Beyond Modelling: Understanding Mental Disorders in Online Social Media

Esteban Andrés Ríssola[1(✉)], Mohammad Aliannejadi[2], and Fabio Crestani[1]

[1] Università della Svizzera italiana, Lugano, Switzerland
{esteban.andres.rissola,fabio.crestani}@usi.ch
[2] Univiersity of Amsterdam, Amsterdam, The Netherlands
m.aliannejadi@uva.nl

Abstract. Mental disorders are a major concern in societies all over the world, and in spite of the improved diagnosis rates of such disorders in recent years, many cases still go undetected. Nowadays, many people are increasingly utilising online social media platforms to share their feelings and moods. Despite the collective efforts in the community to develop models for identifying potential cases of mental disorders, not much work has been done to provide insights that could be used by a predictive system or a health practitioner in the elaboration of a diagnosis.

In this paper, we present our research towards better visualising and understanding the factors that characterise and differentiate social media users who are affected by mental disorders from those who are not. Furthermore, we study to which extent various mental disorders, such as depression and anorexia, differ in terms of language use. We conduct different experiments considering various dimensions of language such as vocabulary, psychometric attributes and emotional indicators. Our findings reveal that positive instances of mental disorders show significant differences from control individuals in the way they write and express emotions in social media. However, there are not quantifiable differences that could be used to distinguish one mental disorder from each other.

1 Introduction

During the last decade, there has been an increasing research interest in the identification of mental state alterations through the exploitation of online digital traces. One of the main reasons is that the capabilities of public health systems to cope with the plethora of cases that emerge on a daily basis are certainly limited. However, the proliferation of online social media platforms is changing the dynamics in which mental state assessment is performed [7,23]. Individuals are using these platforms on a daily basis to share their thoughts as well as to disclose their feelings and moods [8].

Research on language and psychology has shown that various useful cues about an individuals' mental state (as well as personality, social and emotional

Work done while Mohammad Aliannejadi was affiliated with Università della Svizzera italiana (USI).

J. M. Jose et al. (Eds.): ECIR 2020, LNCS 12035, pp. 296–310, 2020.
https://doi.org/10.1007/978-3-030-45439-5_20

conditions) can be discovered by examining the patterns of their language use [6]. As a matter of fact, language attributes could act as indicators of the current mental state [22,25], personality [19,26] and even personal values [2,4]. The main reason, as argued by Pennebaker et al. [21], is because such latent mental-related variables are encoded in the words that individuals use to communicate.

The constraints dictated in reality, such as cost and time, make the efficient process of personal diagnosis unfeasible. Initiatives such as the Strategic Workshop on Information Retrieval in Lorne [11] (SWIRL) are already proposing the application of principles of core Information Retrieval for the development of decision-making systems applied to fields that years back were not easy to conceive or imagine. In particular, they highlight the potential for cross-disciplinary collaboration and impact with a number of scientific fields, including psychology. In this respect, the Early Risk Prediction on the Internet (eRisk) [14,15], as well as the Computational Linguistics and Clinical Psychology (CLPsych) [9] workshops were the first to propose benchmarks to bring together many researchers to address the automatic detection of mental disorders in online social media.

These initial efforts to address the automatic identification of potential cases of mental disorders in social media have mainly modelled the problem as classification [16]. Researchers participating in these workshops have examined a wide variety of methods to identify positive cases [27,30]; however, not much insight has been given as to why a system succeeds or fails. Moreover, the models and features used in those studies could be analysed and motivated more deeply. Therefore, we argue that even though achieving an effective performance is important, being able to track and visualise the development of the mental disorder also is. This means that an accurate system can be more useful if it provides a way of understanding the explanatory factors that lead to a certain decision.

For this reason, it is necessary to carry out experiments providing insights on how the use of language is distinctive among social media users suffering from mental disorders as well as between different disorders. Moreover, it is useful to find ways to better visualise such development. Thus, systems oriented at visualisation for risk-assessment and decision-making could be complemented with preliminary step-by-step directions for practitioners to identify high-risk individuals based on statistical and visual analyses.

In this work, we conduct a thorough study of various dimensions of language to characterise users affected by mental disorders. Also, we provide several methods for visualising the data in order to provide useful insights to psychologists. To this end, we first compare users affected by a particular disorder against control individuals. Secondly, we are interested to know whether different mental disorders share the same characteristics or they are clearly different in terms of the dimensions analysed. Our main research questions, therefore, are:

- **RQ1:** How different is the language of users with mental disorder compared to control individuals in online social media?
- **RQ2:** To what extent is the language of depression, anorexia and self-harm cases different in online social media?

