International Journal of Data Science and Analytics
https://doi.org/10.1007/s41060-018-0135-9

REGULAR PAPER

Characterizing negative sentiments in at-risk populations via crowd computing: a computational social science approach

Jesus Garcia-Mancilla[1] · Jose E. Ramirez-Marquez[2] · Carlo Lipizzi[2] · Gregg T. Vesonder[2] · Victor M. Gonzalez[1]

Received: 22 February 2017 / Accepted: 2 June 2018

Abstract

Drawing on psychological theory, we created a new approach to classify negative sentiment tweets and presented a subset of unclassified tweets to humans for categorization. With these results, a tweet classification distribution was built to visualize how the tweets can fit in different categories. The approach developed through visualization and classification of data could be an important base to measure the efficiency of a machine classifier with psychological diagnostic criteria as the base (Thelwall et al. in J Assoc Inf Sci Technol 62(4):406–418, 2011). Nonetheless, this proposed system is used to identify red flags in at-risk population for further intervention, due to the need to be validated through therapy with an expert.

Keywords Crowd computing · Depression characterization · Twitter

1 Introduction

Traditional approaches for negative emotion identification, like surveys and interviews applied to large populations, are expensive and time-consuming [27]. For example, in the study from Gonzalez-Celis and Padilla [17], which measured quality of life in elderly people in Mexico City, 196 men and women over 60 years old were interviewed, with each interview lasting 60 minutes. Such a research approach, as an example, is one of the most common practices in cognitive psychology and requires significant resources in terms of time, people, and money. Furthermore, such an approach generally results in a limited number of test subjects and in the best cases, only large or federal organizations (like the Centers for Disease Control and Prevention [CDC]) can perform them at a significant scale. Novel, innovative and less resource-intensive approaches are needed to understand state of mind in larger populations.

Given the magnitude of the data that have to be collected to analyze big complex problems involving humans, social networks and their corresponding information output socio-computational processes can provide an alternative and a powerful platform to mine information at a fraction of the resources needed by traditional approaches, and give useful insights from different areas [6]. This capability has been deployed with the use of people acting as social sensors to provide information about events and natural disasters [3,21,32], event detection [14], social mobility [27], and to provide an estimation of general public health [4,7,13,15,29]. Specifically, Twitter has been used to identify more complicated human behavior like mental health [8], including emotion and sentiment analysis [5,9,16,18,20,24,25].

Negative sentiment identification, with the purpose of diagnosing mental health disorders like depression using information generated in social networks, is not an easy task. The symptoms related to these disorders are usually identified after long conversations with patients, which become complex to establish in social media like Twitter because the predominance of single-message or few message communications. Our goal then is to provide an analysis framework on the characterization of different negative sentiments and

✉ Jose E. Ramirez-Marquez
jmarquez@stevens.edu

Jesus Garcia-Mancilla
research@jegama.io

Carlo Lipizzi
clipizzi@stevens.edu

Gregg T. Vesonder
gvesonde@stevens.edu

Victor M. Gonzalez
victor.gonzalez@itam.mx

1 Department of Computer Science, ITAM, Mexico City, Mexico

2 School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ, USA

Published online: 23 June 2018

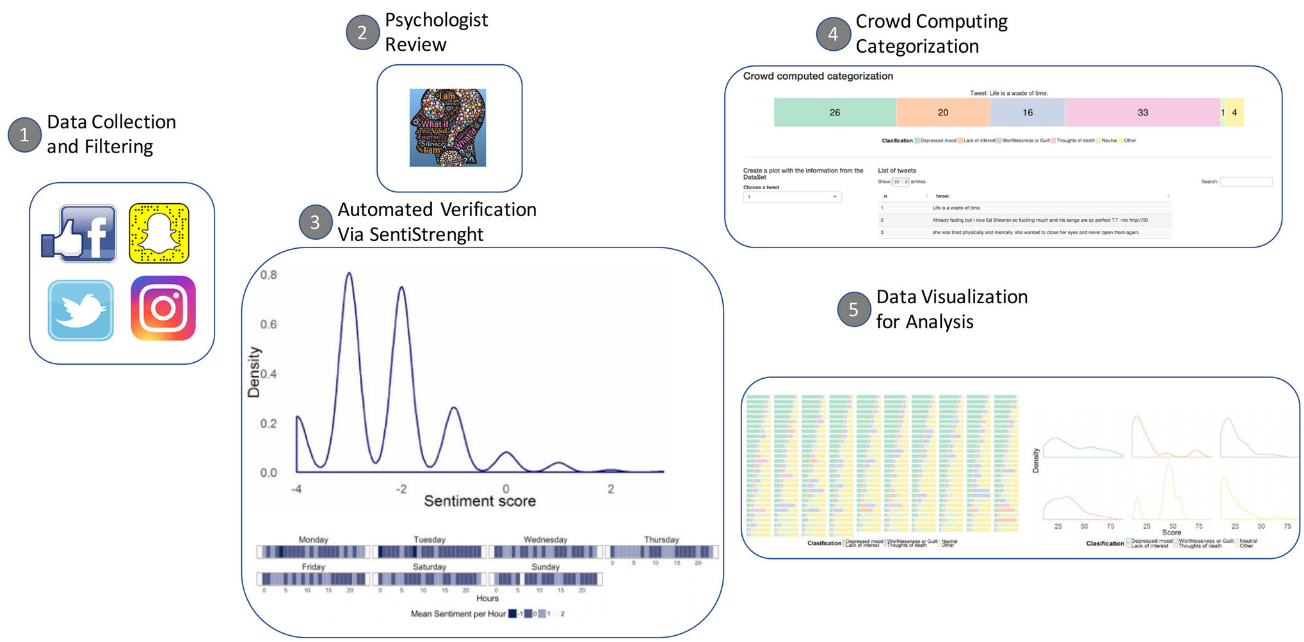


Fig. 1 Proposed approach

how they are exposed in Twitter messages (tweets) by users, and through comparative analysis of large samples to be able to identify symptoms of the disorders. This approach could be helpful in the development of new methodologies to help identify people with potential mental health disorders by text mining of their communications as an alternative for diagnosing at early stages. In no measure is this research intended to supplant any health care or medical professional in the arduous process of identifying depression per se.

Despite the efforts to identify emotional state in large populations via the use of social media (and Twitter specifically) [10–12], to the best of our knowledge, the application of psychological theory has largely been overlooked other than to list general behaviors of people with depression. Thus, for example, the link between how depression is diagnosed by psychologists and the use (and availability) of social network data to identify such ailments in large populations is still unexplored. This paper intends to provide an initial understanding of such a link. To address this goal, the team of researchers is multidisciplinary with expertise in data mining, sociology, psychology, data visualization, machine learning, and human–computer interaction. The approach described in Fig. 1, developed through visualization, classification, and clustering of data could be an important base to measure the efficiency of a machine classifier with psychological diagnostic criteria as a baseline to aid in the human interpretation of the signs of depression. Our results show the beginnings of a computational methodology based on psychological theory, crowd computing, and visualization to identify a negative emotional state (i.e., depression) among a population of users in a social network and to do it efficiently over a large population and through time and space.

2 Literature review

Among the many mental health disorders, we decided to focus on depression as a case to develop a specific context for our analysis. Other mental health disorders could be analyzed using a similar approach.

