Sentiment Perception of Readers and Writers in Emoji use

Jose Berengueres  
College of IT  
UAEU  
Al Ain, UAE  
jose@uaeu.ac.ae

Dani Castro  
IT Department  
Happyforce S.L  
Barcelona, Spain  
dani@myhappyforce.com

Previous research has traditionally analyzed emoji sentiment from the point of view of the reader of the content not the author. Here, we analyze emoji sentiment from the point of view of the author and present a emoji sentiment benchmark that was built from an employee happiness dataset where emoji happen to be annotated with daily happiness of the author of the comment. The data spans over 3 years, and 4k employees of 56 companies based in Barcelona. We compare sentiment of writers to readers. Results indicate that, there is an 82% agreement in how emoji sentiment is perceived by readers and writers. Finally, we report that when authors use emoji they report higher levels of happiness. Emoji use was not found to be correlated with differences in author moodiness.

Keywords—Sentiment; Emoji; Happiness.

I. INTRODUCTION

A few studies that correlate emoji with sentiment and other variables exist. For example, [1] correlated the emoji people use with the economic development of the country of the tweet. Emojis can also be used to profile the gender of the author of a tweet [2]. Regarding sentiment, a few papers correlate emoji with sentiment by asking humans to rate tweets where a given emoji appears [3, 4]. Perhaps, the most comprehensive list is [4]. In [4] the sentiment ($s$) of an emoji is defined as the mean rating of the tweet where the emoji appears. The scale used is $\{-1, +1\}$ where -1 is most negative and +1 is most positive. Finally, different phone makers and operating systems will display the same emoji code differently. For example, Apple's emoji images are different from the original NTTdocomo emoji, this impacts how emoji are perceived. [5] studied the variations in interpretation depending on the phone and operating system.

A. Author sentiment

Publicly available benchmarks provide data on sentiment of the reader, usually inferred from a human asked to rate a tweet. However, none has considered the sentiment of the author due to lack of data rating the sentiment of the writer of a text or tweet. Here, we analyze author sentiment and at we present a list of emoji with their correlation to author sentiment as calculated by [4]. We do this by taking advantage of a dataset that contains comments with emoji that happen to be annotated for happiness of the writer of the comment. This text is organized as follows:

B. Overview

In the next section we explore the dataset answering simple questions such as, is people that use emoji happier than the rest? In the analysis section, we correlate emoji use with two variables:

(i) the happiness score at the time of posting, and
(ii) the happiness variation

Finally, in the discussion section we compare writer sentiment to reader sentiment and highlight the differences.

II. DATA

In a closely related paper by our group [6], we developed a model that predicts employee turnover from how an employee used a happiness self-reporting mobile application in the context of company feedback. Then, we determined the top risk factors of employees that churn in their jobs taking advantage of the top predicting features. Here, we use the same data source with new samples that have been added since the time of publication [6]. The text of comments is completely anonymized (words have been removed, only emoji remain). The csv data and R code are available at https://github.com/orioli/emoji-writer-sentiment

A. Source

Data used here was collected from 2014-05-10 to 2017-03-08. The data was generated by employees that worked at 56 different companies. The companies belong to one of the following sectors: e-payment startup, IT consulting services, retail, manufacturing, services, tourism or education. About half of the companies are multinational and the other half are Barcelona-based companies. Employee data was collected in the framework of corporate feedback program as part of their companies’ kaizen initiative. The bulk of the employees used the this mobile application in Spain (Barcelona area). More than 90% are Spanish nationals. The comments are written in various languages: 97% in Spanish, 2% in English, 1% Catalan. The data consisted of four tables: votes, comments, likes, employee churn. Here, we will focus on two of them: votes and comments. An in-depth review of the dataset is available at [6]

B. User flow

A happiness vote was obtained when an employee opened the app and answered the question: How happy are you at work today? To vote, the employee indicates their feeling by touching one of four icons that appeared on the screen. See 1st screen, Table 1. The UI of the English version is shown in the table. The default UI was in Spanish language. After the employee indicates their happiness level, a second screen appears where they can input a text explanation (usually a suggestion or
comment), this is the comments table. Finally, in a third screen the employee can see their peers’ comments and like or dislike them. Following, we describe and visualize key metrics of each of the tables.

