Emotion Classification of App Reviews - Guidelines for Annotation

Reading time: 10 minutes

Introduction

Emotion classification is the automatic extraction or identification of a particular emotion from a text document [1]. The goal is to assign a specific tag from a given set of predefined emotional labels (e.g., joy, sadness, anger, disgust) to text segments based on their emotional content. These labels are typically defined by a taxonomy of emotions, used to shape the limits and constraints of each emotion and facilitate the annotation process.

App reviews are a common application area for emotion classification, where user feedback is analysed to understand their emotional responses to different features or aspects of an app. By categorising these emotions, developers and companies can gain valuable insights into user experience [2], identify potential areas for improvement [3], and enhance overall user experience [4]. However, research in app review opinion mining is mostly devoted to binary polarity assessment (i.e., positive vs. negative). Multiclass emotion detection has undergone intense study in areas such as customer reviews for products and services [5], microblogging platforms [6] and artistic-related content like movies [7] and books [8]. However, emotion analysis in app reviews is very limited, and there are no public benchmark datasets that can be used for the evaluation of automated approaches.
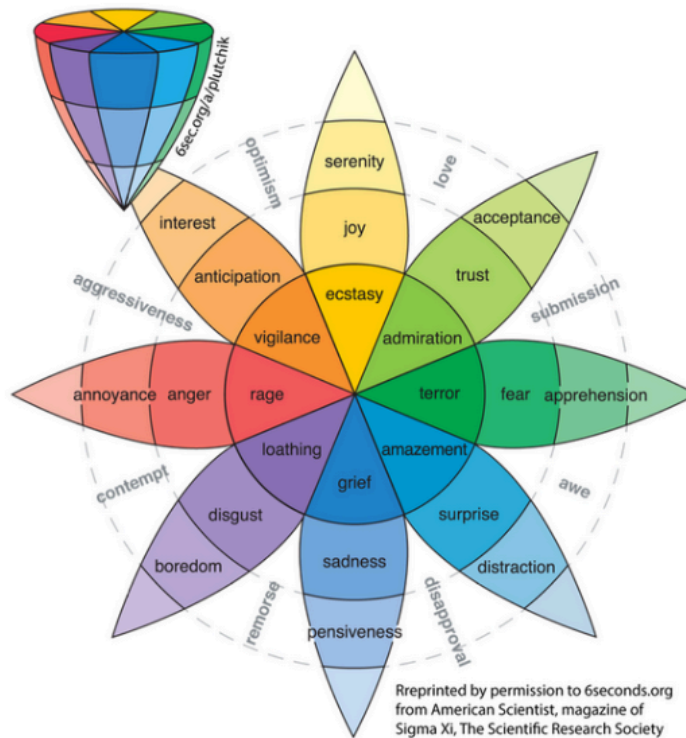
This document describes the **guidelines for annotating a dataset of mobile app reviews using a multiclass emotion taxonomy**. The results are intended to serve as a benchmark dataset for automated NLP emotion classification tasks.

Plutchik's Wheel of Emotions

The *Plutchik's Wheel of Emotions* is a model developed by psychologist Robert Plutchik to identify emotions and illustrate the relationships between them [9,10]. It organises emotions into a wheel, with eight primary emotions: *joy*, *trust*, *fear*, *surprise*, *sadness*, *disgust*, *anger*, and *anticipation*. These emotions are arranged in pairs of opposites and are displayed in varying intensities, with more intense emotions at the centre and less intense emotions at the periphery. The wheel visually represents how emotions relate to each other.

PLUTCHIK'S WHEEL OF EMOTIONS
NAME A FEELING & ENHANCE EMOTIONAL LITERACY

Rreprinted by permission to 6seconds.org
from American Scientist, magazine of
Sigma Xi, The Scientific Research Society

For the annotation process, we will use the set of 8 primary emotions defined by Plutchik's. This decision is motivated by the following reasons:

- The literature review highlights Plutchik's primary set of emotions as the most frequently used for app review multi-class emotion analysis.

