Content Labeling Based on Unstable Emoji Reactions

Oleksiy Mints
SMK College of Applied Sciences
Klaipeda, Republic of Lithuania
e-mail: oleksii.mints@smk.lt

***Abstract*—This paper proposes using unstable emoji reactions, which are available on some online communication platforms, as a tool for crowdsourced content labeling. The study was conducted using the Russian-speaking Telegram community as an example. The paper analyzes 220,972 comments with reactions from August 2021 to April 2025. The study examined the drift in the meaning of frequently used reactions: 👍, 👎, ❤, and 🤡. The results showed that the meaning of emojis is not constant and can change over time. This is especially true for emoji reactions that are not standard for online communication platforms. Changes in the meaning of emoji reactions are most likely during periods of high social turbulence. The results show that unstable emoji reactions can be used to label content on online platforms only if they are analyzed individually based on their meaning in a specific community.**

***Keywords* — Emoji reactions, content labeling, sentiment analysis, Telegram, unstable emoji.**

I. Introduction

Sentiment analysis is important for ensuring safe and comfortable online communication. At its core, it allows us to quantify and interpret the emotional tone of messages from individual users and entire online communities of various sizes – from workplace chat rooms to global news channels. In different communities, sentiment analysis can perform functions such as monitoring emotionally charged and polarizing content, identifying potential manipulation, moderating content, and monitoring participants' morale.

Traditional sentiment analysis is lexicon-based. In this case, emotional meanings are assigned to textual elements (n-grams), usually relying on predefined sentiment dictionaries. The main advantages of lexicon-based models are computational simplicity and ease of implementation.

Further development of large language models has led to multimodal approaches in sentiment analysis being created. These approaches aim to overcome the limitations of methods based only on textual lexicons. Multimodal models integrate linguistic and visual features, enabling a more nuanced understanding of emotions in digital communication [1], [2]. Pre-trained neural networks demonstrate promising results in emotion recognition. For instance, the authors of [3] employed a BERT architecture that incorporated CNN, RNN, and Bi-LSTM. However, despite their accuracy, multimodal methods have not yet been widely adopted due to their computational complexity and critical dependence on large-scale, context-labeled training datasets.

Therefore, classical sentiment analysis is still preferred for practical use in many online communication applications because it is less computationally intensive.

However, lexicon-based models only perform well under conditions of contextual stability, and when the meaning of lexical items is universal. These models are less effective in environments with nuanced communication because they often overlook the evolving, community-specific semantics of additional communication symbols, such as emojis.

Emojis provide valuable cues for detecting differences between the emotional tone implied by the surface level of text messages and their contextual content. In sentiment analysis research, emojis are often seen as additional symbols that authors use to emphasize certain nuances of their tone, such as sarcasm. However, emojis may not only be present in texts authored by the message sender but also denote reactions of other community users to the content. Authors of [4] found that Facebook reaction frequencies serve as strong indicators of users' emotional responses. For example, a high number of angry reactions versus love reactions reliably signals a negative versus positive overall reception. This allows emoji reactions to be used for crowdsourced content labeling.

The number of emoji reactions on Facebook is quite limited. Therefore, the meanings of these reactions can be considered stable. At the same time, some communication platforms, such as Telegram or Discord, offer their users a large number of emojis to use as reactions. In such conditions, the contextual meaning of emojis can vary at different times and in different communities. The author's observations, supported by an analysis of recent publications [5], demonstrate that some emojis can have unstable meanings and exhibit contextual plasticity in certain online communities. This occurs when the meaning of an emoji changes over time, becoming different from its commonly accepted meaning.

The purpose of this work is to identify the possibility of using such unstable emoji reactions to content labeling on platforms with a wide selection of emoji reactions.