– **RQ3:** How can language-specific and emotional information be visualised to be utilised by psychologists during the diagnosis process?

To the best of our knowledge, this is the first study where the expression of mental disorders in social media is analysed at this depth. Our main findings reveal that positive instances of mental disorder significantly differ from control individuals[1]. More interestingly, we discover that considering the dimensions of language analysed it is not possible to establish a difference between depression, anorexia and self-harm.

The remainder of the paper is organised as follows. Section 2 summarises the related work; Sect. 3 details the approach followed to answer the research questions; Sect. 4 describes the data used in this work; Sect. 5 presents the corresponding results and analyses; Sect. 6 concludes the work.

2 Related Work

The majority of the works in the area have been mostly focused on the automatic identification of mental disorders in social media. Here, we outline those where some effort have been devoted to better understanding the relationship between language and mental disorders in social media and are relevant to our work.

Park et al. [20] provided a preliminary study towards verifying whether online social media data were truly reflective of users' clinical depressive symptoms. To this end, they analysed the expression of depression among the general Twitter[2] population. Over a period of two months, they collected tweets which contained the word "depression". A subsequent analysis showed that depression was most frequently mentioned to describe one's depressed status and, to a lesser extent, to share general information about depression.

De Choudhury et al. [5] presented an early work on automatic depression detection by using crowd-sourcing to collect assessments from several Twitter users who reported being diagnosed with depression. They a built a depression lexicon containing words that are associated with depression and its symptoms.

Coppersmith et al. [9] used Twitter data to carry out an exploratory analysis to determine language features that could be useful to distinguish users experiencing various mental disorders from healthy individuals. Despite they were able to determine a set of useful features, they observed that language differences in communicating about the different mental health problems remain an open question.

Gkotsis et al. [12] analysed various mental disorder communities on Reddit[3] (better known as *subreddits*[4]) to discover discriminating language features

[1] In this work, the term *positive* refers to subjects who have been diagnosed with depression, anorexia or self-harm. While, *control* refers to individuals not affected by any of the aforementioned mental disorders.

[2] See: https://twitter.com.

[3] See: https://www.reddit.com.

[4] Titled forums on Reddit are denominated *subreddits*.

between the users in the different communities. They found that, overall, the subreddits that were topically unrelated had condition-specific vocabularies as well as discriminating lexical and syntactic characteristics. Such study of Reddit communities might not result in accurate discrimination between users affected by mental disorders and healthy individuals. The main reason is that many of the participants of such specific forums are individuals concerned about the disorder because they had a close relative or friend suffering from it. We are interested in studying user's language features regardless of the topic discussed.

Overall, the presented works are concerned about the ability to predict whether users in online media platforms are positive instances of a mental disorder. Little effort has been devoted to understanding and providing insight and measuring the attributes which differentiates users affected by mental disorders from healthy individuals as well as between diverse mental disorders.

3 Objectives and Method

In this section we describe how we design the experiments to study social media posts in order to answer the research questions posed in the introduction. We outline what can be learned from each experiment, focusing on the language of diagnosed subjects and how their differences can be quantified.

3.1 Open Vocabulary

Vocabulary Uniqueness: One variable we analyse to answer **RQ1** is the similarity and diversity of the unique sets of words which compose the vocabulary of positive and control classes. Analysing such dimension tell us up to which extent classes have a common vocabulary and which words, if any, could be specifically used by users belonging to a certain class.

Considering each vocabulary as a set, we inspect the relative size of the their intersection. To this end, we use Jaccard's index to measure the similarity between finite sample sets. Formally, let P be the unique set of words obtained from positive users, *e.g.* self-harm, and C be the unique set of words obtained from control users. We compute Jaccard's index as follows:

$$J(P, C) = |P \cap C|/|P \cup C|.$$

As we see, the index gives us the ratio of the size of the intersection of P and C to the size of their union. The index ranges from 0 to 1, where an index of 1 indicates that the sets completely intersect, and thus, have the same elements. As the value approaches to 0 the sets are a more diverse among themselves.

Word Usage: An important aspect when studying the language of different groups, in addition to vocabulary similarities and differences, is to understand the patterns of word usage. Here, we attempt to answer **RQ1**, **RQ2**, **RQ3** by computing and comparing the language models for each class. The goal of this

analysis is to quantify the differences that might emerge between the classes in terms of the probability of using certain words more than others.