- Plutchik's Wheel of Emotions identifies opposite pairs, making it easier to compare and apply this emotion model to app reviews.

- Plutchik's Wheel of Emotions models contiguous emotions, enabling the identification and discussion of app reviews that might fall between two emotions.
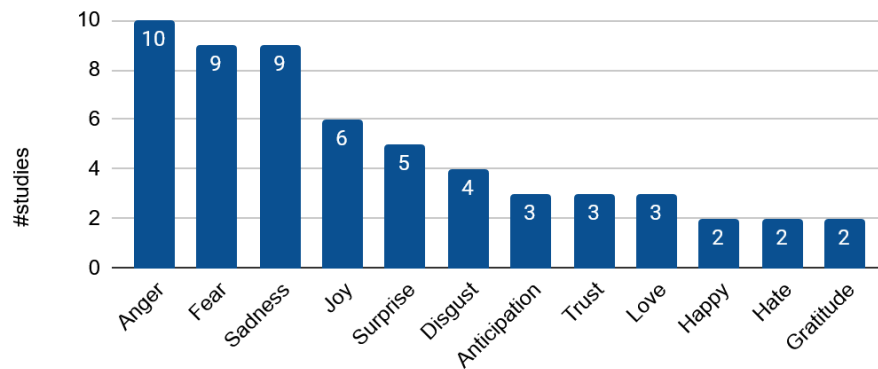
According to this taxonomy, emotions are defined as follows:

- **Joy:** A feeling of great pleasure and happiness.

- **Trust:** A feeling of confidence in or reliance on someone or something.

- **Fear:** An unpleasant emotion caused by the belief that someone or something is dangerous.

- **Surprise:** A feeling of mild astonishment or shock caused by something unexpected.

- **Sadness:** A feeling of sorrow or unhappiness.

- **Disgust:** A strong feeling of aversion or repulsion.

- **Anger:** A strong feeling of annoyance, displeasure, or hostility.

- **Anticipation:** A feeling of excitement about something that is going to happen.

Applying Plutchik's Wheel of Emotions to App Reviews

In this section, we discuss and illustrate with examples how the aforementioned emotions might apply to different sentences belonging to app reviews.

Joy

Joy expresses **excitement** or **pleasure** towards an app, a feature, a release or a user experience in general. It implies a sense of **possibility** and **positiveness.**

Joy reviews express positive aspects about an app:

[Ex1] Minimalistic, but useful app without ads.

Users also express *Joy* when they mention what they like about the app, such as features or characteristics:

[Ex2] It's good and great and it's has a meal planner it's quite helpful.

[Ex3] Excellent task and list management tool

Joy can also be linked to positive user experience:

[Ex4] Good exercise and good food suggestions morning exercise give a special energy😊

Trust

Trust expresses **acceptance** and **connection** from the user to the app. It implies a sense of **safeness** and **warmness**, illustrating explicitly the user's willingness to use the app.

Users express *Trust* when they explicitly mention **their personal experience** with the app **in a personal way**, reporting how the app satisfies their goals or needs:

[Ex5] It has really made a big difference in my health and wellness.

[Ex6] One of the best note-taking I've tried so far.

[Ex7] To learn GPS navigation is my second favorite option

Trust can also be about the user **relying on the app to conduct a specific task** according to its expectations:

[Ex8] I set up a new, independent account on MY email server specifically for this app to connect to, to backup my phone SMS/call logs, etc.

[Ex9] My favourite features are recurring tasks - brilliant for implementing new habits and calendar where I can schedule future to do things on specific days.

Trust is also expressed when the user **relies on the app over other alternatives in the market**:

[Ex10] By far the best app I've used for keeping track of macros.

[Ex11] Although I use My Diary as a journal, & RPG Notes for TTRPG notes, no app I've used beats this one when a simple & efficient note pad is needed to jot ideas & make to-do lists.