**TABLE I. SCREENS OF THE APP USED TO COLLECT DATA**

| Table (Rows) | Feed-back | UI flow |
|--------------|-----------|---------|
| **Happiness votes** (398k) | How happy are you at work today? | ![1st screen](image) |
|              | - Great: +1 points  |
|              | - Good: 0 points   |
|              | - So-so: -0.5 points |
|              | - Pretty Bad: -1 points |
| **Comments** (68,476) | Comment box | ![2nd screen](image) |
| **Likes / Dislikes** (599,103) | Anonymous forum Users can: | ![3rd screen](image) |
|              | - view comments  |
|              | - like a comment |
|              | - dislike a comment |

**C. Self-reported happiness**

Table 2 shows basic statistics about happiness votes aggregated by user. A four-grade scale is used to convert the happiness text categories to numeric. The scale is chosen to map to values as similar as possible to the polarity scale commonly used in sentiment analysis. Therefore, the original happiness labels are converted to numeric as follows: \{bad:-1, so-so: -0.5, good: 0, great: +1\}. The most common answer, good, is assigned zero value. Using this numeric scale, the mean of happiness of all votes is **0.0688** (mean all-time happiness).

Fig. 1 is a histogram of votes recorded. The August and January month dips correspond to the Summer and Christmas vacation period in Spain. Weekly dips are weekends. Fig. 2 is a histogram of votes (includes votes without comments). Company names are anonymized with a one or two-letter code. Fig. 3 is a histogram of app usage derived from the votes timestamps. 0.5 means that the employee used the app once every two days, (days include working days, weekends and holidays). About 10% of the employees used the app every day to report their happiness.

**TABLE II. STATS ON HAPPINESS**

| Item                        | Max | Min | Mean | Median |
|-----------------------------|-----|-----|------|--------|
| happiness vote              | +1  | -1  | 0.008 | 0.00   |
| votes per user              | 905 | 1   | 57   | 24     |
| mean happiness by user      | +1  | -1  | 0.12  | 0.01   |
| sd happiness by user        | 1.4 | 0   | 0.41  | 0.43   |

**Fig. 1** Count of votes casted daily. Each color is a company. The app is being used by an increasing number of companies.

**Fig. 2** Count of answers to the question: ‘How happy are you at work today?’ (1st screen). X axis is index of the categories: \{1: bad, 2: so-so, 3: good, 4: great\}. Period: first 2.5 years.

**Fig. 3** Usage of the app by vote activity. About 10% of employees use the app every day.

**D. Comments & Emoji Extraction**

Table 3 visualizes the top emoji used by the employees in their comments in the forum. Table 4 visualizes basic statistics of the comments table. In total, 68k comments were recorded.
with a median length of 58 chars per comment. 5% of comments contain emoji. 962 users used emoji at least once, 4893 users never used emoji. Out of 63k comments, 3,506 comments contain at least one emoji. 358 different types of emoji are used.

| TABLE III. | TOP 15 EMOJI |
|------------|-------------|
| Emoji code | count | % over all emoji | Description |
| 44f        | 884   | 8.8       | Clapping hands |
| 4aa        | 879   | 8.7       | Flexed biceps |
| 3fb        | 701   | 6.9       | Fitzpatrick-1 lighter tone* |
| 618        | 471   | 4.7       | Face throwing a kiss |
| 44d        | 435   | 4.3       | Thumbs up |
| 389        | 347   | 3.4       | Party popper |
| 3fc        | 303   | 3.0       | Fitzpatrick-3 light skin tone* |
| 64c        | 214   | 2.1       | Two hands up |
| 483        | 194   | 1.9       | Dancer woman |
| 382        | 182   | 1.8       | B-day cake |
| 381        | 168   | 1.7       | Wrapped present |
| 914        | 166   | 1.6       | Thinking face |
| 917        | 147   | 1.5       | Hugging face |
| 36d        | 144   | 1.4       | Fitzpatrick-4 dark skin tone* |