Language models are processes that capture the regularities of language across large amounts of data [10]. In its simplest form, known as a unigram language model, it is a probability distribution over the terms in the corpus. In other words, it associates a probability distribution of occurrence with every term in the vocabulary for a given collection. In order to estimate the probability for a word w_i in a document D in a collection of documents S we use

$$P(w_i|D) = (1 - \alpha_D)P(w_i|D) + \alpha_D P(w_i|S),$$

where α is a smoothing coefficient used to control the probability assigned to out-of-vocabulary words. In particular, we use the linear interpolation method[5] where $\alpha_D = \lambda$, *i.e.*, a constant. To estimate the probability for word w_i in the collection we use $s_{w_i}/|S|$, where s_{w_i} is the number of times a word occurs in the collection, and $|S|$ is the total number of words occurrences in the collection. In this work, D identifies all the documents in a specific class, *i.e.*, we concatenate all the documents of a particular class such as self-harm. While S is the union of all the documents of two classes in a corpus, *i.e.*, positive and control.

Once we computed the language models for each class, we plot the probability distributions obtained and analyse to which extent the distributions differ. Furthermore, we support our observations by computing the Kullback-Leibler divergence (KL), a well-known measure from probability theory and information theory used to quantify how much two probability distributions differ. In essence, a KL-divergence of 0 denotes that the two distributions in question are identical. The KL-divergence is always positive and is larger for distributions that are more different. Given the *true* probability distribution P and control distribution C, the KL-divergence is defined as:

$$KL(P||C) = \sum_x P(x) log \frac{P(x)}{C(x)}.$$

3.2 Psychometric Attributes and Linguistic Style

A common method for linking language with psychological variables involves counting words belonging to manually-created categories of language [5,6,9]. Conversely to the experiment described in Sect. 3.1, such method is known as "closed vocabulary" analysis [28]. In essence, we address **RQ1**, **RQ2** and **RQ3** by studying "function words[6]", and topic-specific vocabulary. On the one hand, the goal of conducting such analysis is to quantify specific stylistic patterns that could differentiate positive instances of a mental disorder from control individuals. For example, individuals suffering from depression exhibit a higher tendency

[5] Also referred to as Jelinek-Mercer smoothing.

[6] A *function word* is a word whose purpose is to contribute to the syntax rather than to the meaning of the sentence.

to focus on themselves [3], and thus, it is expected that the use of personal pronouns such as "I" would be higher. On the other hand, certain positives classes might exhibit a higher use of specific topically-related words. As we show later for the case of anorexia when compared to depression and self-harm.

It should be noted that we decide to keep the stop-words since many words such as pronouns, articles and prepositions reveal part of people's emotional state, personality, thinking style and connection with others individuals [6]. As a matter of fact such words, called *function words*, account for less than one-tenth of a percent of an individual's vocabulary but constitute almost 60 percent of the words a person uses [6].

The *Linguistic Inquiry and Word Count* [29] (LIWC)[7], provides mental health practitioners with a tool for gathering quantitative data regarding the mental state of patients from the their writing style. In essence, LIWC is equipped with a set of dictionaries manually constructed by psychologist which covers various psychologically meaningful categories and is useful to analyse the linguistic style patterns of an individual's way of writing. In our study, we measure the proportion of documents from each user that scores positively on various LIWC categories (*i.e.*, have at least one word from that category). In particular, we choose a subset of the psychometric categories included in LIWC where we found significant differences between positive and control users. Subsequently, we plot the distributions obtained using box-plots and compare them.

3.3 Emotional Expression

Individuals usually convey emotions, feelings, and attitudes through the words they use. For instance, gloomy and cry denote sadness, whereas delightful and yummy evoke the emotion of joy. Here, we address **RQ1**, **RQ2**, and **RQ3** by studying how individuals, suffering from mental disorders, emotionally express themselves in their social media posts. Furthermore, we investigate how such emotional expression could differentiate between affected and non-affected users.

We utilise the emotion lexicons built by Mohammad et al. [17,18] where each word is associated with the emotions it evokes to capture word-emotion connotations. In addition to common English terms, the lexicons include words that are more prominent in social media platforms. Moreover, they include some words that might not predominantly convey a certain emotion and still tend to co-occur with words that do. For instance, the words *failure* and *death* describe concepts that are usually accompanied by sadness and, thus, they denote some amount of sadness.