Furthermore, *Trust* is also related to **positive praises of security and privacy concerns**:

[Ex12] It has advanced privacy protection

Joy and *Trust* can often be confused, as they are contiguous emotions according to Plutchnik's Wheel. In this context, when in doubt, we will apply the following distinction:

- If the user makes **no explicit mention of their own experience**, limiting the feedback to positive aspects and likes, with no involvement of their expectations or their use of the application, then we will consider the review as *Joy*.
- If the user makes explicit praise in the review about their **own user experience** with the application from a **personal point of view**, including the satisfaction of their goals and needs, then we will consider the review as *Trust*.

For instance, in the following *Joy* examples:

[Ex13] It's very intuitive, and lightweight which makes it a must have for list lovers.

[Ex14] I like the size options and you can control transparency

[Ex15] Does exactly what it needs to in an elegant way and doesnt force monetization or clunky features on you.

The author of the review implies positive aspects of the app that they enjoy. However, there is no information about the suitability and acceptance of the app with respect to the personal goals and needs of the user.

However, in the following *Trust* examples:

[Ex16] Best habit tracker I've ever used, free, no ads, data can be exported, overall great.

[Ex17] Instagram is my app to browse or watch videos and more

[Ex18] Use it as my Diary, Really handy 😃 👍.

The author explicitly makes reference to their use of the app, stating their acceptance and reliability and emphasizing positive aspects reinforcing this use.

Fear

Fear expresses **stress** or **anxiety** towards an app or a particular event. It implies a sense of **agitation**, expressing that the user is **scared** about something.

Fear can refer to users **complaining or reporting** an incident that **affects their trust** for an app:

[Ex19] People are adding fake members via python by scrapping to make their group look huge.

Fear also implies users **asking for help**, **showing frustration or anxiety** towards a situation they cannot control:

[Ex20] I having Google Pixel 3 phone, but call recording option is not available on this app please help me.

[Ex21] Facing fast gifs speed problem and gif folder limit reach . Anyone please help

Surprise

Surprise expresses **shocked** or **unexpected** emotions towards an app. It implies a sense of **unpredictability** related to something new, altering user expectations (either positively, negatively or neutral).

Surprise can refer to **issues encountered during the use of the app** that **cannot qualify as bugs** but **did not suit user expectations**:

[Ex22] wanted to go to an event later in the month showed only 3 attendees but rsvp was closed not showing canceled?

[Ex23] Another weird issue is I got a transaction SMS from the bank that had the words "For lost/stolen cards call xxxx" and Truecaller gave a notification with the title "Stolen Card" and the last 4 digits of my card which was weird

Surprise can also refer to **events or situations** the user encounters that do **not suit their judgment**:

[Ex24] 3 stars for asking me to rate a planner/task app I installed 1hr ago so I don't know how well it works for a few days minimum before asking for review.

Sadness

Sadness expresses **disappointment** or a **sense of loss** towards the expectations of the user with respect to the app. It implies a sense of **heaviness** and **severeness**, implying a decrease in the average user experience.

Sadness refers to negative statements about the user experience with respect to user expectations:

> **[Ex25]** *It makes my workflow almost unusable :(*

Sadness also refers **to bug reports or faults**:

> **[Ex26]** *Just that RCS/Chat function has regular failures unfortunately.*

> **[Ex27]** *Stories are replaying and starting over again and again.*

Sadness can also refer to **new releases causing disruption in the traditional activity** of the user:

> **[Ex28]** *Latest update blocks ability to choose a bus-only routing in London without moving to the premium version.*

> **[Ex29]** *Great app but sadly you can't connect my fitness pal and no food/calorie/macro tracker in app 🥱 only thing it's missing.*

Disgust

Disgust expresses a **lack of trust** and **rejection** towards an app. It implies a sense of **bitterness** and **refusal**, implying a judgment from the user that something in the app is wrong, unsafe or not compliant with rules.

