Understanding Depression from Psycholinguistic Patterns in Social Media Texts

Alina Trifan(✉), Rui Antunes, Sérgio Matos, and Jose Luís Oliveira

IEETA/DETI, University of Aveiro, Aveiro, Portugal
alina.trifan@ua.pt

Abstract. The World Health Organization reports that half of all mental illnesses begin by the age of 14. Most of these cases go undetected and untreated. The expanding use of social media has the potential to leverage the early identification of mental health diseases. As data gathered via social media are already digital, they have the ability to power up faster automatic analysis. In this article we evaluate the impact that psycholinguistic patterns can have on a standard machine learning approach for classifying depressed users based on their writings in an online public forum. We combine psycholinguistic features in a rule-based estimator and we evaluate their impact on this classification problem, along with three other standard classifiers. Our results on the Reddit Self-reported Depression Diagnosis dataset outperform some previously reported works on the same dataset. They stand for the importance of extracting psychologically motivated features when processing social media texts with the purpose of studying mental health.

Keywords: Mental health · Depression · Social media · Machine learning · Psycholinguistic features

1 Introduction

Suicide ideation, anxiety and depression are some of the most spread mental health diseases among adolescents and young adults, with a little under 800 000 people dying by suicide each year [1,2]. Fortunately, in the recent years there has been an increasing acknowledgement of this reality and a better understanding of the importance of enabling young people to improve their mental resilience, from early stages on.

Communication is at the core of society and currently the written digital communication is one of the most popular forms of expressing ourselves. We use social networks to detail our activities or routines, to describe our feelings, mental states, hopes and desires [6]. Young adults suffering of mental illness are more likely to express themselves online, either through blogging, social networks or specific public forums [15]. As we write digitally more and more, these large

J. M. Jose et al. (Eds.): ECIR 2020, LNCS 12036, pp. 402–409, 2020.
https://doi.org/10.1007/978-3-030-45442-5_50

volumes of data can be processed automatically with the purpose of inferring relevant information about one's well-being, such as mental health status.

Prevention and early identification of mental health diseases by means that are complimentary to traditional medical approaches have the ability to mitigate the under-supply of mental health facilities by advancing different types of counseling or support for the ones in need, such as connecting a depressed person to resources or peer support when they most need it [10]. Using social data has yet another advantage with respect to the stigma associated to mental health screening, as such approaches can provide new opportunities for early detection and intervention and have the potential to open new insights on research of the causes and mechanisms of mental health [4,5].

In this work we address the challenge of identifying depressed users of the Reddit social platform. We present encouraging results that demonstrate that social data has the potential for complementing standard clinical procedures. We base our methodology on a combination of a tf-idf weighting scheme for bag of words features and a rule-based estimator, that takes into account several psycholinguistic features that characterize depressed users. Our goal is to assess the extent up to which standard classifiers take into consideration psycholinguistic patterns and if by specifically contemplating them in a classification pipeline we can obtain better results. The dataset in use was the one proposed by Yates et al. [25].

This paper is structured in 4 more sections. We present a background on the subject in Sect. 2, followed by the methods we have employed in Sect. 3. Detailed results are presented and discussed in Sect. 4. Finally, Sect. 5 concludes the paper.

2 Background

A large volume of written data in the form of accessible common language is available through social media. This attracted the attention of natural language processing researchers. Among them, those who study the language of individuals in relation to their mental health conditions. Social media data has been identified as an emerging opportunity for revolutionizing in-the-moment measures of a broad range of people's thoughts and feelings [16]. Several studies focusing on mental health understanding through social network data have been conducted using Twitter[1] texts. Coppersmith et al. [9] presented a method for gathering data for a range of mental illnesses along with proof-of-concept results that focus on the analysis of four mental disorders: post-traumatic stress disorder, depression, bipolar disorder, and seasonal affective disorder. Their ultimate goal was to enable the ethical discussion regarding the balance between the utility of such data and the privacy of mental health related information. Later on, Coppersmith et al. [10] released a Twitter dataset of users who have attempted suicide, matched by neurotypical control users. Language modeling techniques were employed to classify these users, along with open government

[1] https://twitter.com/.

data to identify quantifiable signals that can relate them to psychometrically validated concepts associated to suicide. Nadeem et al. [18] used the same dataset to predict Major Depressive Disorder among online personas based on a Bag of Words approach and several statistical classifiers. More recently, Vioulès et al. [24] combined natural language processing features with a martingale framework to detect Twitter posts containing suicide-related content. The results were comparable to traditional machine learning classifiers.