To have a better understanding of the characteristics of depression behavior, different diagnostic manuals have been consulted, to help us understand and differentiate such a behavior from other mental health problems with similar characteristics, like substance abuse side effects. Thus, we decided to use the diagnostic and statistical manual of mental disorders (DSM-V). This manual is currently in its fifth edition from the American Psychiatric Association, the leading psychiatric organization in America. According to this manual, different behaviors can be identified in a person with depression; these are: depressed mood (self-reported or reported by others), diminished interest or pleasure in activities, significant weight loss or gain, insomnia or hypersomnia, psychomotor agitation or retardation, fatigue or loss of energy, feelings of worthlessness or excessive or inappropriate guilt, diminished ability to think or concentrate, and recurrent thoughts of death. We refer to this set of behaviors as the list hence on [1].

Other behavioral characteristics of depression could be identified, but we are not considering them because the literature indicates that they are not as relevant as the ones in "the list." They include: significant distress or impairment in social, occupational, other areas of functioning; depression is not attributable to the physiological effects of a substance or to another medical condition; the occurrence of the major depressive episode is not better explained by other psychological conditions; and there had never been a manic or hypomanic episode [2]. All of these characteristics are part of an in person psychiatric diagnostic that certainly cannot be done at a massive population level nor through the use of social media. Thus, for the purpose for this work, we used behaviors from "the list" to identify different characteristics of a population with potentially depressive tendencies which can be detected through the analysis of Twitter messages. From the list, at least five behaviors have to be present and experienced for at least six months to conclude that the person is suffering from depression.

Similarly, while substance abuse is a major problem that has to be taken into account in psychotherapy sessions, following the literature, we did not consider it as factor for depression because there is no immediate way to identify this behavior with the profiles of the Twitter users and more importantly, it is often related to other mental health problems.

During the last few years, research done using social media information (and Twitter specifically) as a social sensor has been shown to be a valid approach that can aid in the detection and classification of a great variety of events in the real world. Regarding depression as a health problem, we identified three relevant papers from the work of Munmund De Choudhury and her colleagues, as important for this research area. One of their first works used a data set collected for a year, with tweets from people where some of them had a clinic diagnosis of depression [12]. In that work, they used Amazon Mechanical Turk (AMT) to identify people who were diagnosed with depression, and then they collected the tweets from those users for a year.

In an additional study by De Choudhury et al. [11], they collected tweets from 376 women while they were pregnant and then after childbirth; with these data, they proposed a method to predict if the woman will have postpartum depression. This is one of the first approaches to create a predictive model to detect whether an individual is at a high risk to have depression in the near future.

The final work from De Choudhury and the closest to our work uses a large tweet corpus of people with depression to train a support vector machine (SVM) classifier to detect depression on a large scale. The results were highly correlated with the data from the Centers for Disease Control and Prevention (CDC) [10]. The factors that they analyzed with the SVM classifier were: emotional expression (positive, negative, activation, and dominance), time (day (6AM–8:59PM) and night (9PM–5:59AM)), linguistic style (22 specific linguistic styles using Linguistic Inquiry and Word Count (LIWC)), user engagement (volume of tweets, replies from a posts author, retweets, links shared, question-centric post from a posts author), egocentric social network properties (number of followers or in-links of the user, and the count of her followers or out-links).

The work of De Choudhury aims to characterize people with clinical diagnose of depression, moving from the fact that they are already experiencing that disorder. In contrast, the approach that we present in this paper goes from an opposite direction: Based on the different identifiers "the list" emerging from the DSM-V, we analyze tweets of a person that could help us to identify a sad mood or depression.

The work from De Choudhury is one that illustrates approaches which are used for other researchers to detect depression within social networks like Twitter. For example, the work by Kumar et al. [22] found that, followed by the celebrity suicide, the content in social networks like Reddit tends to have more negative emotions and less social concern. This finding could help to understand patterns on social networks and help to filter data sets loaded with negative emotions.

One important point to highlight is that most of the work done in this area is completely classified and analyzed by computers with little human intervention. One of the studies that relied on human classification is by ODea et al. [30]; after the tweet mining, they gave them those tweets to experts that labeled them as: *strongly concerning*, *possibly concerning*, or *safe to ignore*. After that, the tweets on the first two categories were used to train a support vector machine (SVM). One of the main difficulties reported by the researchers was the differing opinions from the experts, making it more difficult to rely on the outcomes of the SVM.

One of the key challenges in detecting depression is its complexity, but geotemporal variables have to be taken into account while testing the proposed methodologies. Tsugawa et al. [35] explain that at least two months of data are necessary to create a significant analysis of a single individual in order to have a good classification. On the other hand, Yang et al. [38] analyzed the geographic patterns of the users and with that information, they improved their depression detection algorithm.

As seen, most of the computational approaches try to infer the state of the person from other variables or extracting characteristics from the text using traditional natural language processing; however, with the increasing popularity of recurrent neural networks (RNNs), but more specific, long short-term memory networks (LSTMs), more robust sentiment analysis can be done over the Twitter data [19]. This new generation of networks can infer context and understand more complex semantic correlations [31].

More traditional approaches are giving us important insights when they are applied to very large amounts of data.

As described by Sueki [33], examining phrases like *want to commit suicide* and *want to die* found an interesting pattern; the first phrase had a stronger correlation with depression than the second. Most people would think that the active desire of death could be the most important, but as found in this work, suicidal ideation is expressed more explicitly in social networks when depression is in a more advanced stage.

Other works [10,28,36,37] on the topic have social engagement, emotion polarity (positive–negative), language, linguistic states, ego network, mentions of antidepressant medication, social activity, and use of social media as common points and can be used to explain how Twitter can reflect the mood of people, and most of the papers have a psychological reference for each of them. Decrease in social activity, increase in negative affect, highly clustered ego networks, heightened relational and medical concerns, greater expression of religious involvement, reports almost the same as the CDC (geographical detection), and teenage girls tend to get more depressed using more social media are their main findings.

3 Method

With the purpose of understanding the evolution of negative emotion development in large-scale cohorts, our method is a five-step process that involves: (1) data collection, (2) machine-based sentiment analysis, (3) data filtering, (4) human-based crowd computing, and (5) visualization. Each of these steps is described as follows:

Step 1 Data collection: The development of our research started with identification of dependent and independent variables that are related to depression detection in Twitter users. This was done by the analysis of what the DSM-V [1] classifies as depression and the different characteristics identified by De Choudhury [10]. Based on the literature, as part of the selection of users for the cohort, they have to be evaluated to guarantee that they meet the criteria, as done by De Choudhury [12].

Identifying a cohort of users that post emotionally charged tweets is necessary to perform a correct analysis. Therefore, our steps are as follows:

i. We identified n major influencers (for our practical example $n = 4$), based on the number of mean followers (for our example Mean $= 113,600$) and the percentage threshold λ of depression-related tweets post per day (for our case $\lambda = 90\%$). For experimentation purposes, the reason to choose only 4 users was to have a fast accessible cohort of tweets with depression tendencies. The first step to characterize the influencer was manual; a random set of 50 tweets were assessed by a clinical psychologist (two of the authors have background in psychology) to verify that we were considering the correct cohort. This step, however, is not mandatory for future replications of the study, because the automated analysis proved to be a reliable way to identify the general tendencies of the cohort, as is later described in Sect. 4.1. Then, a month of data were collected, and an automated sentiment classifier was used to confirm our data set.

ii. After identifying the cohort, we retrieved the tweets, including the retweets and replies, using the methodology proposed by Lipizzi et al. [26]. All the tweets were collected using the open Twitter API and stored using MongoDB. We gathered tweets for a month, and the data set had 175 thousand entries. Each tweet was processed in order to calculate general sentiment (using *SentiStrength* algorithm), word diversity, and sender diversity. These last metrics were used to verify that the accounts used are human based and not automated. Thus, word diversity quantifies how many different words on average were used (higher diversity is associated to human based) and, for the sender diversity—how many different users were tweeting and having a conversation (not just retweeting the same tweet multiple times which is a typical bot behavior).