* a modifier, not an emoji.

| TABLE IV. | OVERVIEW OF COMMENTS |
|------------|----------------------|
| Item       | Max | Min  | Mean | Median |
| Length of comment in chars | 1k* | 1    | 101.7 | 60 |
| Posting date | 2017-09 | 2015-02 | 2016-12 | 2017-02 |
| Count of Emoji per comment that contains emoji | 14b | 1 | 2.4 | 2 |
| Author happiness when comment does contain emoji | +1 | -1 | 0.38 | 1 |
| Author happiness when comment does not contain emoji | +1 | -1 | 0.17 | 0 |
| Count of all Emoji used per user that uses emoji | 268 | 1 | 10 | 4 |
| Emoji per comment of users that use emoji | 49 | 1 | 4.17 | 2 |
| % of comments with emoji by users that used emoji at least once | 90% | 0.4% | 20.1% | 15% |
| Happiness of votes when the user also posted a comment after voting | +1 | -1 | 0.18 | 0 |

* comments of longer length are excluded  
 b comments with more than 15 emoji are excluded

Fig 4 visualizes the distribution of count of emoji in the 3850 comments with emoji that were detected. Comments with more than 15 emoji or longer than 1k char were omitted as outliers. The mean like ratio of a comment with emoji, computed as (likes received / (likes + dislikes)) is 7.95. The like ratio for a comment without emoji is 8.09, the difference is not significant.

Emoji as html entities where extracted with a regex with the following regex expression '€\#x1[a-zA-Z1-9]{5}€'. 10088 occurrences were detected. As stated before indirectly, of the 63k comments 59k contain no emoji and the rest contain varying amounts of emoji. Fig. 5 compares seven of these groups. Top: shows how the ratio of likes to dislikes that a comment receives is correlated with the number of emojis in a comment. Bottom: is same looking at the total number of likes received. In both cases, statistically significant differences between groups where not found. In other words, emoji in a comment does not influence the sum of likes a comment gets nor the likes to dislikes ratio (social approval).

Fig 4 3850 comments have emoji, 43% include only one.

Fig 5 Left: X is the like ratio. Right: X is count of likes received. Y is density of comments with colored by number of emoji. The number of emoji used in a comment is not correlated with the likability of the comment.

### III. ANALYSIS

Regarding author happiness, we look at two parameters: (i) the happiness of the author at the time of posting and (ii) the change in happiness compared to the author’s historic mean (baseline happiness). We call this change in happiness that occurs daily for everybody bias.
A. Quantization noise

Note that because a 4 point happiness categorical scale is used in the answers {bad, so-so, good, great}, quantization noise will be introduced when converting them to numeric. From basic signal processing, we can assume this is a typical analogic to digital signal quantization scenario where if we assume that the error introduced, (when categorical text happiness was quantified to scale \{-1, -0.5, 0, +1\}), is not correlated with the happiness value then, we have that the quantization error introduced always follows a uniform noise distribution. For a step of size 1.0, the mean of such noise is 0, the variance = 1/12, and sd=0.29. The variance for a step size 0.5 is 1/16, and the sd=0.25. In our scale, we have two steps of size 0.5 and one of size 1. In addition, two of the steps are bounded on one side. Therefore, we can take 1/16 as a lower boundary of introduced in our happiness variance measurements. Since the mean is zero, we will not worry about this noise except in mean aggregates with sample sizes <24. On the other hand, the lower boundary of the variance can be subtracted from the measured variance to find the a closer approximation to the variance of the happiness before quantization noise was introduced, see table 6 footnotes. This substraction is accurate if we assume both distributions are gaussian and not correlated [7].