While the previous examples proved that even short texts from Twitter can provide some insight into the relation between language and mental health conditions, longer-form content is nowadays explored for further insight into this matter. Yates et al. [25] introduced a large Reddit[2] dataset of self-reported depressed users, matched by similar control users. The release of the dataset was coupled with some preliminary results on their classification. The same research group included exact temporal spans that relate to the date of diagnosis, in an attempt to show that this type of diagnosis is not static [17]. Another Reddit dataset recently released by Cohan et al. [8], based on which authors investigated extended self-diagnoses matching patterns derived from mental health-related synonyms with focus on nine different mental health conditions. De Choudhury et al. [13] applied a logistic regression classifier that led to high accuracy results in order to predict suicidal ideation in Reddit users. One of the first demonstration of suicidice risk assesment through Reddit posts, matched with clinical knowledge was very recently reported by Shing et al. [23].

Initiatives such as CLEF Early Risk[3] or CLPsych[4], just to name a few, emerged over the last years as a proof of the importance of this research interest. These projects fostered collaborative work on the topic of mental health and social data and push forward new discoveries and insights. As a practical outcome that these initiative encouraged, triaging content in online social networks or public forums enabled the identification of content that requires the attention of moderators to ensure that urgent content can be responded to more quickly and consistently. Over the last years, the focus of these shared tasks was the early identification of people with suicidal inclinations or people susceptible to depression.

3 Methods

The dataset used in this work is the one proposed by Yates et al. [25], publicly available based on a signed user agreement with emphasis on data protection and proper acknowledgements. In order to get an understanding on the impact of the previously described patterns of depressed users we experimented standard feature extraction methods, which we have complemented with the design of a rule-based estimator that solely relies on these psycholinguistic features. We have experimented several classifiers that have been identified in the literature

[2] https://www.reddit.com/.
[3] http://early.irlab.org/.
[4] http://clpsych.org/.

as appropriate for this classification task. All experiments were managed using the scikit-learn machine learning framework (release version 0.21) [20]. The texts in the dataset were curated for any direct link to mental health. We considered this curation relevant as it relates to the possibility of identifying people that are unaware of their mental health status through heterogeneous texts.

Dataset Description. The dataset consists of all Reddit users who made a post between January and October 2016, matching high-precision patterns of self-reported diagnosis (e.g. "I was diagnosed with depression"). The depressed users were matched by control users, who have never posted in a subreddit related to mental health and never used a term related to it. In order to avoid a straight-forward separation of the two groups, all posts of diagnosed users related to depression or mental health were removed. In the end, 9210 diagnosed users were matched by 107 274 control users. Each user in the dataset has an average of 969 posts (median 646) and the mean post length is 148 tokens (median 74).

Data Preprocessing. The preprocessing of the Reddit posts follows standard approaches in text classification. The posts are lowercased and tokenized, after removing all non-alphabetic characters. Stopwords are filtered, by using an altered version of the stopwords list of the Natural Language Toolkit[5]. The alteration consists of removing from the original stopwords list self-related words and words that belong to the list of absolutist words, as described next. We are not interested in discarding these words as they may convey valuable psycholinguistic content, as detailed in the following subsections.

3.1 Experiments

The dataset was originally split into similar size chunks of training, validation and test samples, each of them containing roughly posts of 39 000 users. Because of the large size of each of these chunks (7 GB), we explored both incremental and online training with the following three classifiers: Multinomial Naive Bayes (MNB), linear Support Vector Machine with Stochastic Gradient Descent (SGD) and Passive Aggressive (PA). For the out of core classification, we trained the classifiers with batches of 500 users data. The batch size is not expected to have an impact on the performance of the classifiers[6]. For each of these classifiers, we have performed a grid search over the validation dataset in order to identify the best parameters that characterize them.

The first approach followed a standard processing stream for text classification. We considered Bag of Words (BoW) features for the three classifiers and we applied counts and tf-idf based feature weighting. A further study into psycholinguistic literature revealed possible patterns in the language of depressed users, that we modelled as features of a rule-based estimator:

[5] https://www.nltk.org/.
[6] https://scikit-learn.org/stable/modules/computing.html.

Absolutist Words. A recent study on absolutist thinking, which is considered a cognitive distortion by most cognitive therapies for anxiety and depression, showed that anxiety, depression, and suicidal ideation forums contained more absolutist words than control forums [3]. The study, conducted as a text analysis of 63 Internet forums with over 6400 members resulted in a validation of an absolutist words dictionary, presented in Table 1. Their usage frequency was considered for the rule-based estimator.