II. Data Preparation and Analysis

The data source for this study is a publicly available Telegram channel primarily used for interaction in Russian, whose members are largely expatriates from Russia, Ukraine, and Belarus living in Georgia. The research period spans from August 1, 2021, to April 1, 2025. This timeframe covers several significant political events in Georgia and will capture potential shifts in communicative norms.

https://doi.org/10.51408/csit2025_07

Personal identifiers beyond the public display name were neither harvested nor stored. After deduplication and verification, the corpus contained 448,502 messages, 422,423 of which were user comments. Of those comments, 220,972 carried at least one emoji reaction. The message with the most reactions received 1,198.

Data collection routines were written in Python 3.12 with the Telethon library. Data preprocessing, statistical analysis, interactive exploration, and clustering were conducted in Orange Data Mining v. 3.38.0, using custom Python scripting nodes for sentiment calculations.

2.1. Data preprocessing

The Orange workflow of the data preprocessing stage is shown in Fig.1.

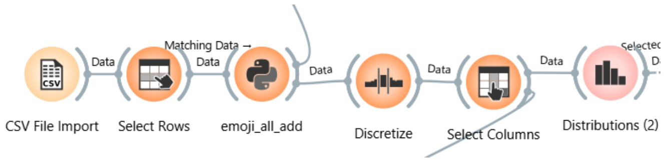

Fig. 1. Workflow of data preprocessing

The following actions were performed at this stage:

– Posts that were not comments were filtered out.
– Reactions of each type were allocated as separate features.
– The statistical characteristics of the data sample were calculated.
– The data was discretized by posting date for further analysis.

A histogram of the distribution of posts in the channel for the period under study was built (Fig. 2).

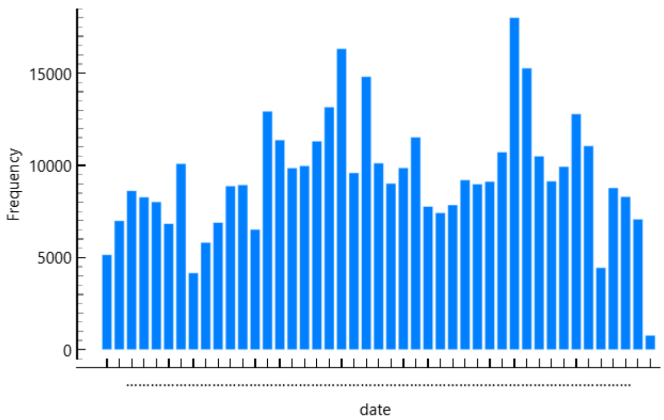

Fig. 2. Distribution of the message's frequencies

Analysis of the distribution histogram shows that the average intensity of comments changed insignificantly from the beginning to the end of the research period. The peaks on the histogram correspond to the following periods: Mar 22; Sep 22; Mar 23; May 23; May 24; and Oct 24. Most of these periods are associated with political events, both internal and external, that strongly impacted Georgia and are reflected in the increased activity of the channel's subscribers.

2.2. Data analysis

We used the unigram-based method proposed in [6] to evaluate the sentiment of the messages. Then, we calculated sentiment statistics for messages marked with the top 5 specific emojis. To eliminate random factors, we only considered messages marked with at least 10 emojis. We found that a large proportion of the studied messages were evaluated as neutral (sentiment = 0). Therefore, statistics were additionally calculated for messages with a non-zero sentiment value. The results are shown in Table 1.

Table 1. Sentiment statistics for the top 5 emojis

Emoji	total msgs with ≥ 10 emoji	mean (all)	dispersion (all)	mean (excl 0)	median (excl 0)	dispersion (excl 0)
all	70410	1.51377	4.67294	2.76337	3	3.3926
👍	44120	1.56504	4.32875	3.03016	3.07765	2.28992
😁	11223	1.61794	4.55714	3.87002	5	2.84611
👎	7993	1.40581	5.2307	3.14332	3.125	3.64852
🤡	8450	1.52482	4.105	2.18176	2.74024	4.20452
❤	5952	1.96434	2.59	3.80136	2.94118	1.72805

Analysis of Table 1 shows that a significant proportion of emojis are weakly related to the emotional evaluation of text. Thus, the mean and median sentiment values for messages with stable opposite emojis, such as 👍 and 👎, are nearly identical. These values are also close to the corresponding characteristics for the entire data sample. We may conclude that these emojis are primarily used to express agreement or disagreement with the message.