B. Happiness drift

In addition, for simplicity we assume that the mean happiness of users does not drift or decay drastically for the users (employees). Decay in employee satisfaction has been found to be correlated to employment length and has three components that we will ignore here in order to simplify. The three components are: honeymoon effect in new employees, hangover effect in new employees, and steady decay in employees that do not change job positions for 2 or more years [8]. This effects have been confirmed to appear in this dataset too (not shown here).

C. Comments with emoji vs. comments without emoji

Fig. 6 compares comments that contain no emoji vs. the rest. Each colored point is a comment. Left chart: the Y-axis is the happiness of the author when they posted the comment. X-axis is the date of the comment. Right chart is a density plot. Moving averages are shown. Table 4 lists the means for each group. To reduce the effect of happiness drift [8] in employees we consider the last 10 months of observations which also contain most data points. The right graph is a density plot. In fig. 5 we saw that emoji does not affect the social approval of a comment. However, when an author uses emoji they are happier than when not. A Kolmogorov test also supports this hypothesis. However, the margin of confidence of the moving averages CI 95, (shown in grey) overlaps for some periods of time. Sample size is N=3850 for comments with emoji and N=60k for comments not containing any emoji.

Fig. 7 is the same as fig. 6 but Y-axis is bias as defined earlier. Dots are not shown in colors for readability. The right chart is the corresponding density plot of the comments that contain at least one emoji vs. all the comments with no emoji. A Kolmogorv test does not support the hypothesis that the two distributions have different means (p-value = e-16). This is, posting a comment with or without emoji is not correlated with a significant greater increase or decrement in happiness on the day of posting (moodiness).

D. Emoji users vs. non-emoji users

We have compared comments to comments. Now we compare authors to authors by computing aggregates at author level. Fig. 8 shows a scatterplot for the 925 users (authors) who used emoji at least once. The blue line is a linear regression. The figure shows how the emoji use rate is related to more happiness. Users that use emoji more than 50% of the time also exhibit high long term happiness.

E. Emoji correlation with bias and happiness

Fig 9 shows a scatter plot of two given specific emoji. Y is the company mean, X is the company sd. A great variability exists between companies. Fig. 10 and 11 shows the happiness and bias and for 4 groups of emoji. The distribution of positive emojis such as the clapping hands seems to resemble a Tracy-widom distribution with a longer right tail and a steeper left tail.

![Figure 6](image-url)

Fig. 6 Are authors happier when they use emoji in a post? Yes, but not all the time. Note that compared to fig. 2, the most common answer is now (4: great) assigned +1 score.

![Figure 7](image-url)

Fig. 7 Y (bias) is the difference in daily happiness minus mean of each author. X is timeline. Labels are same as in fig 6. Red is for comments that contain no emoji. Only 7k points are shown due to computational restrictions.

| Item | Does author use emoji? |
|------|------------------------|
| Item | no | yes | used emoji today? |
| never | no | yes |
| Average happiness | 0.10 | 0.27 | 0.38 |
| SD happiness | 0.73 | 0.73 | 0.69 |
### F. Estimating sentiment from emoji

How well can emoji that appear in a comment be used to estimate the sentiment of the author on the day he posted the comment? Fig. 12 is a scatter plot of sd vs. mean happiness for the most common emoji in our dataset. Fig. 13 is the same for bias.