Step 2 Sentiment analysis: Once the data were collected, we did a sentiment analysis using *SentiStrength* [34]. By sentiment, we mean an emotional state, and the algorithm identifies sentiments such as happiness, excitement, or amusement as an increasing positive value, whereas sadness, mirth, or anger as a decreasing negative value. This algorithm estimates the strength of positive and negative sentiment, reporting in two polarities: -1 (not negative) to -5 (extremely negative), 1 (not positive) to 5 (extremely positive), and -1 to 1 neutral. With these characteristics and how the algorithm classifies text sentiment make it a good match for tweets given their length and also the ability to understand the categories it classifies. With this algorithm, we aimed to review the collected tweets, determining whether they had a general negative emotion and how strong the sentiment was, as a mechanism to create a distribution of the sentiments. Other general metrics as proposed by Lipizzi et al. [26] were analyzed, such as word and sender diversity, and the amount of tweets per day and their time of day. With this information, a visualization was generated to understand the collected data, which is shown in Fig. 3 of Results section.

Step 3 Data filtering: To have a better base for our new classification, we created a new data set, in which most of the replies and the retweets were removed. The reason to do so was based in our experience: In the case of the replies, those replies would create noise in the analysis due to the length

Table 1 Description of the classification

Classification	Description
Depressed mood	Shows sadness, emptiness, hopelessness, or subjective report of his kind of feelings
Lack of interest	Reports markedly diminished interest or pleasure in all, or almost all, kinds of activities
Worthlessness or guilt	Feelings of worthlessness, or excessive or inappropriate guilt (not guilt about being sick)
Thoughts of death	Recurrent suicidal ideation with or without a specific plan to suicidal attempt
Others	Angry mood, significant weight loss or gain, insomnia or hypersomnia, fatigue or loss of energy
Neutral	None of the above

of the responses, being too short, or specifically related to something, and without the context. And, the retweets on the other hand were removed because it was not necessary to have duplicate evaluations of the same tweet. After this process, the remaining tweets were read by a psychologist one at a time, and we kept the most relevant ones based on the following criteria: tweets length (more than 50 characters), perceived emotional intensity, and variety of topics. These criteria were selected based on previous works where these characteristics were considered important to provide a deeper analysis and understanding of the tweets [10,26]. This process was done twice to optimize the cohort of tweets for our goals, leaving a final data set of 500 unique tweets. Our manual process can be of course improved, and we are currently working on automating it for efficiency.

Step 4 Crowd computing: Once the new data set with the 500 filtered tweets was obtained, it was uploaded to AMT. In this environment, the turk was tasked with classifying each tweet into one and only one of the categories in “the list.” The categories selected were based on the behaviors described in the DSM-V and are detailed in Table 1. Our purpose here is to classify the filtered data set, using crowd sourcing on the AMT platform, and generate a distribution for each tweet. An AMT campaign was run to have the tweets classified within a single category by 100 different individuals, so as to have a human-based characterization of the tweets that could then be used in future works to build a more robust algorithm. AMT was used because of its immediate availability and it has been used by researchers to enlist a large amount of people to do small, repetitive tasks [23]. Moreover, AMT provided us the opportunity to understand how a large number of people understand and classify tweets (or people statements).

Step 5 Visualization: With the data set classified by AMT, we developed a set of visualizations to facilitate, to understand, and to characterize the differences of how a tweet can be categorized: i. Individual tweet distribution. ii. Tweets distribution. iii. Distribution of moods in the cohort. iv. Distribution of moods per user.

This approach provides a visual tool to gain insights of the complexity of the task and how the same tweet can be understood from different perspectives. More importantly, the population distribution can be used to visualize the behavior of the population through time and in comparison with other cohorts.

```
Method abstract;
Data: Literature analysis and understanding of the DSM-V
        diagnosing criteria.
Categorization model based on psychology theory.
Depression-related influencers on social media.
for tweet ← 0 to 150, 000 do
    Retrieve the tweets using the methodology proposed by
    Lipizzi et al. [26].
    General sentiment analysis using SentiStrength to verify the
    collected data.
    Analysis of the word and sender diversity, and other general
    metrics to discard bots.
end
Filter of 500 unique tweets to do a deeper analysis.
Start a AMT campaign to classify the filtered data set using
crowd computing.
while un-analyzed data do
    Analysis and visualization of the crowd-computed
    classification distributions.
    Unsupervised training to test classification.
end
```

4 Discussion of results

4.1 Sentiment analysis

To test the accuracy of the selected negative emotions cohort and the sentiment algorithm, the sentiment visualization was conducted in two different cohorts; the first one was our selected negative emotions cohort that was later used for the crowd computing analysis, and the second one was from 446, 172 tweets collected from 12 major influencers (with more than 106,000 followers and picked form TIME’s magazine lists of online influencers). This *healthy* cohort consists of three adults and three young people, with an even number of men and women. The purpose of having the second cohort is to understand if there is a visual difference in the sentiment analysis between the selected depressed cohort and a healthy one.

Two different visual analyses were done to identify patterns of the general sentiment calculated with *SentiStrength*

