by [10] it is clear that, at least in the case of LIBLINEAR and Multinomial Naïve Bayes classification, a training set of 250,000 tweets is optimal for emotion mining, with only slight improvements of accuracy in significantly larger training sets. Thus, the intent is to expand the ETTD to accommodate such training sets while maintaining balance among the seven basic emotions. Like [5], the use of emoticons in the classification process shall be implemented, with hopes of improving the reliability of classification methods when emoticons are present.

In addition to these improvements, a user interface component shall be implemented to allow review by healthcare or public health providers. This interface will be used to generate a geographic heat map (as a snapshot or in real time) that correlates tagged tweets to their location (provided such information is available). The motivation behind this task is that such a representation will assist in serving the intended purpose of this research, which is to assist mental health professionals and local governments in identifying areas with high concentrations of tweets that reflect negative emotions (*guilt* and *sadness* in particular). Further analysis of this information shall allow local governments to pinpoint the relationship of these emotions to issues in specific communities, local legislation, and other events.

Conclusion

Determining the emotion in a piece of social media text is critical since it is commonplace for a large portion of the global population to communicate their thoughts and feelings on such media in response to daily events. The concept of emotion mining, though young, is a fascinating challenge which incorporates a plethora of professional disciplines beyond computer science. This study addressed this challenge with the ETTD, a corpus of 24,500 tweets labeled with one of seven emotional tags: *joy*, *fear*, *surprise*, *anger*, *sadness*, *disgust*, and *guilt*. Next, a lexical approach, in which we use weighted word-emotion association vectors, was tested as a preliminary attempt to accurately and reliably classify tweets. After this, a supervised machine learning-based approach to the problem was assessed. The algorithms used in this approach were trained on 75% of the ETTD to extract patterns from the training data and then use such patterns to classify new tweets. Each of the methods showed promising results on the ETTD and achieved similar results when applied to the CBET dataset and the dataset created by [14]. Significant overlap of *guilt* with *sadness*, *anger*, and *disgust* is shown in the confusion matrices produced by this work, which may suggest that *guilt* is in fact a combination of such emotions.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Conflicts of interest

None declared.

ORCID iDs

Suboh M Alkhushayni  <https://orcid.org/0000-0002-3191-1275>

References

- [1] Shahraki A G 2015 Emotion mining from text *Master Thesis, Dept. Com. Sci.* University of Alberta (Edmonton, AB, Canada)
- [2] Burton N 2016 What are basic emotions? *Psychology Today* 07-Jan-2016 [Online] Available [Accessed: 16-Oct-2020] (<https://www.psychologytoday.com/us/blog/hide-and-seek/201601/what-are-basic-emotions>)
- [3] Ekman P, Friesen W V and Ellsworth P 1972 *Emotion in the Human Face: Guidelines for Research and an Integration of Findings.* (Oxford: Pergamon)
- [4] Ortony A and Turner T 1990 What's basic about basic emotions? *Psychological Review* **97** 315–31
- [5] Shahraki A G and Zäiane O R 2017 Lexical and learning-based emotion mining from text. In *Proceedings of the International Conference on Computational Linguistics and Intelligent Text Processing* (Vol. 9, pp. 24–55). Chicago
- [6] Izard C E 1991 *The Psychology of Emotions.* (New York City : Spring Science & Business Media)
- [7] Kinsley H 2020 Combining Algorithms with NLTK retrieved from (<https://pythonprogramming.net/combine-classifier-algorithms-nltk-tutorial/>) [Accessed: 16-Oct-2020]
- [8] Jabreel M and Moreno A 2019 A deep learning-based approach for multi-label emotion classification in tweets *Applied Sciences* **9** 1123

- [9] Jianqiang Z, Xiaolin G and Xuejun Z 2018 Deep convolution neural networks for twitter sentiment analysis *IEEE Access* **6** 23253–60
- [10] Wang W, Chen L, Thirunarayan K and Sheth A P 2012 Harnessing twitter 'big data' for automatic emotion identification 2012 *Int. Conf. on Social Privacy, Security, Risk and Trust (PASSAT) and 2012 Int. Conf. on Social Computing (SocialCom)* pp 587–92 (Piscataway, NJ) (IEEE)
- [11] Mohammad S M 2012 #emotional tweets *Proc. of the First Joint Conf. on Lexical and Computational Semantics (*SEM)* pp 246–55 (Association for Computational Linguistics)
- [12] Hasan M, Agu E and Rudensteiner E 2014 Using hashtags as labels for supervised learning of emotions in twitter messages *Health Informatics Workshop (HI-KDD)*
- [13] Dhawan S, Singh K and Sehrawat D 2014 Emotion mining techniques in social networking sites *International Journal of Information & Computation Technology* **4** 1145–53
- [14] Lövhelm H 2012 A new three-dimensional model for emotions and monoamine neurotransmitters *Medical Hypothesis* **78** 341–8
- [15] Shaver P, Schwartz J, Kirson D and O'connor C 1987 Emotion knowledge: further exploration of a prototype approach *Journal of Personality and Social Psychology* **52** 1061
- [16] Pedregosa F et al 2018 Scikit-learn: Machine Learning in Python [online] Jmlr.csail.mit.edu. Available at: [Accessed 16 - Oct. 2020] (<http://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html>)
- [17] Ravindra 2017 Emotion-Classification-of-Twitter-Dataset (<https://github.com/ravindra-s/Emotion-Classification-of-Twitter-Dataset>)