Table 1. Absolutist words validated by Al-Mosaiwi et al. [3].

Absolutely	Constant	Every	Never
All	Constantly	Everyone	Nothing
Always	Definitely	Everything	Totally
Complete	Entire	Full	Whole
Completely	Ever	Must	

Analysis of Lexical Categories. Depressed users frequently use negative emotion words and anger words on social networks [12,19]. Empath [14] is a text categorization open-source software that analyzes text across 200 built-in, pre-validated categories that were generated from common topics in a web dataset. Empath's categories have been human validated and are highly correlated ($r = 0.906$) with similar categories in the Linguistic Inquiry and Word Count [21]. As depressed users tend to have an overall more negative connotation of their texts, we used Empath's lexical category (version release 0.41) of a user's overall set of posts for the rule based estimator.

Self-related Speech. Depressed users tend to use them more often self-related words (such as: I, me, myself, mine) [7,22].

Posts Length. Depressed and suicidal people tend to write more words than control users [11]. We consider this information relevant for an heuristic that takes into consideration the number of tokens of a user for the rule-based estimator.

4 Results

The statistical analysis of the training dataset revealed that on average, depressed users have 770 mentions of absolutist words in their writings, while the average mentions for a control user are 210. Posts belonging to depressed users contain 2888 self-related words, while posts of control users contain on average 716 of them. The average number of tokens for a control user is at 20 551 tokens, while for a depressed one reaches 69 000. The categorization of the posts by means of Empath has little impact on the final results, given that most posts express negative emotions. The most relevant results are summarized in

Table 2. Comparative results on detecting depressed Reddit users based on multiple approaches. The following abbreviations are considered: MNB = Multinomial Naive Bayes, PA = Passive Aggressive Classifier, SGD = Support Vector Machine with Stochastic Gradient Descent, RE = Rule-based Estimator.

Method	Prec.	Rec.	F1	Acc
Tf-idf MNB (alpha=1)	0.61	0.47	0.53	0.94
Tf-idf PA (loss=*sqrt_hinge*, tol=e^{-3}) batch	**0.82**	0.64	**0.72**	**0.96**
Tf-idf PA (loss=*sqrt_hinge*, tol=e^{-3}) online	0.64	0.64	0.64	0.94
Tf-idf SGD (l1=0.95, loss=*hinge*) batch	0.76	0.62	0.68	0.95
Tf-idf SGD (l1=0.95, loss=*hinge*) online	0.70	0.65	0.68	0.95
RE PA (loss=*sqrt_hinge*, tol=e^{-3})	0.63	0.13	0.22	0.95
Feature Union PA (loss=*sqrt_hinge*, tol=e^{-3})	0.68	**0.72**	0.70	0.95
[25] CNN	**0.75**	0.57	0.65	N/A
[25] FastText	0.37	**0.70**	0.49	N/A

Table 2. Apart from assessing each model separately, we considered a feature union of equal weights for tf-idf and the output of the rule-based estimator, combined with a Passive Aggressive classifier.

We achieve the best results in terms of precision when using tf-idf weighting with a Passive Aggressive classifier. One explanation for this result may reside in the fact that the Passive Aggressive classifier is an online classifier that learns sequentially. Batch training led to better results for SGD and PA. We believe this happens because the models are being updated along the way and are less prone to overfit. When taking into account the psycholinguistic features we manage to improve the results in terms of recall, while not losing too much precision.

5 Conclusions

Analysis of social media texts has the potential to provide methods for understanding a user's mental health status and for the early detection of possible related diseases. We have presented in this paper preliminary results on the use of hand-crafted psycholinguistic features as possible improvements to standard classification approaches of depressed online personas.

As future work, we are interested in extending these psycholinguistic features with others that can be further revealed by clinical publications. Moreover, we want to analyze statistical information about the weights that each of these feature has in the final classification results. We also plan to investigate the lexical variability of posts of depressed users so as to infer possible new insights on the matter. If an automated process can predict or detect depressive users, they can be targeted for further medical assessment or they could be provided with alternative means of support and treatment. A great outcome of automatic processing of social networking data is the detection of users that are unaware of their condition.

Acknowledgements. This work was supported by the Integrated Programme of SR&TD SOCA (Ref. CENTRO-01-0145-FEDER-000010), co-funded by Centro 2020 program, Portugal 2020, European Union, through the European Regional Development Fund. Rui Antunes is supported by the Fundação para a Ciência e a Tecnologia (PhD Grant SFRH/BD/137000/2018).