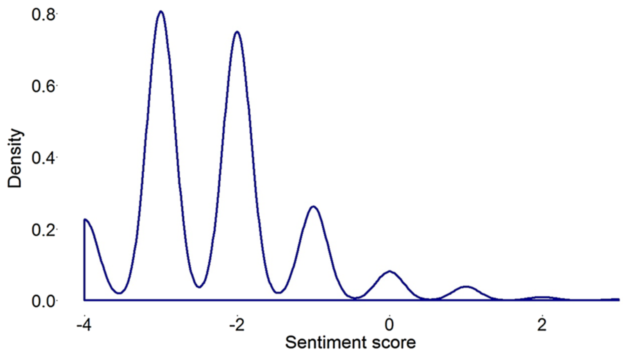


Fig. 2 Cohort with depressive tendencies: *Tweets'* sentiment density

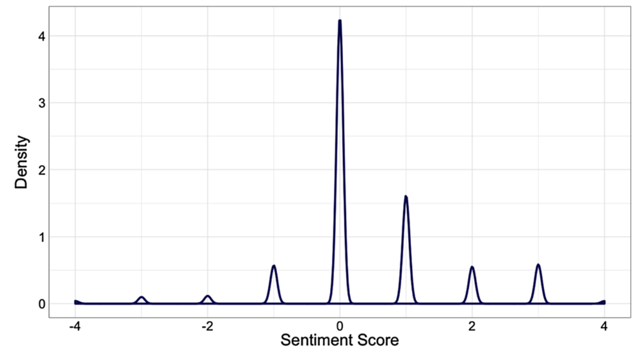


Fig. 4 Healthy cohort: *Tweets'* sentiment density

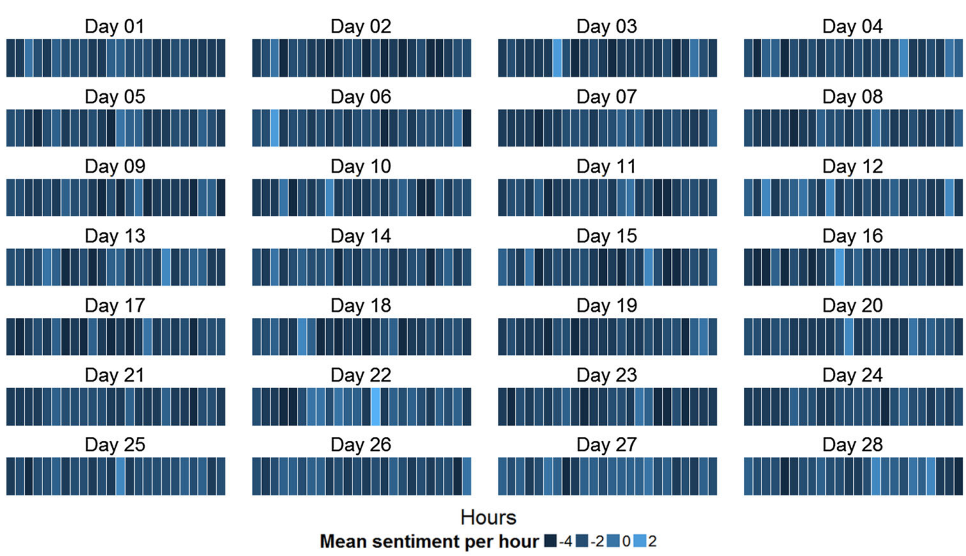
[34] for all the tweets collected from the major influencers, both in our selected cohort and the *healthy* cohort. Figure 2 shows the density scores for the first cohort (negative emotions), in which most of the tweets tend to have -2 and -3; they are negative sentiment scores. Only a few were positive, suggesting that our selected cohort has a high negative sentiment which could be related to depression.

A second analysis on Fig. 3 is relevant to show the power of visualization. The visual describes the mean sentiment per hour of each day providing a visual sentiment distribution for each of the 28 days described. While De Choudhury et al. [10] found that the general population tended to have negative sentiment during the night, our analysis of users depression-related tweets found that negative sentiment is distributed uniformly throughout the day. Note that from the visual if we suspect that daylight tweets tend to be more positive, a clear visual change of pattern would be seen in Fig. 3. However, if no significant difference is found between day and night (i.e., the tweets during the day are not significantly more negative than at night), then the visual will be mostly of a single constant color.

The two main goals for the sentiment analysis were: (1) to provide a comparison and an alternative metric to support that all the tweets were depression related, and (2) to see if having only unfiltered depression-related tweets from accounts that only tweet this type of messages could reflect another distribution than the one presented in previous studies [10], where it has been proposed that most of the depressive tweets are posted in the night.

Finally, based on the comparisons between Figs. 2 and 4, sentiment score tends to be mostly negative for our target cohort, while fairly positive and neutral for the "healthy" population. And between Fig. 3—mostly negative sentiment throughout the days—and Fig. 5—mostly neutral or positive tweets, respectively, for targeted and healthy cohorts, we claim that the sentiment distribution of normal healthy users does not exhibit the depressed mood in their tweets, and thus the focused cohort can be moved to the following steps.

Fig. 3 Cohort with depressive tendencies: *Tweets'* mean sentiment distribution by day by hour



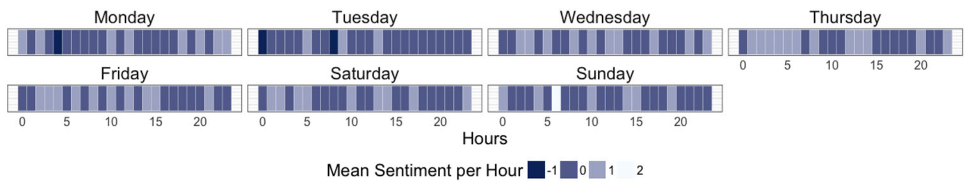


Fig. 5 Healthy cohort: *Tweets'* mean sentiment distribution by day of the week by hour

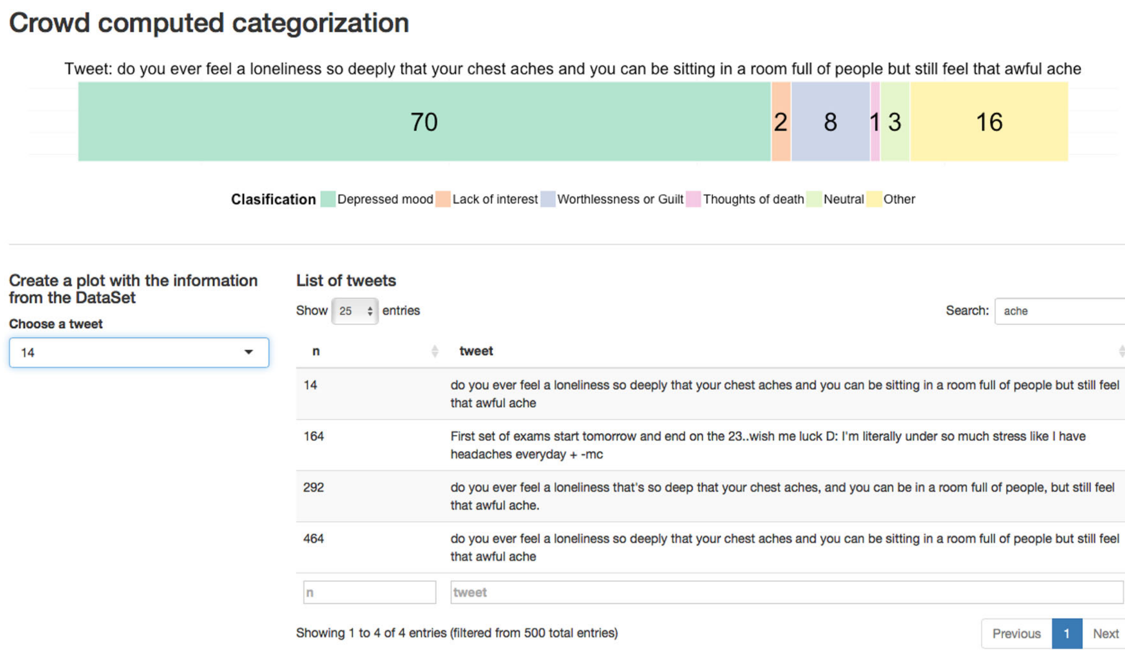


Fig. 6 Screenshot of the interactive classification from the AMT classification

4.2 Crowd computing

Given the complexity of the classification task, the results of the crowd computing campaign in Amazon Mechanical Turk provide an important insight onto how tweets from a depression-related cohort can be ambiguous or unrelated to depression. The risk is that the individual classifying a tweet might provide the wrong assessment and, thus we minimize such risk by validating with the assessment of many other human classifiers; finding those differences is a central part of this work. The assessment of 100 human classifiers provides us with the classification distribution of the tweet. Comprehensive view of all tweets and the crowd-computed classification through an interactive visualization can be found online.[1] A screenshot is shown in Fig. 6.

The interactive visualization serves as a search tool that can be used for identifying a specific tweet as a complement to the plot visualization. The table in the lower side can change depending on the volume of entries to be shown with the respective id for the plot.

4.3 Visualization analysis

4.3.1 Individual tweet distribution

As shown in Fig. 7, most tweets are relatively straightforward to classify; in the case of tweet number 14, 70% of people labeled it as *Depressed Mood*, and in the case number 16 and 212, 67% and 68% classified them as related to *Worthlessness or Guilt*. But in some tweets, the difference can be very subtle, and the classification cannot be made so clear, as seen in number 11. The same amount of people (31%) think that the tweet: Im not strong -mc can be classified as *Depressed Mood* or *Worthlessness or Guilt*.