In Table 6, for each emoji we list the mean happiness of author (writer sentiment) when they wrote the comment where the emoji appears. Next to it, we list reader sentiment provided by [4]. This sentiment was computed by humans reading and rating tweets via Amazon Mechanical Turk using a scale from -1 to +1, s is the mean sentiment asigned by several human raters to the same tweet (or tweets) where a given emoji appears. The diff column is the difference. Fig. 14 is a scatter plot of reader sentiment vs. writer sentiment. The grey line is y=x. Points above indicate emoji sentiments where writer rating > reader rating. Notice that there are two emoji that are rated negative sentiment by the writers but positive sentiment by readers. These emoji are: crying face and disappoint but relieved face. Another highlight is the pouting face emoji whose writer sentiment is double the reader sentiment, both negative.

| N datapoints | 36934 | 22286 | 3850 |
|--------------|-------|-------|------|
| N unique authors (user id) | 3714 | 925 |

![Fig. 8 Each dot represents an emoji user (an author that used emoji at least once), the sample size N=0.9k X-Axis is the ratio of number of comments with emoji divided by all the comments the author wrote (emojiphilia). Y-axis is the long term self reported happiness of the author. Line is a linear regression. R-squared 0.006. Empty quarter: users with a ratio greater than 50% are seldom unhappy.](image8)

![Fig. 9 Comparing companies to companies by two emoji. Only a subset of companies is shown for clarity.](image9)

![Fig. 10 Distribution of bias. When pouting face emoji is used the author happiness level was less than his average 75% of the time.](image10)

![Fig. 11 Bar density plot of 3 emoji and the top 30 emoji together. When clapping hands emoji is used, the author was at a happiness level of 4 more than 70% of the time.](image11)

### G. Significance

To estimate the significance of difference in means between reader and writer for a given emoji such as the emoji ‘disapointed but relived face’, we have that reader sentiment is: 0.122, SD=NA, N= 341 and writer sentiment is -0.275, sd= 0.634, N=60. We assume sd reader = sd writer as we don’t have the value. Then, applying a Kolmogorov test with simulated samples with the said statistics we have that the difference in means is significant with an alternative hypothesis p-value = 0.0001048.
H. Regression

A linear regression to estimate writer sentiment from reader sentiment (not shown in fig. 14, but close to the y=x line) yields a Multiple R-squared: 0.8244. Meaning that 82% of the sentiment variation in the writer can be inferred from the ratings by the readers provided by [4]. However, the intercept is -0.04830 meaning that if reader sentiment is used to estimate writer sentiment, it will underestimate on average by an average of 0.04 points, or normalizing by the sd of writers (0.41), by about 10% of their sd.

IV. DISCUSSION

We have analyzed emoji sentiment of Spanish workers from three points of view: comment level, author level and emoji level. The main conclusions are:

A. Emoji is correlated with happier users

Authors that use emoji self-report happier levels than authors that never use emoji. While both show similar variability, authors that never use emoji self-report an average happiness of +0.10 while authors that use emoji report +0.27 on days that they did not use emoji in their comments and +0.38 on days they do use at least one emoji in their comment. However, correlation does not imply causality. In fig. 2 we saw that the most common vote for happiness is {3: good} assigned 0 points in our scale. However, when we look at the same histogram considering only users that commented at least one, the most common vote is {4: great} assigned +1 here (see fig. 6). A self-selection of happier users seem to be happening. We suspect that the same is happening when a user includes an emoji in a comment.

Entering an emoji requires an extra effort by the writer as changing the keyboard from QWERTY to emoji and back is required. Thus, the cause of increased happiness in users that use emoji, (as in users that care enough about their jobs to input text as opposed to just voting), might not be the emoji but the fact that users that care enough to ameliorate or beautify their feedback with emoji might be more engaged than the average. Engagement was found to be a predictor of less turnover in [6] and less turnover can be interpreted as happier worker. The number of user that use the app > number of users that vote and comment > number of users that vote, comment and use emoji > number of users that vote, comment and use emoji and use emoji in today's comment

B. Moodiness

As we saw in fig. 7, bias (the difference between daily happiness and an author's average long term happiness) is not
correlated with emoji use on a given day. In other words, emoijphilia is not correlated with higher or lower moodiness.

C. Predictive power

Certain emoji are correlated with high levels of unhappiness. For example, the pouting face emoji is correlated with the lowest self-reported happiness \(-1\) in more than 50% of cases it is used. For comparison when a random emoji is used the author also reports the maximum happiness possible \(+1\).