4 Data

Here, we study various collections released at different editions of the eRisk workshop [14,15]. The main goal of the workshop is to provide a common evaluation framework for researchers to address the early identification of depression,

[7] See: http://liwc.wpengine.com/.

anorexia and self-harm. The collections consist of a set of documents posted by users of Reddit, consisting of two groups of users. Positive cases of a particular mental disorder, such as anorexia, as well as control individuals. We choose to conduct our study using these collections since they have been developed and used through the various editions of eRisk and, therefore, have been extensively curated and validated. Furthermore, they are publicly available for research.

Following the methodology proposed by Coppersmith et al. [7], users of the positive class (*i.e.*, depression, anorexia or self-harm) were gathered by retrieving self-expressions of diagnoses (*e.g.*, the sentence "I was diagnosed with depression") and manually verifying if they contained a genuine statement of diagnosis. Control users were collected by randomly sampling from a large set of users available in the platform. The maximum number of posts per user is 2,000. A summary of the eRisk's collections studied in this work is shown in Table 1.

Table 1. Summary of eRisk's collections. The activity period represents the number of days passed from the first to last the document collected for each user. On average, a user's corpus spans over a period of roughly one year and half. The oldest documents in the collections date from the middle of 2006, while the latest ones are from 2017.

	Depression		Anorexia		Self-harm	
	Positive	Control	Positive	Control	Positive	Control
# of subjects	214	1,493	61	411	41	299
# of documents	89,999	982,747	24,776	227,219	7,141	161,886
Avg. # of documents/subject	420.5	658.2	406.16	552.84	174.17	541.42
Avg. # words/document	45.0	35.3	64.6	31.4	39.3	28.9
Avg. activity period (days)	≈658	≈661	≈799	≈654	≈504	≈785

5 Results and Analyses

In this section we present the results obtained from the experiments outlined in Sect. 3 on the various collections described in Sect. 4. Moreover, we analyse the corresponding outcomes towards answering the proposed research questions.

Our analyses mainly focus on language and its attributes. Nonetheless, we would like to highlight certain differences observed in terms of social engagement behaviour between the different groups.

Basic Statistics. Table 1 shows various statistics of different eRisk's collections. Interestingly, we note that among all the collections, an average positive user generates less documents than an average control user. However, as we see the average length of the documents is longer for the positive cases. In particular, it is interesting to highlight the case of users who suffer from anorexia. They even write longer documents than users affected by depression and self-harm.

Finally, it is worth noting that we spotted no meaningful differences among the various control groups. This observation is repeated for various experiments and is expected since each control group is a random sample of Reddit posts at different temporal periods. Therefore, each control group is a representative sample of users on Reddit.

5.1 Open Vocabulary

Vocabulary Uniqueness: Table 2 compares the vocabulary of the different groups of positive users (*i.e.* depression, anorexia and self-harm) against that of control users. We analyse their union, intersection and difference. We observe that positive cases of depression and anorexia exhibit more similarity to their respective control groups with a Jaccard index of 59% and 65%, respectively. Self-harm positive cases, on the other hand, use a more diverse set of words in their posts when compared to the control group (44%).

Moreover, the vocabulary size of the various positive groups gives us an idea about the words that have never been used by control users, but used by the positive groups. Among the terms that are unique to the positive groups, we find the following ones interesting: selfharm, trazodone[8] (Depression); anorexics, depersonalization[9], emetrol, pepto[10] (Anorexia).

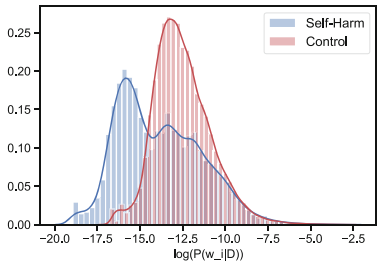
Table 2. Vocabularies comparison between positive and control users. KL-divergence computed across the language models obtained for the documents of positive and control users. As a reference, the KL-Divergence is also calculated between the different control groups. For instance, we observe an average divergence of 0.08 between the control group of the depression dataset and the other two control groups (*i.e.*, self-harm and anorexia).

	Depression	Anorexia	Self-harm
# of unique words positive	41,986	21,448	11,324
# of unique words control	70,229	31,980	25,091
Jaccard's index (positive vs. control)	0.59	0.65	0.44
Difference size (positive vs. control)	218	229	49
Difference size (control vs. positive)	28,461	10,761	13,816
KL (postive||control)	0.18	0.18	0.18
KL (control||positive)	0.21	0.31	0.20
KL (control||control)	0.08	0.07	0.10

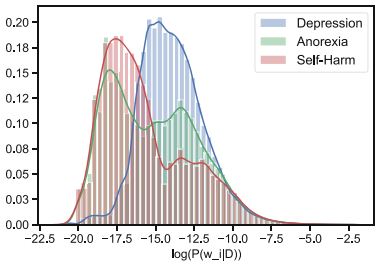
[8] *Trazodone* is an antidepressant medication.