As shown in Table 1, only ❤ and partially 😁 among popular emojis can adequately indicate text sentiment. Their use correlates with higher sentiment values, so they can be considered expressions of gratitude for the content of the message (❤) or thanks for funny content (😁).

Now, let's look at the statistics for the 🤡 (clown) emoji. If we consider only messages with non-zero sentiment scores, the mean and median values for this emoji are much smaller than those for others. At the same time, the dispersion is larger. The frequency histogram of use of this emoji during different periods of community development (Fig. 3) is also distinctly different from the others.

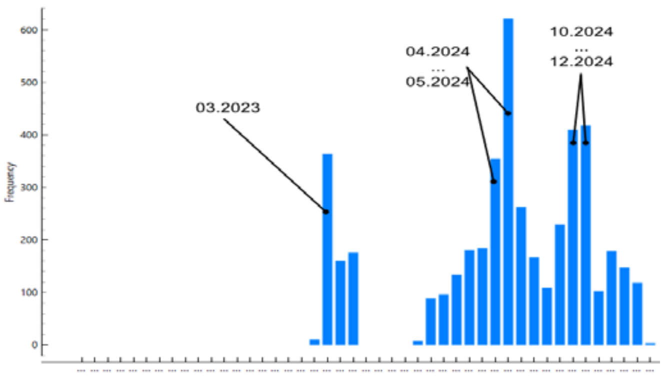

Fig. 3. Distribution of the 🤡 emoji frequencies

Fig. 3 shows that the emoji "clown" became popular for the first time in March 2023. During this period, mass protests against the "Law on Transparency of Foreign Influence" began in Georgia. The law was removed from the agenda at the end of March 2023, and the protests ended. The next two peaks in Fig. 3 are also connected with political events in Georgia.

During these events, there was increased activity by bots – program agents or people posting repetitive, manipulative messages in online community. They were quite difficult to detect automatically. But active participants in online communities learned to recognize these "bots" quite easily and started labeling them with the rarely used emoji 🤡. Over time, this practice spread to the rest of the community. Thus, here this emoji began to be used as a label for bot activity.

It should be noted that many messages marked with the emoji 🤡 were later deleted and did not get into the dataset, which significantly violates the statistics. However, as Fig.3 shows, from about the end of 2023, emoji 🤡 began to be used more actively. Let's consider the peculiarities of its use with the following workflow (Fig. 4).

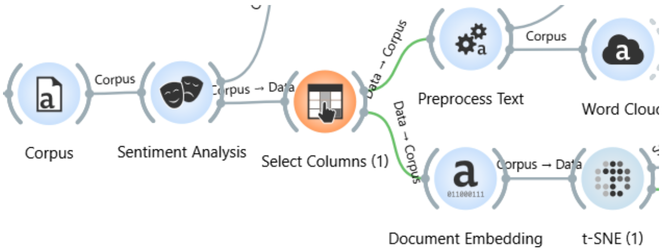

Fig. 4. Workflow of “clown” emoji analysis

This workflow calculates sentiment predictions for messages containing the emoji 🤡 and clusters them using t-SNE to identify the characteristics of different message groups. The analysis of these clusters revealed the following:

– Among messages with high sentiment scores, many can be regarded as sarcastic. In this case, the emoji 🤡 is an indicator of sentiment-context mismatches.

– Among messages in the same cluster (Fig.5), there are often nearly identical messages. This confirms the observation of high bot activity in the studied community, even though many of these messages were deleted.

