



Crowdsourcing a More Realistic Emotional Lexicon Process

Robert Ball^(✉), Joshua Jensen, and Samuel Romine

Weber State University, Ogden, UT 84408, USA
{robertball, joshuajensen1}@weber.edu

Abstract. Even though sentiment analysis and emotion analysis are relatively new subfields of natural language processing, they are increasingly being used by businesses and academics. Their primary objective is to automatically capture the emotions and opinions of people as they wrote their text. Whether the text is from social media, literature, or any other form of text, it is a difficult problem. We present a novel approach by introducing the idea of capturing a range of universal primary emotions (i.e., sadness, happiness, anger, disgust, surprise, and fear). By crowdsourcing the emotion lexicon, it creates a diverse lexicon of emotions. This type of lexicon shows how many people within the same culture and language can have both convergent emotions where everyone universally every agrees that a word has a single emotion behind it (e.g., yummy = happy) and divergent emotions where there is a range of feelings for the same word (e.g., hated = angry, disgusted, and sad). Finally, we empirically prove that emotions are not binary as is often used by modern approaches to this problem (e.g., not sad \neq happy), showing that people's emotions are complex and nuanced.

Keywords: Sentiment analysis · emotion analysis · natural language processing · emotion · sentiment · opinion mining

1 Introduction

Sentiment analysis is a subset of natural language processing (NLP) that analyzes sentiment, emotion, opinion, and attitudes in text [16]. This subfield of NLP is relatively new and can be traced back to the beginning of the millennium. Sentiment analysis is primarily concerned with polarity, how positive or negative a set of text is [10]. Emotion analysis has the same goal as sentiment analysis but is a newer term that is reserved for analysis that strictly analyzes text for emotions (e.g., sad, happy, etc.) [13].

Sentiment analysis involves various areas of data science, human-computer interaction, and NLP for the purposes of decision making, finding opinions, getting the “feel” of a situation, web scraping, data mining, predicting stock market fluctuations, understanding how much people like or dislike a restaurant, etc. Many modern business decisions are made based on sentiment analysis [8, 10, 11].

However, the previous body of knowledge on this topic makes one faulty assumption – that the relationship between emotion and opinion is consistent across people

and culture. This assumption is false as explained by Jackson et al., that “emotions can vary systematically in their meaning and experience across culture and language” [9]. In other words, not everyone shares the same opinions nor backgrounds. This creates a fundamental flaw in the current approach to emotion analysis. We propose a new way of doing sentiment analysis and emotion analysis. We account for the realities where people view the world and express themselves differently from one another because individual people have a diversity of thought and emotion from one another that stems from their unique individual differences and experiences in life.

Due to individuality, even with people that agree with the same opinion, they might use different language. Because people are inherently diverse in their backgrounds, word choices, cultures, etc. the same word may evoke different emotions for different people.

2 Related Works

Many scientists through the centuries noted that emotions are directly tied to changes in facial muscles [3]. Perhaps the most influential work surrounding Facial Feedback Theory is Paul Ekman’s work where he identified six universal emotions: anger, disgust, fear, happiness, sadness, and surprise. Ekman traveled the world both taking and presenting images of people at different emotional states. He found that in both literate and illiterate societies (where there was no outside exposure of media to confound the study) that people were able to recognize the six universal emotions through facial expressions [15]. Ekman later postulated that there might be more universal emotions than just six that exist [4]. Indeed, work by other scientists found this to be true [1, 2].

Exactly how many emotions are there? Robert Plutchik expanded on Ekman’s theory with what is referred to as the wheel of emotions. First, Plutchik added two emotions to Ekman’s six: trust and anticipation. Also, each of Plutchik’s eight emotions can have a range of strength or intensity. For example, anger can be a range from simple annoyance to extreme rage. In addition, Plutchik suggests that by combining emotions that you get new emotions. [14].

In 2019 Jackson et al. show that aside from the fact that there are primary emotions, “[t]here is a growing recognition, however, that emotions can vary systematically in their meaning and experience across culture and language.” In addition, they found that in surveying the emotion theory and linguistics literature that studies that had imposed training phases and forced choice paradigms found more evidence of universal recognition of emotion, but studies that had fewer constraints had more cultural variability [9].

In other words, Jackson et al. found that emotions and words are not universal but have a great deal of cultural and contextual variability. The idea is like what many children ask: “When you see red, do you see the same red as I do?” For emotions, the question would be the following: “When you feel sad, do you feel that same sad that I feel?” The answer is no, which leads to the conclusion that sentiment and emotion lexicons should not have only one sentiment or emotion associate with just one word.