Other issue that was identified in the first analysis was that not all the tweets are depression related. As shown in Fig. 7, the tweet 2: Already fading but I love British Singer so fucking much and his songs are so perfect T.T -mc http://00[2] is an example that new approaches are needed to classify and identify depression in social networks in addition to mining data from a cohort at risk of depression. Focusing solely on tweets for a cohort with depression and immediately process-

[1] https://jegama.shinyapps.io/MTurk.

[2] The name of the singer and the url were modified to protect the user's privacy.

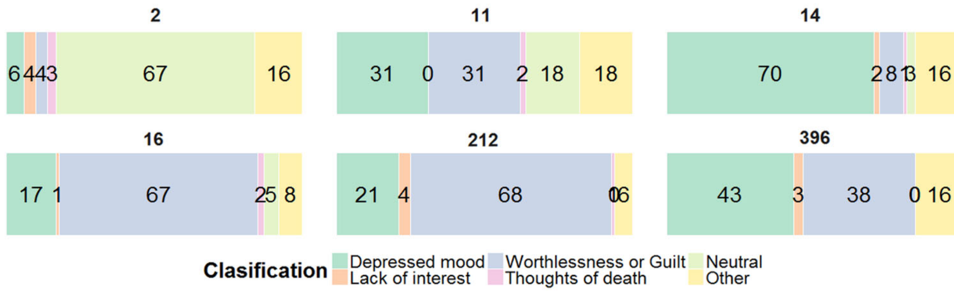


Fig. 7 Examples of how different *tweets* were classified

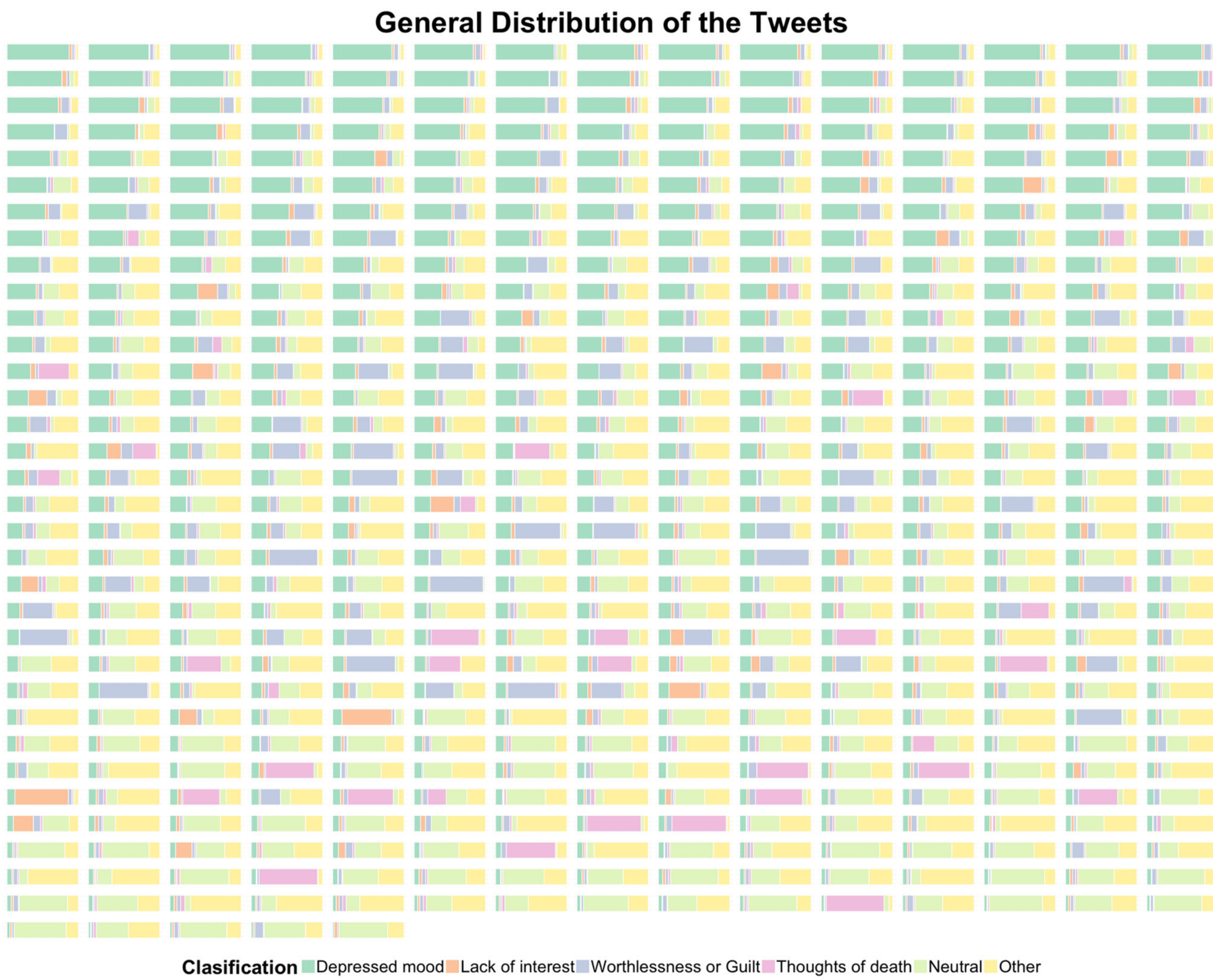


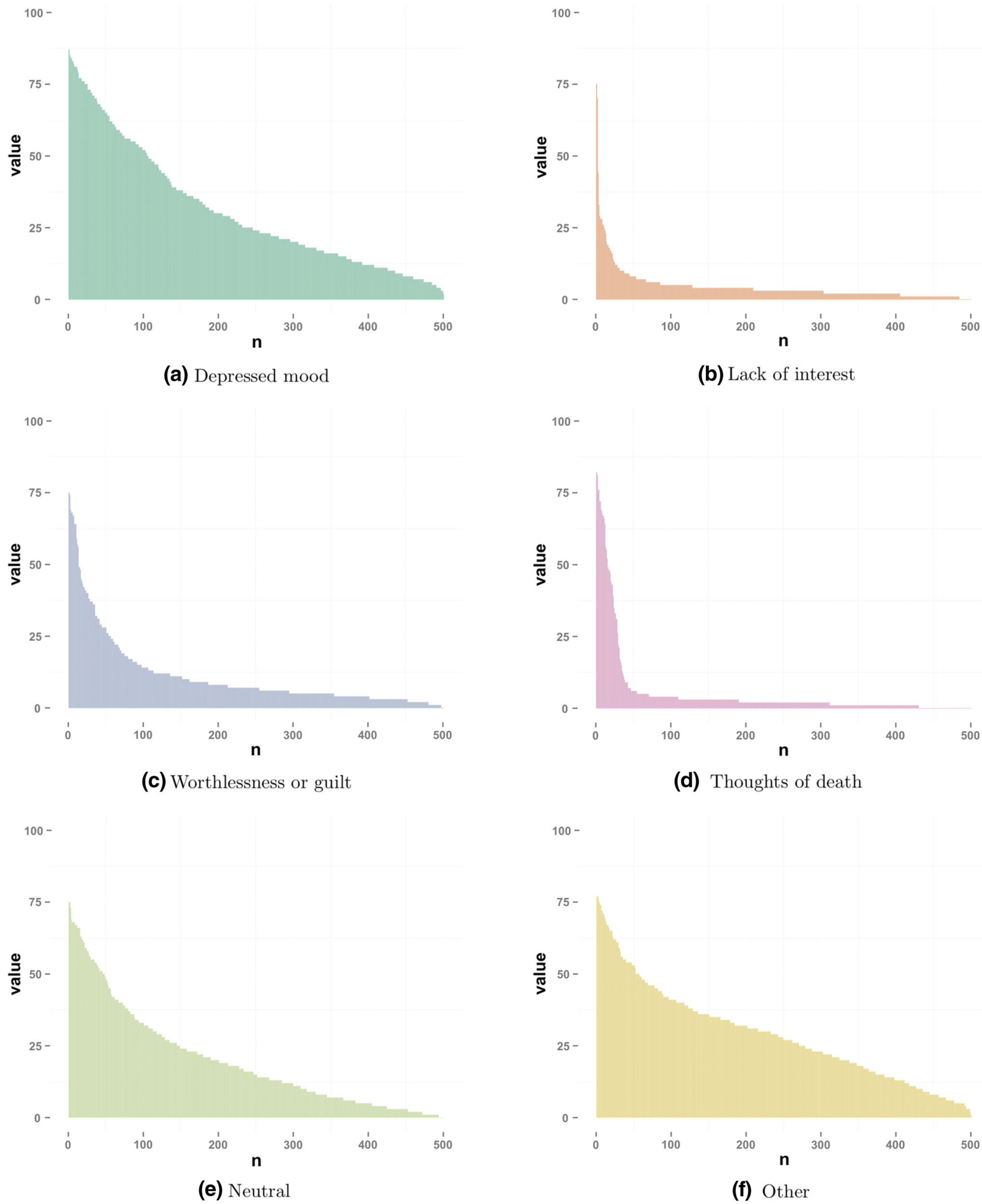
Fig. 8 Distributions of all the *Tweets*

ing the information via a machine learning algorithm could generate a lot of noise and give false positive results, since not all tweets are depression relevant.

4.3.2 Tweets distribution

These examples can be seen as some of the main representative distributions of the tweets. Figure 8 serves three purposes:

1. It shows how the population behaves in relationship with the DSM-V (i.e., how the 500 tweets fit in all the categories proposed in this work).
2. It can serve as a baseline to compare future cohorts or users at risk via a visual comparison.
3. It can be used to track the evolution of emotion of this cohort though time and space.



(a) Depressed mood
(b) Lack of interest
(c) Worthlessness or guilt
(d) Thoughts of death
(e) Neutral
(f) Other

Fig. 9 Histogram of each of the categories of the classification

As shown in Fig. 8, the *Worthlessness or Guilt* value is not directly correlated with the *Depressed Mood* one; however, *Thoughts of death* tend to be higher where *Depressed Mood* values are lower. Other expected results visible on Fig. 8 are related to the *Neutral* category, which can be seen as the opposite of the *Depressed Mood*, the lower this value, the higher the *Neutral* one.

4.3.3 Distribution of moods in the cohort

Figure 9 on page 11 shows how the *tweets* were voted on each one of the categories, helping to understand how all the proposed categories fit in our cohort.

4.3.4 Distribution of moods per user

Once we had a better understanding of the analysis of the cohort, we filtered the tweets from our major influencers. However, only three major influencers were considered due to the small number of tweets from one of the influencers.

Two different visualizations were done to understand the behavioral distribution of each user's tweets. The first one is a general distribution similar to Fig. 10, which helps to understand how the behavioral categories are present in each user.