Disgust is clearly expressed when the user claims that they refuse to use the app:

> **[Ex30]** *I find this app frustrating as it has stopped working and not for the first time , today I was on week 7 workout 1 and it stopped working on my 25 minute run so I had to try and gauge myself when to stop, will be using a running watch from now on*

Disgust can also be expressed by stating that an app is not suited for purpose, either with sarcasm or literally:

> **[Ex31]** *Might as well just call it "Weather Radar" and ditch the widget part of the name because its non-functional.*

> **[Ex32]** *It is unnecessary for an athletic networking site to wade into politics.*

Sadness and *Disgust* can often be confused, as they are contiguous emotions according to Plutchnik's Wheel. In this context, when in doubt, we will apply the following distinction:

- If the focus of the review is on **disappointment concerning user expectations** in terms of functional correctness, usability, efficiency and other quality aspects, then we will consider the review as *Sadness*.

- If the focus of the review is on **the explicit rejection and distrust of the user for the app** or a set of its features, then we will consider the review as *Disgust*.

For instance, in the following *Sadness* examples:

[Ex33] Alarm routine won't start gargle podcasts

[Ex34] My days aren't all exactly the same so today I had a zoom meeting and couldn't access to it from my phone and had to restart it.

The user is clearly reporting a disappointment comparing the performance of the app with respect to their expectations, without expressing any additional or more grave judgment about the acceptance of the app by the user.

However, in the following *Disgust* examples:

[Ex35] Until you change the privacy settings, I won't update

[Ex36] The only thing you get out of the app are bots, spamming you and trying to get you to a paid site.

The user is either reporting a clear rejection of the app or making a strong statement about the app being useless to satisfy any purpose or need.

Anger

Anger expresses **fury** and **rage** experienced as a result of the use of an app. It implies a sense of **fierceness** and **hate**, implying that the user has encountered an obstacle causing huge disruption.

Anger is clearly expressed through hate:

[Ex37] I hate telegram

[Ex38] I hate that it disconnects if the app is not running in the background and sometimes it misses to push notifications.

Anger is often transmitted through hate speech, including insults or extremely negative vocabulary:

[Ex39] And you can see from the support replies, all you get is some worthless canned response from some mindless wage slave, at best.

[Ex40] Stupid Java supporters.

Finally, *Anger* can also be expressed as a complementary emotion when users express fury through writing style, either with capitalization, exclamation marks or emojis.

[Ex41] The food log needs a BARCODE SCANNER!

Anger and *Disgust* can also be confused sometimes for the same reason, as they are contiguous emotions according to Plutchnik's Wheel. In this context, when in doubt, we will apply the following distinction:

- If the user expresses **hate, fury, rage** or any anger-related emotion, **even if it is also rejecting the use of the app** (implicitly or explicitly), then we will consider the review as *Anger*.
- If the user **does not express any anger-related emotion**, then we will consider the review as *Disgust*.

For example, in the following *Anger* example:

[Ex42] *It wasn't what I needed and I absolutely HATE that they don't even tell you the timers are for premium memberships only.*

While the user is clearly stating that the app does not meet its expectations, the use of hate-related vocabulary implies *Anger* as the predominant emotion.

On the other hand, in the following *Disgust* example:

[Ex43] *Until you change the privacy settings, I won't update*

There is no negative or hateful judgement in addition to the rejection of the app by the user.

Anticipation

Anticipation expresses **curiosity** and **consideration** about something unknown or unclear to the user. It implies a sense of **alertness** and **exploration**, related to looking ahead to what could come with uncertainty and/or expectation.