[9] Depersonalization is a mental disorder in which subjects feel disconnected or detached from their bodies and thoughts.

[10] *Emetrol*, and *Peptol* are medications used to treat discomfort of the stomach.



(a) Self-harm vs. control



(b) Anorexia vs. depression vs. self-harm

Fig. 1. Language models probability distribution comparison (best viewed in colour). (Color figure online)

Word Usage: Figure 1 compares the language models obtained for the different classes. Note that the smoothing is necessary since, as shown before, there are terms which are present only in the positive class vocabulary but not in the control one and vice-versa. Figure 1(a) contrasts the language model of self-harm users against control individuals. We note that there are clear differences in terms of language use.

This observation is supported by the computation of KL-divergence. Table 2 shows the value of KL-divergence computed across the language models obtained for the documents of positive and control users. We note that the KL-divergence confirms the difference between the positive and control language models observed in the plots. We note similar patterns for depression and anorexia. In fact, as we compare the distribution of different control groups, we observe smaller KL-divergence values (0), indicating that these distributions are very similar.

Finally, comparing the language models of the three positive classes as in Fig. 1(b) is harder to identify noticeable differences between the distributions. Depression and anorexia language models are rather similar. While the largest noticeable difference is observed for self-harm when compared with either depression or anorexia. This reinforces the idea that the word probability distribution between the different positive classes is very similar, and thus, the way they use words. In this way, a system could compare the language model of a patient with both control and positive groups to provide the psychologists assistance in determining whether they are positive or not. The psychologist can further examine the patient to determine which disorder they are diagnosed with.

5.2 Psychometric Attributes and Linguistic Style

Using selected set of categories[11] from LIWC, we demonstrate that language use of Reddit users, as measured by LIWC, is statistically significantly different

[11] For a comprehensive list of LIWC categories see: http://hdl.handle.net/2152/31333.

between positive and control individuals. Figure 2 shows the proportion of documents from each user that scores positively on various LIWC categories (*i.e.*, have at least one word from that category). Selected categories includes function words (like pronouns and conjunctions), time orientation (like past focus and present focus) and emotionality. Bars are coloured according to the positive and control classes they represent.

The most remarkable case is the difference found in the use of the pronoun "I" between positive and control users, which in the case of depression replicates previous findings for other social media platforms [5,9]. Moreover, the proportion of messages using words related with positive emotions (*posemo*) is larger than negative (*negemo*) ones, even for the positive classes. Such circumstance could be related to the fact that English words, as they appear in natural language, are biased towards positivity [13]. Except for categories *we* and *she/he* differences reach statistical significance using Welch two sample t-test (p-value < 0.001) from each corresponding control group in Fig. 2(a). No significant differences between depression, anorexia and self-harm were found for any of the LIWC categories analysed in Fig. 2(a).

Figure 2(b) depicts categories related to *biological process*. In essence, this category includes words directly associated with the body and its main functions. We note that individuals within the anorexia group show a certainly different behaviour compared with depression and self-harm groups. Intuitively, this result is expected, given that anorexia is characterised by an intense fear of gaining weight and a distorted perception of weight. Overall, people with anorexia place a high value on controlling their weight and shape, using extreme efforts that tend to significantly interfere with their lives. Therefore, it is reasonable that such individuals talk more about themes related with their body and its function. Statistical significance (p-value < 0.001) between individuals suffering from anorexia and those affected by depression is achieved for categories *body*, *health* and *ingest*. While only for the latter category the differences are statistically significant when comparing self-harm and anorexia users.

In addition to the default categories that LIWC includes, we study other domain-specific lexicons. The first of them is the well-known depression lexicon[12] released by Choudhury et al. [5]. It consists of words closely associated with texts written by individuals discussing depression or its symptoms in online settings. The use of such words is not significantly different with respect to anorexia and self-harm groups. This suggest that those words are also frequently used by users affected by anorexia or self-harm. The same situation was observed when considering the set of absolutist terms[13] derived from the work of Al-Mosaiwi et al. [1], who concluded that the elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation.

[12] Examples of words included in the lexicon: *insomnia*, *grief*, *suicidal*, *delusions*.
[13] Examples of absolutist terms: *absolutely*, *constantly*, *definitely*, *never*.