The second one is a density distribution that for visualization purposes, two different graphs were proposed to understand how this approach could help to characterize negative emotions like depression. By filtering the values lower than 10, it becomes easier to understand patterns in the categories. Otherwise, some categories could lead to miss interpretations on earlier stages. Figures 11 and 12 are examples of this potential pitfall: Without the filter (11), the amount of tweets with scores close to 0 makes it difficult to visualize that in fact, some tweets have a *Lack of interest* with high scores (> 75, see Fig. 12). This number is significant, considering the consistent pattern of tweets with *Depressed mood* too.

For the first user (A), the distribution described in Fig. 10 shows that several tweets have high values of "Worthlessness or Guilt," which could be interpreted as a potential concern because, in the frequency graph (Fig. 12, it was found a considerable amount of tweets with values between 40 and 60 with "Thoughts of death." Analyzing both visualizations and the proposed classification, it could be possible to identify the depression progress in a person, which could be, in this case, at early stages.

On the second user case (B), a red flag could be raised, considering that, between all the three users, this one has the higher frequency of tweets with high scores of "Thoughts of death," and a higher spike in frequency between all users in Fig. 13, 14, and 15 whereby a specialist could intervene to get a better understanding in person. It is also recommended to follow the publications to identify a pattern because another category with high scores is "Neutral," which could be related to create a diversion from reality, pretty common after expressing thought of death.

Fig. 10 Sentiment distribution from user A

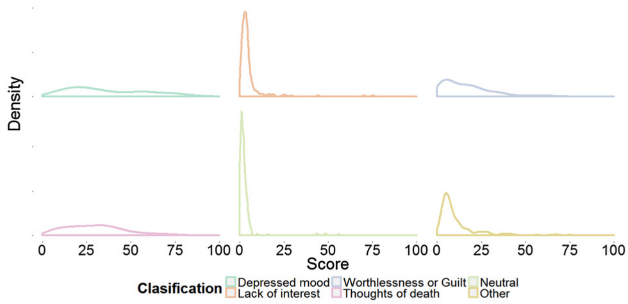


Fig. 11 Sentiment Distribution from user A (not filtered)

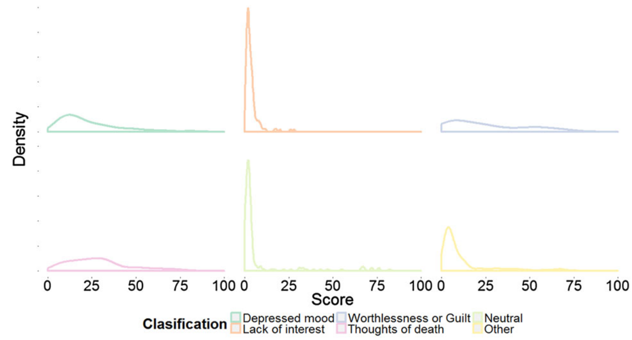


Fig. 14 Sentiment Distribution from user B (not filtered)

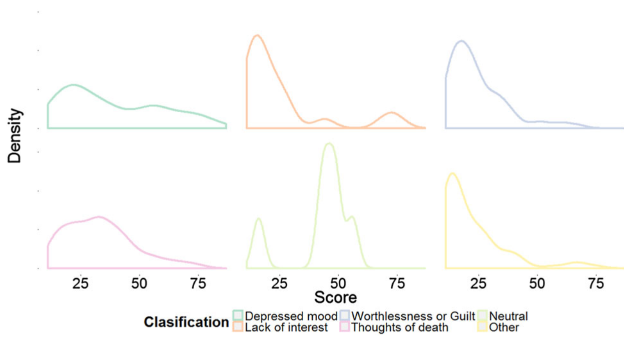


Fig. 12 Sentiment Density from user A (filtered)

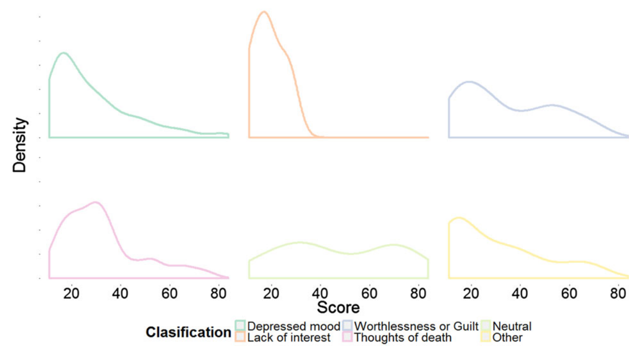


Fig. 15 Sentiment Density from user B (filtered)

The last user (C) is considered to have the higher risk due to the high percentage of tweets with high scores of "Thoughts of death," which is easily visible the general distribution at Fig. 16, 17, and 18. On the other hand, the user also has a high frequency of tweets with high scores of "Depressed Mood," hence, a closer follow up is recommended. However, more tweets are necessary to do a more detailed description.