Anticipation can relate to inquiries and questions about particular upcoming changes, also expressed through feature requests or change proposals:

[Ex44] *Something's I'd have liked to see - get reports of specific projects - ability to add subtasks to tasks*

[Ex45] *I would love a barcode scanner to be included.*

Additionally, *Anticipation* also relates to **considerations and questions raised by users about something confusing**, not clear, for which they **do not know how to proceed**:

[Ex46] *But then again, you are not supposed to pause a pomodoro timer, right?*

While they are opposite emotions according to Plutchnik's Wheel, *Surprise* and *Anticipation* can often be confused, especially when the sentence implies uncertainty or concerns about something. In this context, when in doubt, we will apply the following distinction:

- If the review entails a reactive emotion from the user's point of view (i.e., raised by a particular event or situation, as a reaction to something that happened), then we will consider *Surprise*.
- If the review entails a proactive emotion from the user's point of view (i.e., raised by the curiosity and active consideration of the user), then we will consider *Anticipation*.

For instance, in the following *Surprise* example:

[Ex47] wanted to go to an event later in the month showed only 3 attendees but rsvp was closed not showing canceled?

The user experienced an unexpected situation as a result of their use of the app, based on an external incident/event.

However, in the following *Anticipation* example:

[Ex48] I'm trying to figure out what it means when someone phone text messaging an a Goggle account are Linked?

While the user expresses uncertainty about something, this is not emerging from a reactive situation from a particular event, rather than being a consideration (i.e., *anticipation*) of the user towards the future of the app.

Finally, *Anticipation* can also be confused with sadness when users ask for a missing feature. In this context, when in doubt, we will apply the following distinction:

- If the review expresses explicit disappointment with the missing feature, then we will consider *Sadness.*
- If the review expresses as a wish for the feature to be included, with no disappointment connotations, then we will consider *Anticipation.*

Annotation instructions

Here is a summary of the guidelines for conducting the annotation:

1. The annotation task is conducted at the sentence level.

2. For each sentence, you will be given the following information:

   a. <u>Metadata</u>: including the mobile app name, the category of the app, the review identifier, the sentence identifier (within the review), the date of publication and the rating of the review.

   b. <u>Feature</u>: a particular feature mentioned within the review (you can ignore this field)

   c. <u>Review</u>: the full text of the review. You can use the full text for context and a better understanding of the sentence to annotate.

   d. <u>Sentence</u>: the sentence to annotate.

3. Annotations are sentence-level. Full review can ONLY be used for context when there is some cognitive/linguistic ambiguity (e.g., elliptic subjects, pronouns…)

4. Assigned emotions cannot be based on judgment or reasoning inference based on the opinion stated by the user. Annotations must be restricted to what the user expresses explicitly.

5. Annotated emotions must refer exclusively to the current state of the app. If a user expresses an emotion referring to a past version of the app or event, this emotion must not be considered for annotation.

6. Each emotion should be assessed independently in a binary manner (Yes/No), meaning a sentence can have multiple emotions assigned if they are explicitly expressed.

7. Exceptionally, in addition to Plutchik's emotions, you can use the following tags:

    a. *Neutral* → to annotate sentences which are either too short to infer any emotion or reflect purely descriptive, objective observations. This includes the situation where a sentence does not include any emotion linked to the current state of the app.

    b. *Reject* → to annotate sentences which are either in a non-English language or cannot be processed from a semantic standpoint (e.g., noisy content, waste reviews, indecipherable content…).

8. Additionally, use the *'comments'* column to leave any comments about the annotation process of that sentence.

References

[1] Bin Lin, Nathan Cassee, Alexander Serebrenik, Gabriele Bavota, Nicole Novielli, and Michele Lanza. 2022. Opinion Mining for Software Development: A Systematic Literature Review. ACM Trans. Softw. Eng. Methodol. 31, 3, Article 38 (July 2022), 41 pages. https://doi.org/10.1145/3490388

[2] Maalej W., Kurtanović Z., Nabil H., Stanik C. 2016. On the automatic classification of app reviews. Requirements Engineering, 21 (3), pp. 311 - 331. https://doi.org/10.1007/s00766-016-0251-9

[3] Guzman E., Maalej W. 2014. How do users like this feature? A fine grained sentiment analysis of App reviews. IEEE 22nd International Requirements Engineering Conference, RE 2014 - Proceedings, art. no. 6912257, pp. 153 - 162. https://doi.org/10.1007/10.1109/RE.2014.6912257