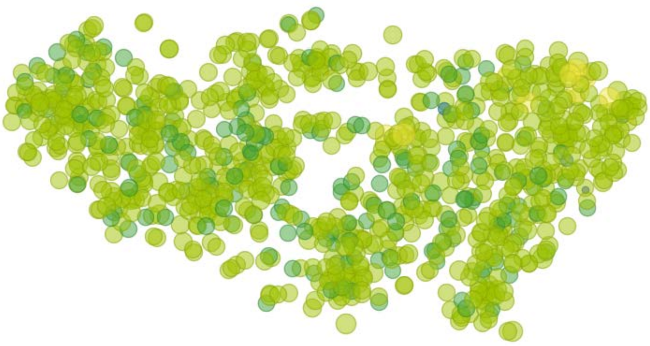

Fig. 5. t-SNE data projection

– A Word Cloud analysis of messages tagged with this emoji revealed a significant number of words with political connotations. These words are among the most frequent nouns and appear 2-3 times more frequently than in the entire sample (Table 2).

These calculations show that the 🤡 emoji is a highly likely sign of polarizing content in the analyzed data that causes conflicts among community members. During the period under review, such content was most often political.

Table 2. Statistics on using politics-related words.

Word	In a subset of the general dataset	In the messages, marked by “clown”	Ratio
Грузия (Georgia)	1093	2218	2,03
Россия (Russia)	269	841	3,13
Украина (Ukraine)	<100	231	>2,31
грузин (Georgian)	271	652	2,41
страна (country)	433	1357	3,13

III. Conclusions

The analysis showed that the communicative role of the emoji reactions changes over time. However, these changes occur relatively rarely and are most likely related to external influences. For example, the online community under study grew approximately threefold during the analyzed period. New participants may bring new rules that combine with existing traditions and cause the community to evolve.

The most obvious contextual shift observed in the study was the transformation of the 🤡 reaction emoji from an ironic symbol to a warning about bot activity and polarizing content. The meanings of stable emojis, which were also studied (👍, 😁, 👎, ❤) either remained unchanged or changed insignificantly over the period under review.

Thus, unstable emoji reactions can be used to label content on online platforms only if they are analyzed individually based on their meaning in a specific community. It is also necessary to take into account that the meaning of these reactions may also change over time. Nevertheless, the use of unstable emoji reactions expands the modalities of sentiment analysis and makes it more accurate. Further research is needed to clarify the conditions under which the meanings of such reactions shift, which will expand the possibilities for their practical application in content labeling.

References

[1] Y. Lou, J. Zhou, J. Zhou, D. Ji, Q. Zhang, “Emoji multimodal microblog sentiment analysis based on mutual attention mechanism”, *Scientific Reports*, vol.14, pp.1--12, 2024.

[2] H. Wang, C. Ren, Z. Yu, “Multimodal sentiment analysis based on multiple attention”, *Engineering Applications of Artificial Intelligence*, vol.140, art.109731, 2025.

[3] J. Cao, C. S. Prakash, W. Hamza, “Attention fusion: a light yet efficient late fusion mechanism for task adaptation in NLU”, *Findings of the Association for Computational Linguistics: NAACL* 2022, pp. 857--866, 2022.

[4] Y. Tian, T. Galery, G. Dulcinati, E. Molimpakis, C. Sun, “Facebook sentiment: reactions and Emojis”, *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*, Valencia, Spain, Association for Computational Linguistics, pp. 11--16, 2017.

[5] E. Morales, J. Hodson, A. Gruzd, P. Mai, “I👍 your Hate: Emojis as Infrastructural Platform Violence on Telegram”, *Proceedings of the 58th Hawaii International Conference on System Sciences*, pp. 2316--2325, 2025.

[6] Y. Chen, S. Skiena, “Building sentiment lexicons for all major languages”, *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, vol. 2 (Short Papers), pp. 383--389, 2014.