In 2004 Hu and Liu published a seminal work on sentiment analysis. Since then, sentiment analysis is typically performed by comparing a corpus (a set of text of any length) to sets of established words in a lexicon. Hu and Liu’s original algorithm counts

all the words in the text that are in the positive set and all the negative words that are in the negative set. The remaining words that are not in the positive or negative sets are then considered neutral. If there are more positive words, then the text is considered positive. The same is true for negative and neutral [7].

In 2010 Mohammad and Turney proposed the idea of using emotions instead of polarization (positive/negative) to create a richer lexicon [12]. They later further fleshed out their idea [13]. Their disruptive concept is twofold: First, use crowdsourcing to create the lexicon and second, determining the intensity of each of the emotions for the word along with the typical polarization (e.g., positive/negative/neutral) to words. Their lexicon is called EmoLex. EmoLex is a good step forward, however, recent research from Jackson et al. shows that the more constrained a study with emotion the less varied the result [9].

Our work is heavily based on Mohammad and Turney’s work with EmoLex. We will show that the idea of crowdsourcing for emotion is arguably one of the best approaches to getting representative emotions. We also show while sentiment analysis is quite useful that emotion analysis also has a place for richer insight into text.

3 Creating a More Realistic Emotion Lexicon

We created an emotion lexicon that was crowdsourced like EmoLex but with the result that shows the diversity of primary emotions for a given subject.

We got our words to add to the lexicon from two sources: First, we used 3,000 books from Project Gutenberg, a repository of tens of thousands of books in the open domain (either not copyrighted or the copyright has expired). Second, we web scraped over 4.5 million reviews from AllRecipes.com.

Using the NLTK collocations package and the PMI score we used the highest scored adjectives and adverbs unigrams from the Gutenberg books and recipe reviews. We did the same for the highest scored trigrams and quad grams for the recipe reviews to test if phrases would have the same results.

We then set up three surveys that we called “quizzes” for the public. The surveys were loosely patterned after BuzzFeed.com quizzes where participants enjoy finding out how much they know or do not about popular culture facts. We used licensed Pixar *Inside Out* images to appeal to popular culture and entice more people to take our surveys.

We had approximately 35 words or phrases per survey. Our intention was not to create a full emotion lexicon, but to test our theories. Figure 1 shows a screenshot of the choice that volunteers were presented with when they took the survey. The “Food Words” survey was the top unigrams from the AllRecipes.com reviews. The “Words from books” survey was the top unigrams from the Gutenberg books. The “Cooking Expressions” survey was the top trigrams and quadgrams from the AllRecipes.com reviews.

We also implemented Google’s free reCAPTCHA on our website to ensure results were from actual humans. reCAPTCHA is a free service from Google that tests if the current user is a human or not. It helps prevent automatic software from filling out our surveys.

As an added security check, we recorded how long it took volunteers to take the surveys and how long it took assign an emotion to each word. In other words, if someone

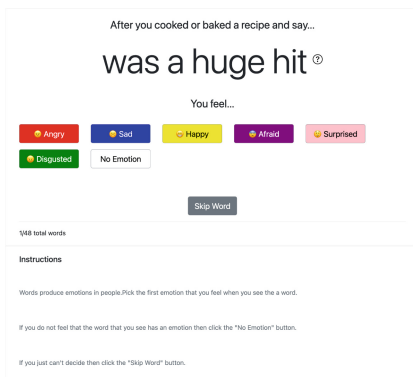


Fig. 1. Screenshot of the “Cooking Expressions” survey. Note that the emotion buttons were presented in a different order for each volunteer. That order was consistent for the duration of the quiz for that person.

simply clicked as fast as they could without reading the words then we can detect that and disregard those answers.

In addition, to have a good experimental design we randomized the word order for each survey so that each volunteer was given the words in a random order. We also randomized the emotion buttons so that all emotions were presented to different people in different order. When the quiz began the buttons were randomly put in an order. That order was consistent for the duration of the quiz for that person. There were two exceptions: the “No Emotion” button was always last and the “Skip Word” button was on the bottom.

After the volunteer assigned an emotion to all the words or phrases, the person was presented with an optional demographics page that asked for their age, gender, and racial/ethnic background. After selecting the “submit” button they were presented with a colorful and fun results page. The intent of the results page was to share it with their friends who might then take the survey themselves.

4 Results

In this section we present several of our findings. We present examples from our emotion lexicon, show that words do not necessarily have an emotional opposite, and discuss crowdsourcing for emotion lexicon creation.

4.1 Example Words and Their Emotions

We created a proof-of-concept algorithm patterned after the open-source `polarity_scores` function used with the VADER lexicon in the Sentiment Intensity Analyzer Python NLTK module. Our function takes in any text and reports back the primary emotions associated with it from our emotion lexicon. We do not claim that our proof-of-concept algorithm is as sophisticated as the VADER function in NLTK, but it is used here as a demonstration of using emotions instead of simply using positive, neutral, and negative.