D. Writer vs. reader sentiment

In table 6 we compare reader sentiment with writer sentiment for emoji with more than 24 counts. The linear regression that estimates writer sentiment from the reader sentiment yields an R-square of 0.82 meaning there is a 18% of the variability in writer sentiment that is not explained by the current benchmark provided by [4], (see column "s.reader"). In addition, even thought the regression R-squared is high the intercept is +0.04. This means that a reader originated benchmark such as [4] consistently overestimates the (sentiment) happiness of authors by 0.04 points (10% of the standard deviation). This number is even higher if we subtract the lower boundary of the variance introduced by the quantization noise. Finally, overall readers and writers seem to broadly agree on emoji sentiment on the most common emojis even though the reader source of comments in [4] is Twitter and the writer source of comments is a work-happiness monitoring app [6].

Finally, the 'Ok sign' is the emoji where readers and writer's sentiment agree the most. The 'pouting face' is the emoji where readers and writer disagree the most. Writers perceive it much more negatively than readers do.

| code  | s.writer | s.reader | sd | writer | count | description                      | s.reader - s.writer |
|------|---------|---------|----|--------|-------|----------------------------------|---------------------|
| 499  | 0.864   | 0.730   | 0.409 | 24     |       | BLUE HEART                       | -0.134              |
| 61a  | 0.786   | 0.710   | 0.418 | 28     |       | KISSING FACE WITH CLOSED EYES    | -0.076              |
| 37b  | 0.766   | 0.512   | 0.523 | 32     |       | CLINKING BEER MUGS              | -0.254              |
| 619  | 0.739   | 0.778   | 0.511 | 44     |       | KISSING FACE WITH SMILING EYES  | 0.039               |
| 483  | 0.719   | 0.734   | 0.483 | 80     |       | DANCING WOMAN                    | 0.015               |
| 3b5  | 0.712   | 0.500   | 0.493 | 26     |       | MUSICAL NOTE                     | -0.212              |
| 64c  | 0.692   | 0.559   | 0.533 | 99     |       | PERSON RAISING BOTH HANDS...     | -0.133              |
| 44f  | 0.692   | 0.520   | 0.513 | 305    |       | CLAPPING HANDS SIGN              | -0.172              |
| 38a  | 0.689   | 0.531   | 0.491 | 37     |       | CHRISTMAS TREE                   | -0.158              |
| 38b  | 0.684   | 0.537   | 0.512 | 38     |       | MULTIPLE MUSICAL NOTES           | -0.147              |
| 31e  | 0.654   | 0.558   | 0.502 | 39     |       | SUN WITH FACE                    | -0.096              |
| 4aa  | 0.646   | 0.555   | 0.525 | 445    |       | FLEXED BICEPS                    | -0.091              |
| 38a  | 0.645   | 0.721   | 0.486 | 51     |       | CONFETTI BALL                    | 0.076               |
| 44d  | 0.642   | 0.521   | 0.545 | 288    |       | THUMBS UP SIGN                   | -0.121              |
| 38l  | 0.632   | 0.738   | 0.488 | 170    |       | PARTY POPPER                     | 0.106               |
| 618  | 0.599   | 0.701   | 0.551 | 232    |       | FACE THROWING A KISS             | 0.102               |
| 61c  | 0.591   | 0.455   | 0.570 | 88     |       | FACE WITH STUCK-OUT TONGUE...    | -0.136              |
| 44c  | 0.561   | 0.563   | 0.599 | 90     |       | OK HAND SIGN                     | 0.002               |
| 382  | 0.451   | 0.613   | 0.500 | 122    |       | BIRTHDAY CAKE                   | 0.162               |
| 38l  | 0.400   | 0.759   | 0.493 | 85     |       | WRAPPED PRESENT                  | 0.359               |
| 38l  | 0.341   | 0.718   | 0.480 | 41     |       | BALLOON                          | 0.377               |
| 525  | 0.317   | 0.139   | 0.835 | 30     |       | FIRE                             | -0.178              |
| 61l  | 0.290   | 0.190   | 0.791 | 40     |       | FACE SCREAMING IN FEAR           | -0.101              |
| 62c  | 0.188   | 0.194   | 0.667 | 40     |       | GRIMACING FACE                   | 0.007               |
| 64f  | 0.180   | 0.417   | 0.646 | 61     |       | PERSON WITH FOLDED HANDS         | 0.237               |
| 44e  | -0.103  | -0.188  | 0.829 | 39     |       | THUMBS DOWN SIGN                 | -0.085              |
| 62d  | -0.195  | -0.093  | 0.647 | 82     |       | LOUDLY CRYING FACE               | 0.102               |
| Code | Expression | Score 1 | Score 2 | Score 3 | Score 4 | Score 5 |
|------|------------|---------|---------|---------|---------|---------|
| 629  | WEARY FACE | -0.209  | -0.368  | 0.526   | 43      | -0.159  |
| 615  | CONFUSED FACE | -0.216  | -0.397  | 0.534   | 37      | -0.181  |
| 625  | DISAPPOINTED BUT RELIEVED FACE | -0.275  | 0.122   | 0.634   | 60      | 0.397   |
| 622  | CRYING FACE | -0.279  | 0.007   | 0.660   | 68      | 0.286   |
| 613  | FACE WITH COLD SWEAT | -0.306  | -0.080  | 0.576   | 49      | 0.226   |
| 614  | PENSIVE FACE | -0.316  | -0.146  | 0.675   | 49      | 0.170   |
| 611  | EXPRESSIONLESS FACE | -0.328  | -0.311  | 0.577   | 32      | 0.017   |
| 623  | PERSEVERING FACE | -0.365  | -0.212  | 0.573   | 37      | 0.153   |
| 61e  | DISAPPOINTED FACE | -0.394  | -0.118  | 0.506   | 71      | 0.276   |
| 621  | POUTING FACE | -0.613  | -0.173  | 0.552   | 53      | 0.440   |