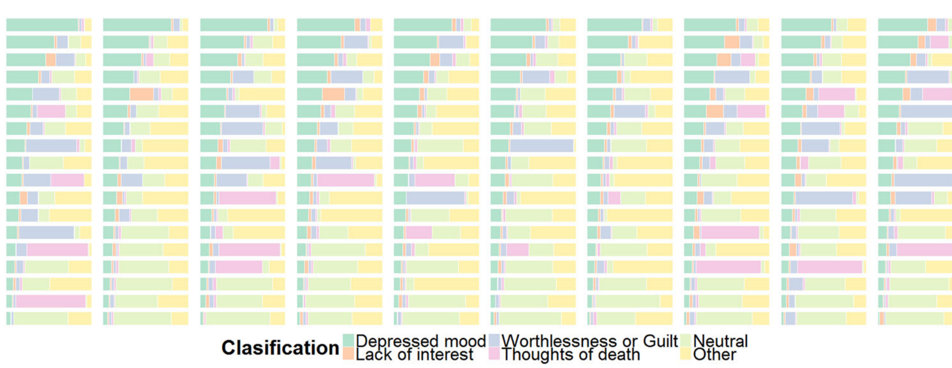
As can be seen in this deeper analysis, the visualizations could help to improve detection of possible depression cases on Twitter users, mainly in profiles where the users post a significant amount of unrelated tweets, leading to lower scores in a machine algorithm. Another possibility is that this kind of analysis could help in the detection in early stages, due to its relatively simple way to visualize change in patterns.

5 Conclusions

Based on psychological theory, we created a crowd-based labeling method for humans to classify tweets. We analyzed these tweets in relation to the original psychological theory classification to determine the efficacy of each of the categories for indicating potential onset of depression signals. The visual distribution of how the classifier behaved provided insight on how it can help in the detection and characterization of depression in a given set of tweets. This approach is promising for testing and validating analyses of the sentiment and emotion of complex human short texts.

By identifying more specific categories or information from the tweets, and moving from traditional positive or

Fig. 13 Sentiment distribution from user B



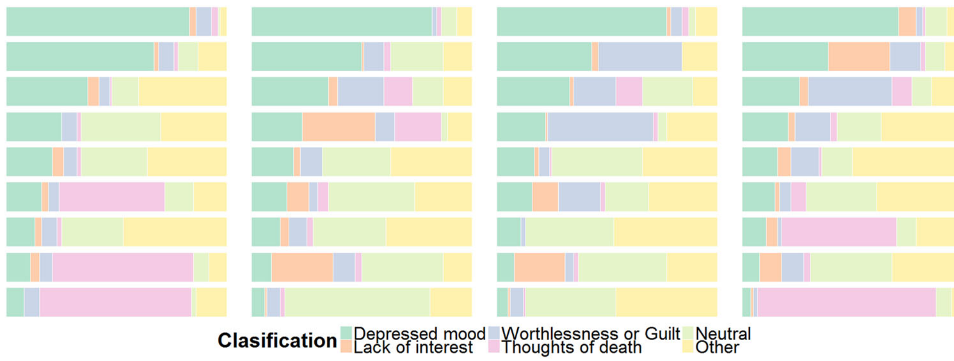


Fig. 16 Sentiment Distribution from user C

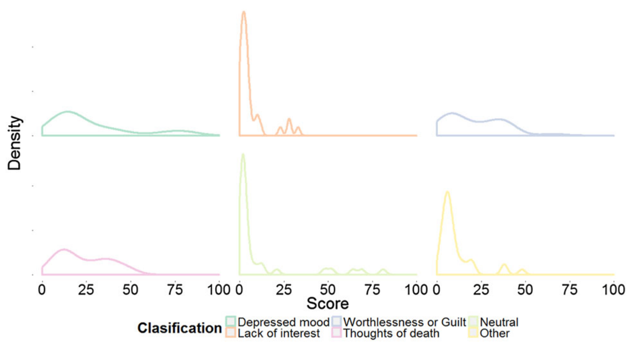


Fig. 17 Sentiment distribution from user C (not filtered)

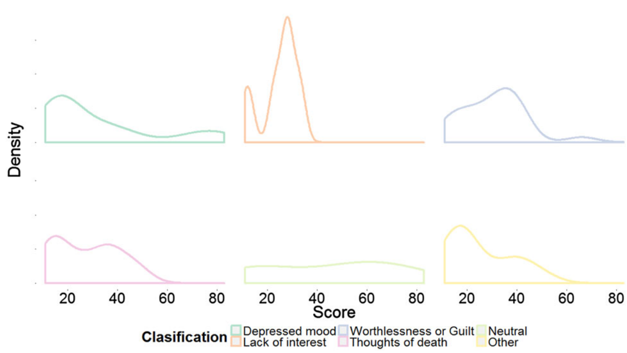


Fig. 18 Sentiment density from user C (filtered)

negative sentiment from tweets, we can be better able to understand crowd-based emotional state and draw a better conclusions. Moreover, providing specific patterns and numbers of the frequency of these categories that are used to diagnose a person with depression or not could help not only for detection in early stages, but also to a psychologist by giving them a more objective observation, instead of relying on subjective comments from their patient.

There are several possible next steps in this work. The first is to extend the data collection stage so that multiple months of data can be collected and analyzed. The second is refining the demographics, if possible, studying the efficacy of the analysis on men versus women, different age-groups, and populations. The third is to train LSTM RNNs to classify tweets using these categories and not just positive or negative. The final step is to extend the methodology proposed in this paper to other domains such as project management, where, for instance, we could examine the communications from participants with the goal of predicting the eventual success or failure of the project.