References

1. World Health Organization Mental Health. http://www.who.int/mental_health/en/. Accessed 10 Oct 2019
2. Mental health atlas (2017). (Geneva: World Health Organization, 2018)
3. 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., 2167702617747074 (2018)
4. Arseniev-Koehler, A., Mozgai, S., Scherer, S.: What type of happiness are you looking for? - A closer look at detecting mental health from language. In: Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, pp. 1–12 (2018)
5. Bruffaerts, R., et al.: Mental health problems in college freshmen: prevalence and academic functioning. J. Affect. Disord. **225**, 97–103 (2018)
6. Calvo, R.A., Milne, D.N., Hussain, M.S., Christensen, H.: Natural language processing in mental health applications using non-clinical texts. Nat. Lang. Eng. **23**(5), 649–685 (2017)
7. Chung, C., Pennebaker, J.W.: The psychological functions of function words. Soc. Commun. **1**, 343–359 (2007)
8. Cohan, A., Desmet, B., Yates, A., Soldaini, L., MacAvaney, S., Goharian, N.: SMHD: a large-scale resource for exploring online language usage for multiple mental health conditions. In: The 27th International Conference on Computational Linguistics (COLING 2018), pp. 1485–1497. ACL (2018)
9. 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, pp. 51–60 (2014)
10. Coppersmith, G., Leary, R., Whyne, E., Wood, T.: Quantifying suicidal ideation via language usage on social media. In: Joint Statistics Meetings Proceedings, Statistical Computing Section, JSM (2015)
11. Coppersmith, G., Ngo, K., Leary, R., Wood, A.: Exploratory analysis of social media prior to a suicide attempt. In: Proceedings of the Third Workshop on Computational Lingusitics and Clinical Psychology, pp. 106–117 (2016)
12. De Choudhury, M., Gamon, M., Counts, S., Horvitz, E.: Predicting depression via social media. ICWSM **13**, 1–10 (2013)
13. De Choudhury, M., Kiciman, E., Dredze, M., Coppersmith, G., Kumar, M.: Discovering shifts to suicidal ideation from mental health content in social media. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 2098–2110. ACM (2016)
14. Fast, E., Chen, B., Bernstein, M.S.: Empath: understanding topic signals in large-scale text. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 4647–4657. ACM (2016)
15. Gowen, K., Deschaine, M., Gruttadara, D., Markey, D.: Young adults with mental health conditions and social networking websites: seeking tools to build community. Psych. Rehabil. J. **35**(3), 245 (2012)

16. Guntuku, S.C., Yaden, D.B., Kern, M.L., Ungar, L.H., Eichstaedt, J.C.: Detecting depression and mental illness on social media: an integrative review. Curr. Opin. Behav. Sci. **18**, 43–49 (2017)
17. MacAvaney, S., et al.: RSDD-Time: temporal annotation of self-reported mental health diagnoses. In: Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, pp. 168–173 (2018)
18. Nadeem, M.: Identifying depression on Twitter. CoRR **abs/1607.07384** (2016)
19. Park, M., Cha, C., Cha, M.: Depressive moods of users portrayed in Twitter. In: Proceedings of the ACM SIGKDD Workshop on Healthcare Informatics (HI-KDD), vol. 2012, pp. 1–8. ACM, New York (2012)
20. Pedregosa, F., et al.: Scikit-learn: machine learning in python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
21. Pennebaker, J.W., Francis, M.E., Booth, R.J.: Linguistic inquiry and word count: LIWC 2001. Mahway: Lawrence Erlbaum Assoc. **71**(2001), 2001 (2001)
22. Rude, S., Gortner, E.M., Pennebaker, J.: Language use of depressed and depression-vulnerable college students. Cogn. Emot. **18**(8), 1121–1133 (2004)
23. Shing, H.C., Nair, S., Zirikly, A., Friedenberg, M., Daumé III, H., Resnik, P.: Expert, crowdsourced, and machine assessment of suicide risk via online postings. In: Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, pp. 25–36 (2018)
24. Vioulès, M.J., Moulahi, B., Azé, J., Bringay, S.: Detection of suicide-related posts in Twitter data streams. IBM J. Res. Dev. **62**(1), 1–7 (2018)
25. Yates, A., Cohan, A., Goharian, N.: Depression and self-harm risk assessment in online forums. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 2968–2978. Association for Computational Linguistics (2017)