* only count < 24 shown

* data source [4]: http://kt.ijs.si/data/Emoji_sentiment_ranking/

* Most agreement

* Most disagreement

* To obtain a more accurate sd estimate subtract the quantization noise variance; \( sd' = \sqrt{sd^2 - 1/16} \)

REFERENCES

[1] Ljubešić, Nikola, and Darja Fišer. "A global analysis of emoji usage." ACL 2016 82 (2016).

[2] Martinc, Matej, et al. "Pan 2017: Author profiling-gender and language variety prediction." Cappellato et al.[13] (2017)

[3] Kiritchenko, Svetlana, Xiaodan Zhu, and Saif M. Mohammad. "Sentiment analysis of short informal texts." Journal of Artificial Intelligence Research 50 (2014): 723-762.

[4] Novak, Petra Kralj, et al. "Sentiment of emojis." PloS one 10.12 (2015): e0144296

[5] Miller, Hannah, et al. "Blissfully happy" or “ready to fight”: Varying Interpretations of Emoji." Proceedings of ICWSM 2016 (2016).Giachanou, Anastasia, and Fabio Crestani.

[6] Jose Berengueres, Guillem Duran and Dani Castro. Happiness an inside Job? Turnover prediction from likeability, engament and relative happiness. IEEE/ACM ASONAM 2017 (in Press)

[7] Daniel Marco and David L. Neuhoff, "The Validity of the Additive Noise Model for Uniform Scalar Quantizers", IEEE Transactions on Information Theory, Vol. IT-51, No. 5, pp. 1739–1755, May 2005. doi:10.1109/TIT.2005.846397

[8] Boswell, W., Shipp, A., Payne, S. & Culbertson, S. (2009). Changes in newcomer job satisfaction over time: Examining the pattern of honeymoons and hangovers. Journal of Applied Psychology, 94 (4), 844-858.