5.3 Emotional Expression

Figure 3 depicts for each class the average number of documents that contain at least one word associated with a particular emotion, including the polarity (positive or negative). We note considerable differences on the expression of emotions between positive and their respective controls. One way to interpret such results is that on average positive users tend to share emotions more regularly than control individuals.

Moreover, we also analyse the frequency correlation of the different emotions for each class. Due to space constraints, we only include self-harm. However, similar observations arise from the remaining positive classes. We note that certain emotions show different correlations depending on the class under observation. For instance, in the control class *surprise* reveals a larger positive correlation with *trust*, *joy* and *positive* and *negative* orientation when compared with self-harm class. Conversely, with *surprise* and *disgust* as well as with *fear* and *disgust*. Interesting to note is that in the case of depression, *sadness* exhibits a negative correlation with *joy* and *positive* orientations. Such correlation does not hold for the corresponding controls. This kind of study allows to better understand how different emotional patterns emerge from the use of emotions (Fig. 4).

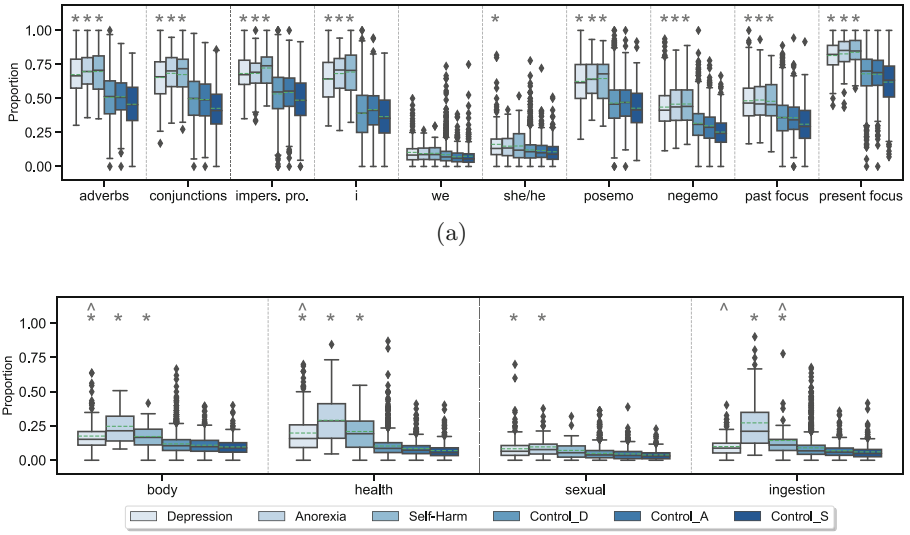


Fig. 2. Box and whiskers plot of the proportion of documents each user has (y-axis) matching various LIWC categories. Statistically significant differences between each positive and their respective control groups are denoted by ***** (p-value < 0.001). Also, statistically significant differences between Depression/Self-Harm and Anorexia are denoted by ˆ (p-value < 0.001).

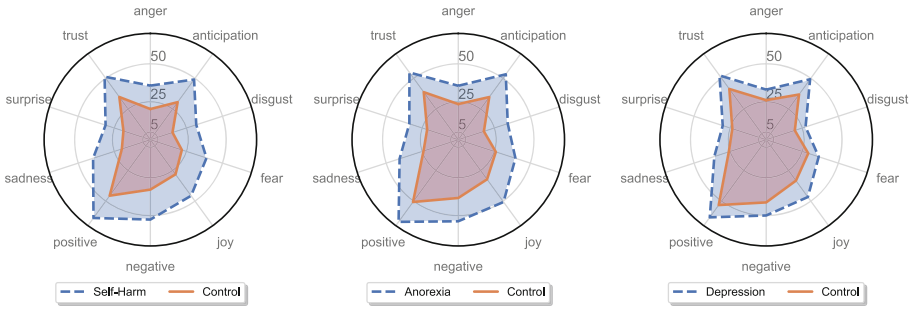


Fig. 3. Radar plot representing the average number of documents that contain at least one word associated with a particular emotion, including the polarity (positive or negative) for each positive group and its respective control.