[4] Panichella S., Di Sorbo A., Guzman E., Visaggio C.A., Canfora G., Gall H.C. How can I improve my app? Classifying user reviews for software maintenance and evolution. (2015) 2015 IEEE 31st International Conference on Software Maintenance and Evolution, ICSME 2015 - Proceedings, art. no. 7332474, pp. 281 - 290. https://doi.org/10.1109/ICSM.2015.7332474

[5] Yana Pashchenko, Mst Farjana Rahman, Md Shamim Hossain, Md Kutub Uddin, Tarannum Islam, Emotional and the normative aspects of customers' reviews, Journal of Retailing and Consumer Services, Volume 68, 2022, 103011, ISSN 0969-6989, https://doi.org/10.1016/j.jretconser.2022.103011

[6] Asghar, M.Z., Khan, A., Bibi, A. *et al.* Sentence-Level Emotion Detection Framework Using Rule-Based Classification. *Cogn Comput* **9**, 868–894 (2017). https://doi.org/10.1007/s12559-017-9503-3

[7] Xiong W, Zhang Y. An intelligent film recommender system based on emotional analysis. PeerJ Comput Sci. 2023 Mar 9;9:e1243. https://doi.org/10.7717/peerj-cs.1243

[8] Elena-Ruxandra Luțan, Costin Bădică. Emotion-Based Literature Books Recommender Systems.Proceedings of the 18th Conference on Computer Science and Intelligence Systems, M. Ganzha, L. Maciaszek, M. Paprzycki, D. Ślęzak (eds). ACSIS, Vol. 35, pages 275–280 (2023). http://dx.doi.org/10.15439/2023F8647

[9] Robert Plutchik, Chapter 1 - A GENERAL PSYCHOEVOLUTIONARY THEORY OF EMOTION, Editor(s): Robert Plutchik, Henry Kellerman, Theories of Emotion, Academic Press, 1980, Pages 3-33, ISBN 9780125587013, https://doi.org/10.1016/B978-0-12-558701-3.50007-7

[10] Imbir, K.K. (2017). Psychoevolutionary Theory of Emotion (Plutchik). In: Zeigler-Hill, V., Shackelford, T. (eds) Encyclopedia of Personality and Individual Differences. Springer, Cham. https://doi.org/10.1007/978-3-319-28099-8_547-1

Appendix - Summary cheatsheet

Emotion	Expresses..	Sense of…	Contains…	Confused with…
Joy	excitement pleasure	possibility positiviness	app appraisal like feature or characteristics	Trust: it's *Joy* if no explicit mention of the user personal experience
Trust	acceptance connection	safeness warmness	personal experience user goals and needs satisfaction positive privacy and security observations	Joy: it's *Trust* if explicit mention of the user personal experience
Fear	stress anxiety	agitation scared	complaining about trust issues ask for help with frustration and anxiety	-
Surprise	shocked unexpected	unpredictability	unexpected issues encountered by users unknown events or situations	Anticipation: is *Surprise* if reactive behaviour, with specific trigger
Sadness	disappointment loss	heaviness severeness	negative user experience bug or fault reports	Disgust: is *Sadness* if disappointment concerning user expectations Anticipation: is *Sadness* if only disappointment about missing feature (no explicit request)
Disgust	lack of trust rejection	bitterness refusal	explicit refusal to use the app claiming app is not suit for purpose	Sadness: is *Disgust* if explicit rejection and distrust of the user for the app Anger: is *Disgust* if no anger-related content or vocabulary
Anger	fury rage	fierceness hate	hate speech insults and extremely negative vocabulary	Disgust: is *Anger* if any anger-related content is present
Anticipation	curiosity consideration	alertness exploration	inquiries about changes feature requests questions about uncertain aspects	Surprise: is *Anticipation* if proactive behaviour, without specific trigger Sadness: is *Anticipation* if feature request is objectively requested