Our function first checks for phrases that match phrases found in our emotion lexicon. For example, if the phrase will make again is found then the functions adds the emotions associated with that phrase. The function then removes all stop words. It then adds all the emotions for unigrams found in the emotion lexicon. Unigrams that are part of phrases that have already been counted are not counted again. Finally, the results are normalized so that they add to 1.0.

Table 1. Results of running the NLTK VADER Sentiment Intensity Analyzer module on All-Recipes.com recipe reviews. This data is provided to show a comparison to our emotion analysis results, which is shown in Table 2.

How many stars the review received	Negative	Neutral	Positive	Compound
1	11%	79%	10%	3
2	7%	80%	13%	24
3	5%	79%	17%	43
4	2%	76%	22%	64
5	2%	71%	27%	72

Table 1 shows the results from the NLTK VADER polarity_scores function on the AllRecipes.com reviews. The reviews were organized from one to five stars with five stars being the most positive reviews. For the 4.5 million reviews that we web scraped, there is not an even distribution of reviews in the dataset, so we randomly selected 90,000 reviews from each star category.

Table 2. Results of running our custom emotion analysis function on the same data from AllRecipes.com as compared to Table 1.

How many stars the review received	Angry	Happy	Disgusted	Afraid	Surprised	Sad
1	19%	20%	19%	2%	4%	36%
2	14%	31%	15%	2%	5%	32%
3	10%	48%	10%	2%	6%	24%
4	3%	81%	3%	1%	6%	6%
5	1%	90%	1%	0%	5%	2%

Considering that the NLTK VADER sentiment analyzer uses a general-purpose lexicon, it did a remarkable job if you consider the trend of negative and positive. There is a clear trend of positive percentage to the number of stars and a clear negative trend to negative percentage to the number of stars. The trend is also clear if you refer to the compound column.

Table 2 shows the results from our custom function using our emotion lexicon. Our results also have positive and negative trends similar to Table 1 except for surprised, which seems to be constant at around 5%.

Interestingly, one can see that happy is very similar to positive from the VADER function. What would be the equivalent of negative? It appears that angry, disgusted, afraid, and sad could all be considered negative. However, Table 2 shows that the different emotions are more nuanced than simply negative and provide a richer and more complete emotion spectrum. This fits the stated goals of emotion analysis much more thoroughly.

How can someone interpret these results? For the VADER results, one possible interpretation is that Table 1 shows how most people would feel in terms of positive and negative sentiment when reading the recipe reviews.

On the other hand, the results from our emotion analysis in Table 2 is that there is a range of emotions that people could feel when reading the lower rated reviews, but for the higher rated reviews there is more of a consensus of a happy feeling. No one person would necessarily feel 19% angry, 20% happy, 19% disgusted, 2% afraid, 4% surprised, and 36% sad when reading the one-star reviews. Our interpretation is that these results are similar to a poll where a wide range of people expressed their feelings.

4.2 Finding that There is no Opposite Emotion

When selecting the words for the surveys we purposely tried to find many words that are opposite in terms of definitions. For example, we used the words *impressive* and *unimpressive* to test how words with opposite meanings correspond to emotions. This is important because modern analysis tools simply negate the resulting emotion when an opposite definition is introduced.

Table 3 shows that most positive words tend to converge on *happy*. On the other hand, negative words have a much larger spectrum of emotion. For instance, even though *awful*, *bad*, and *disappointing* could all be considered negative words, they all differ in the degree of anger, disgust, and sadness. For example, *awful* is 48% anger, 36% disgust, and 16% sad. However, *disappointing* is 26% anger and 74% sad with 0% disgust.

There are many papers that show that sentiment analysis can be improved by using negations such as *not* and *but* (e.g., [5, 6]). The idea is that if a positive word, such as *good* is negated, like *not good*, then the meaning becomes negative. This has been proven to be useful in sentiment analysis many times.

However, we show through our empirical results that simple negation of emotion is insufficient and misleading. Table 3 shows a series of opposite words in sentiment analysis, but that are not automatically opposite in our emotion lexicon. For example, the opposite of *hated* could be *loved*. If *hated* is negative then the opposite, *loved*, should be positive, which it is. However, our results of emotion analysis show that *hated* has an emotion spectrum anger, disgust, and sad, while *loved* has only happy.

One of the reasons that Mohammad and Turney (EmoLex) cited for using Plutchik's emotion theory is that Plutchik's wheel of emotions has opposites [13]. For example, Plutchik's wheel of emotions shows that happy is the opposite of sad.