References

1. Association, A.P.: Diagnostic and Statistical Manual of Mental Disorders, 5th edn. American Psychiatric Association, Washington (2013)
2. Beck, A., Ward, C., Mendelson, M., Mock, J., Erbaugh, J.: An inventory for measuring. Arch. Gen. Psychiatry **4**, 561–571 (1961)
3. Becker, H., Naaman, M., Gravano, L.: Beyond Trending Topics: Real-World Event Identification on Twitter. Technical reports, Columbia University (2011)
4. Bhattacharya, S., Tran, H., Srinivasan, P.: Discovering health beliefs in twitter. In: AAAI Fall Symposium: Information Retrieval and Knowledge Discovery in Biomedical Text. Arlington, Virginia, USA (2012)
5. Bouazizi, M., Ohtsuki, T.: A pattern-based approach for multi-class sentiment analysis in twitter. IEEE Access **5**, 20617–20639 (2017)
6. Brogueira, G., Batista, F., Carvalho, J.P.: A smart system for twitter corpus collection, management and visualization. Int. J. Technol. Hum. Interact. (IJTHI) **13**(3), 13–32 (2017)
7. Culotta, A.: Estimating county health statistics with twitter. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1335–1344. ACM Press, Toronto, ON, Canada (2014). https://doi.org/10.1145/2556288.2557139
8. De Choudhury, M.: Can social media help us reason about mental health? In: Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion, pp. 1243–1244. International World Wide Web Conferences Steering Committee, Seoul, Republic of Korea (2014)
9. De Choudhury, M.: You're happy, i'm happy: diffusion of mood expression on twitter. In: Proceedings of HCI Korea, pp. 169–179. Hanbit Media, Inc., Seoul, Republic of Korea (2014)
10. De Choudhury, M., Counts, S., Horvitz, E.: Social media as a measurement tool of depression in populations. In: Proceedings of the 5th Annual ACM Web Science Conference, pp. 47–56. ACM, Paris, France (2013)
11. De Choudhury, M., Counts, S., Horvitz, E.J., Hoff, A.: Characterizing and predicting postpartum depression from shared facebook data. In: Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, pp. 626–638.

ACM Press, Baltimore, Maryland, USA (2014). https://doi.org/10.1145/2531602.2531675
12. De Choudhury, M., Gamon, M., Counts, S., Horvitz, E.: Predicting depression via social media. ICWSM **13**, 1–10 (2013)
13. De Choudhury, M., Morris, M.R., White, R.W.: Seeking and sharing health information online: comparing search engines and social media. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1365–1376. ACM Press, Toronto, ON, Canada (2014). https://doi.org/10.1145/2556288.2557214
14. Doulamis, N.D., Doulamis, A.D., Kokkinos, P., Varvarigos, E.M.: Event detection in twitter microblogging. IEEE Trans. Cybern. **46**(12), 2810–2824 (2016)
15. Gill, J., Harrison, B., Wood, T.J., Ramnanan, C.J., Jalali, A.: Investigating the use of social networking tools among medical students. Educ. Med. J. **6**(4), e82–e86 (2014)
16. Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford **1**(12) (2009)
17. González Celis, A.L., Padilla, A.: Calidad de vida y estrategias de afrontamiento ante problemas y enfermedades en ancianos de ciudad de méxico. Universitas Psychologica **5**(3) (2006)
18. Hasan, M., Agu, E., Rundensteiner, E.: Using hashtags as labels for supervised learning of emotions in twitter messages. In: ACM SIGKDD Workshop on Health Informatics, New York, USA (2014)
19. Huang, M., Cao, Y., Dong, C.: Modeling rich contexts for sentiment classification with lstm. arXiv preprint arXiv:1605.01478 (2016)
20. Kobayashi, Y., Mozgovoy, M., Munezero, M.: Analysis of emotions in real-time twitter streams. Informatica **40**(4), 387 (2016)
21. Kogan, M., Palen, L., Anderson, K.M.: Think local, retweet global: Retweeting by the geographically-vulnerable during hurricane sandy. In: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, pp. 981–993. ACM Press, Vancouver, BC, Canada (2015). https://doi.org/10.1145/2675133.2675218
22. Kumar, M., Dredze, M., Coppersmith, G., De Choudhury, M.: Detecting changes in suicide content manifested in social media following celebrity suicides. In: Proceedings of the 26th ACM Conference on Hypertext & Social Media, pp. 85–94. ACM Press, Guzelyurt, TRNC, Cyprus (2015). https://doi.org/10.1145/2700171.2791026
23. Layman, L., Sigurdsson, G.: Using amazon's mechanical turk for user studies: eight things you need to know. In: 2013 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, pp. 275–278. IEEE, Baltimore, Maryland, USA (2013). https://doi.org/10.1109/ESEM.2013.42
24. Li, J., Wang, X., Hovy, E.: What a nasty day: exploring mood-weather relationship from twitter. In: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, pp. 1309–1318. ACM Press, Shanghai, China (2014). https://doi.org/10.1145/2661829.2662090
25. Lin, H., Tov, W., Qiu, L.: Emotional disclosure on social networking sites: the role of network structure and psychological needs. Comput. Hum. Behav. **41**, 342–350 (2014)
26. Lipizzi, C., Iandoli, L., Marquez, J.E.R.: Extracting and evaluating conversational patterns in social media: a socio-semantic analysis of customers reactions to the launch of new products using twitter streams. Int. J. Inf. Manag. **35**(4), 490–503 (2015)
27. Martinez, V.R., Mancilla, A., Gonzalez, V.M.: Tweeting over the border: an empirical study of transnational migration in San Diego and Tijuana. arXiv preprint arXiv:1507.05989 (2015)
28. Mok, W.T., Sing, R., Jiang, X., See, S.L.: Investigation of social media on depression. In: 2014 9th International Symposium on Chinese Spoken Language Processing (ISCSLP), pp. 488–491. IEEE, Singapore (2014). https://doi.org/10.1109/ISCSLP.2014.6936690
29. Nambisan, P., Luo, Z., Kapoor, A., Patrick, T.B., Cisler, R., others: Social media, big data, and public health informatics: Ruminating behavior of depression revealed through twitter. In: 2015 48th Hawaii International Conference on System Sciences (HICSS), pp. 2906–2913. IEEE, Kauai, Hawaii (2015)
30. O'Dea, B., Wan, S., Batterham, P.J., Calear, A.L., Paris, C., Christensen, H.: Detecting suicidality on twitter. Internet Interv. **2**(2), 183–188 (2015)
31. Peters, M.E., Ammar, W., Bhagavatula, C., Power, R.: Semi-supervised sequence tagging with bidirectional language models. arXiv preprint arXiv:1705.00108 (2017)
32. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors. In: Proceedings of the 19th International Conference on World Wide Web, pp. 851–860. ACM, Raleigh, NC, USA (2010)
33. Sueki, H.: The association of suicide-related twitter use with suicidal behaviour: a cross-sectional study of young internet users in Japan. J. Affect. Disord. **170**, 155–160 (2015)
34. Thelwall, M., Buckley, K., Paltoglou, G.: Sentiment in twitter events. J. Assoc. Inf. Sci. Technol. **62**(2), 406–418 (2011)
35. Tsugawa, S., Kikuchi, Y., Kishino, F., Nakajima, K., Itoh, Y., Ohsaki, H.: Recognizing depression from twitter activity. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 3187–3196. ACM Press, Seoul, Republic of Korea (2015). https://doi.org/10.1145/2702123.2702280
36. Tsugawa, S., Mogi, Y., Kikuchi, Y., Kishino, F., Fujita, K., Itoh, Y., Ohsaki, H.: On estimating depressive tendencies of twitter users utilizing their tweet data. In: Virtual Reality (VR), 2013 IEEE, pp. 1–4. IEEE, Orlando, Florida, USA (2013)
37. Wilson, M.L., Ali, S., Valstar, M.F.: Finding information about mental health in microblogging platforms: a case study of depression. In: Proceedings of the 5th Information Interaction in Context Symposium, pp. 8–17. ACM Press, Regensburg, Germany (2014). https://doi.org/10.1145/2637002.2637006. http://dl.acm.org/citation.cfm?doid=2637002.2637006
38. Yang, W., Mu, L.: Gis analysis of depression among twitter users. Appl. Geogr. **60**, 217–223 (2015)