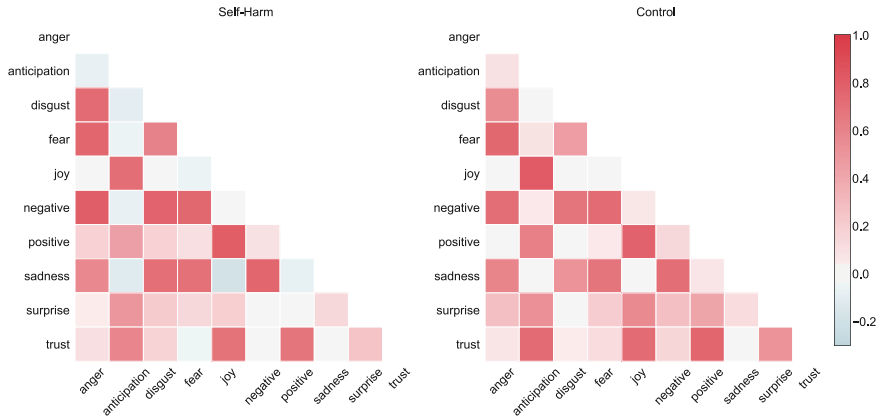


Fig. 4. Heatmap depicting the frequency correlation of the different emotions for self-harm (left) and control (right) groups (best viewed in colour). (Color figure online)

6 Conclusions

The wealth of information encoded in continually-generated social media is eager for analysis. In particular, social media data naturally occurs in a non-reactive way becoming a valuable complement for more conventional methods (such as questionnaires) used to determine the potential presence of mental disorders.

In this work, we reported results from a thorough analysis to show how users affected by mental disorders differ significantly from control individuals. We investigated the writing style, as well as how people express their emotions on social media via visualising certain probabilistic attributes.

To this aim, we analysed and visualised the activity, vocabulary, psychometric attributes, and emotional indicators in people's posts. Studying and visualising such dimensions, we discovered several interesting differences that could

help a predictive system and a health practitioner to determine whether someone is affected by a mental disorder. Across different mental disorders, however, we could not find any significant indicators. Therefore, we can conclude that analysing social media posts could help a system identify people that are more likely to be diagnosed a mental disorder. On the other hand, determining the exact disorder is a much more difficult task, requiring expert judgement. Also, we found that psychometric attributes and emotional expression provides a quantifiable way to differentiate between individuals affected by mental disorders from healthy ones. The study we presented in this work has high practical impact since research should be steered towards building new metrics that can correlate with a disease before traditional symptoms arise and which clinicians can use as leading indicators of traditional later-onset symptoms.

For the future, we are interested in investigating whether the findings also hold in other social media, such as Twitter, where users are restricted to other types of constraints, such as space limitations. This could prove that the language use by individuals affected by mental disorders is independent from the social media platform they participate. Also, it would be interesting to uncover the differences in people's behaviour on different social media platforms. Based on the restrictions, objectives, and features people tend to behave differently on different platforms. Therefore, it is of high importance to see if our findings can be generalised to other social media platforms or not. Also, studying other modalities of data such as video and image can be very effective in detecting people with mental disorders [24].

Acknowledgements. We thank the anonymous reviewers for the constructive suggestions. This work was supported in part by the Swiss Government Excellence Scholarships and Hasler Foundation.

References

1. Al-Mosaiwi, M., Johnstone, T.: In an absolute state: elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. Clin. Psychol. Sci. **6**(4), 529–542 (2018)
2. Aliannejadi, M., Crestani, F.: Venue suggestion using social-centric scores. In: Proceedings of ECIR Workshop on Social Aspects in Personalization and Search (2018)
3. Association, A.P.: Diagnostic and Statistical Manual of Mental Disorders, 5th edn. American Psychiatric Publishing, Washington (2013)
4. Boyd, R.L., Wilson, S.R., Pennebaker, J.W., Kosinski, M., Stillwell, D.J., Mihalcea, R.: Values in words: using language to evaluate and understand personal values. In: Proceedings of the Ninth International Conference on Web and Social Media, ICWSM 2015, Oxford, UK, pp. 31–40 (2015)
5. Choudhury, M.D., Gamon, M., Counts, S., Horvitz, E.: Predicting depression via social media. In: Proceedings of the Seventh International Conference on Weblogs and Social Media, ICWSM 2013, Cambridge, USA (2013)
6. Chung, C., Pennebaker, J.: The psychological functions of function words. In: Fiedler, K. (ed.) Social Communication. Frontiers of Social Psychology. Psychology Press, New York (2007)