However, this does not appear to always be the case. For example, the word *loved* is a simplistic word that only has happy associated with it but *hated* is more complex and

Table 3. Example of some words in the emotion lexicon. These examples show a convergence of *happy* with traditionally positive words but show more of an emotion spectrum with traditionally negative words.

	Angry	Hap py	Disgusted	Afrai d	Surprised	Sad
delicious	0%	100%	0%	0%	0%	0%
wonderful	0%	100%	0%	0%	0%	0%
yummy	0%	100%	0%	0%	0%	0%
awful	48%	0%	36%	0%	0%	16%
bad	25%	0%	21%	0%	0%	54%
disappointing	26%	0%	0%	0%	0%	74%

has a range of primary emotions of anger, disgust, and sadness. Looking at the phrases at the bottom of Table 3 shows similar complexities.

We are not prepared to say that Plutchik’s wheel of emotions is wrong in terms of opposites, however, we are prepared to say that given our empirical results that the opposite of one emotion is not always simplistically another one. For example, *not sad* does not always equal *happy* for all people, but it does for some people.

Our findings are that there are not universal emotional opposites for all people. Since people come from a broad range of backgrounds and cultures it shows that emotion theory be necessarily as complicated as the people that exhibit the emotions.

5 Discussion

There are now many sentiment lexicons and emotion lexicons. However, each lexicon has a different purpose. For example, Hu and Liu’s Lexicon is simple and easy to interpret: Is the given text more positive or negative?

Our proposed approach provides a *range* of primary emotions that different people might feel given a text. In this paper we have provided a novel approach to enrich the field of emotion analysis. We advocate crowdsourcing the original authors of the text or similar authors for statistical sampling. We also show that although negation works well with sentiment analysis that negation does not work in the same way with emotion analysis.

Lastly, we have shown that emotion analysis is not simple. Specifically, we have empirically proved that people are different and that they do not all agree on a single

emotion for a single word. Human beings are complex creatures and should be treated thus.

References

1. Cordaro, D.T., Keltner, D., Tshering, S., Wangchuk, D., Flynn, L.M.: The voice conveys emotion in ten globalized cultures and one remote village in Bhutan. *Emotion* **16**(1), 117–128 (2016)
2. Cordaro, D.T., Sun, R., Keltner, D., Kamble, S., Huddar, N., McNeil, G.: Universals and cultural variations in 22 emotional expressions across five cultures. *Emotion* **18**(1), 75–93 (2018)
3. Davis, J.I., Senghas, A., Ochsner, K.N.: How does facial feedback modulate emotional experience? *J. Res. Pers.* **43**(5), 822–829 (2009)
4. Ekman, P., Cordaro, D.: What is meant by calling emotions basic? *Emot. Rev.* **3**(4), 364–370 (2011)
5. Hatzivassiloglou, V., McKeown, K.: Predicting the semantic orientation of adjectives. In: *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-1997)* (1997)
6. Hiroshi, K., Nasukawa, T.: Fully automatic lexicon expansion for domain-oriented sentiment analysis. In: *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2006)* (2006)
7. Hu, M., Liu, B.: Mining and summarizing customer reviews. In *Proceeding of SIGKDD KDM-04* (2004)
8. Hutto, C.J., Gilbert, E.E.: VADER: a parsimonious rule-based model for sentiment analysis of social media text. In: *Eighth International Conference on Weblogs and Social Media (ICWSM 2014)*, Ann Arbor, MI, June 2014 (2014)
9. Jackson, J.C., et al.: Emotion semantics show both cultural variation and universal structure. *Science* **366**(6472), 1517–1522 (2019)
10. Liu, B.: *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions (Studies in Natural Language Processing)*, 2nd edn. Cambridge University Press, Cambridge (2020)
11. Liu, B.: Sentiment analysis and subjectivity. In: *Handbook of Natural Language Processing*, vol. 2, pp. 627–666 (2010)
12. Mohammad, S., Turney, P.D.: 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*, Los Angeles, CA, pp. 26–34 (2010)
13. Mohammad, S., Turney, P.D.: Crowdsourcing a word–emotion association lexicon. *Comput. Intell.* **29**(3), 436–465 (2013)
14. Plutchik, R.: *Emotions in the practice of psychotherapy: clinical implications of affect theories*. American Psychological Association, Washington, DC (2000)
15. Shiota, M.N.: Ekman’s theory of basic emotions. In: Miller, H.L. (ed.) *The Sage Encyclopedia of Theory in Psychology*, pp. 248–250. Sage Publications, Thousand Oaks (2016)
16. Liu, B.: Sentiment analysis and opinion mining. *Synth. Lect. Hum. Lang. Technol.* **5**(1), 1–167 (2012)