7. Coppersmith, G., Dredze, M., Harman, C.: Quantifying mental health signals in Twitter. In: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, Baltimore, USA (2014)
8. Coppersmith, G., Dredze, M., Harman, C., Hollingshead, K.: From ADHD to SAD: analyzing the language of mental health on Twitter through self-reported diagnoses. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. Association for Computational Linguistics (2015)
9. Coppersmith, G., Dredze, M., Harman, C., Hollingshead, K., Mitchell, M.: CLPsych 2015 shared task: depression and PTSD on Twitter. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, Denver, USA (2015)
10. Croft, B., Metzler, D., Strohman, T.: Search Engines: Information Retrieval in Practice, 1st edn. Addison-Wesley Publishing Company, Boston (2009)
11. Culpepper, J.S., Diaz, F., Smucker, M.D.: Research frontiers in information retrieval: report from the third strategic workshop on information retrieval in Lorne (SWIRL 2018). SIGIR Forum **52**(1), 34–90 (2018)
12. Gkotsis, G., et al.: The language of mental health problems in social media. In: Proceedings of the 3rd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality (2016)
13. Kloumann, I.M., Danforth, C.M., Harris, K.D., Bliss, C.A., Dodds, P.S.: Positivity of the English language. PLoS ONE **7**(1), 1–7 (2012)
14. Losada, D.E., Crestani, F., Parapar, J.: Overview of eRisk: early risk prediction on the internet. In: Conference and Labs of the Evaluation Forum. CEUR-WS.org (2018)
15. Losada, D.E., Crestani, F., Parapar, J.: Overview of eRisk 2019 early risk prediction on the internet. In: Crestani, F., Braschler, M., Savoy, J., Rauber, A., Müller, H., Losada, D.E., Heinatz Bürki, G., Cappellato, L., Ferro, N. (eds.) CLEF 2019. LNCS, vol. 11696, pp. 340–357. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-28577-7_27
16. Masood, R.: Adapting models for the case of early risk prediction on the internet. In: Azzopardi, L., Stein, B., Fuhr, N., Mayr, P., Hauff, C., Hiemstra, D. (eds.) ECIR 2019. LNCS, vol. 11438, pp. 353–358. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-15719-7_48
17. Mohammad, S.: Word affect intensities. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation, LREC 2018, Miyazaki, Japan (2018)
18. Mohammad, S., Turney, P.D.: Crowdsourcing a word-emotion association lexicon. Comput. Intell. **29**(3), 436–465 (2013)
19. Neuman, Y.: Computational Personality Analysis. Introduction, Practical Applications and Novel Directions. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-42460-6
20. Park, M., Cha, C., Cha, M.: Depressive moods of users portrayed in Twitter. In: Proceedings of the ACM SIGKDD Workshop on Healthcare Informatics (2012)
21. Pennebaker, J.W., Mehl, M.R., Niederhoffer, K.G.: Psychological aspects of natural language use: our words, our selves. Annu. Rev. Psychol. **54**(1), 547–577 (2003)
22. Preoţiuc-Pietro, D., et al.: The role of personality, age and gender in tweeting about mental illnesses. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality (2015)

23. Prieto, V.M., Matos, S., Alvarez, M., Cacheda, F., Oliveira, J.L.: Twitter: a good place to detect health conditions. PLoS ONE **9**(1), 1–11 (2014)
24. Reece, A.G., Danforth, C.M.: Instagram photos reveal predictive markers of depression. EPJ Data Sci. **6**(1), 15 (2017)
25. Ríssola, E.A., Bahrainian, S.A., Crestani, F.: Anticipating depression based on online social media behaviour. In: Cuzzocrea, A., Greco, S., Larsen, H.L., Saccà, D., Andreasen, T., Christiansen, H. (eds.) FQAS 2019. LNCS (LNAI), vol. 11529, pp. 278–290. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-27629-4_26
26. Ríssola, E.A., Bahrainian, S.A., Crestani, F.: Personality recognition in conversations using capsule neural networks. In: 2019 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2019, Thessaloniki, Greece, 14–17 October 2019, pp. 180–187 (2019)
27. Sadeque, F., Xu, D., Bethard, S.: Measuring the latency of depression detection in social media. In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM 2018 (2018)
28. Schwartz, H.A., et al.: Personality, gender, and age in the language of social media: the open-vocabulary approach. PLoS ONE **8**(9), e73791 (2013)
29. Tausczik, Y.R., Pennebaker, J.W.: The psychological meaning of words: LIWC and computerized text analysis methods. J. Lang. Soc. Psychol. **29**(1), 24–54 (2009)
30. Trotzek, M., Koitka, S., Friedrich, C.M.: Word embeddings and linguistic metadata at the CLEF 2018 tasks for early detection of depression and anorexia. In: Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum, Avignon, France, 10–14 September 2018 (2018)