UACH-INAOE participation at eRisk2017

Alan A. Farías-Anzaldúa[1], Manuel Montes-y-Gómez[2]
A. Pastor López-Monroy[3], and Luis C. González-Gurrola[1]

[1] Universidad Autonóma de Chihuahua, Mexico
alan.alexis.fa@gmail.com, lcgonzalez@uach.mx
[2] Instituto Nacional de Astrofísica, Óptica y Electrónica, Mexico
mmontesg@inaoep.mx
[3] University of Houston, USA.
alopezmonroy@uh.edu

Abstract. Being depression a mental disorder that affects 1 in 10 people world-wide, it is important to react in an effective way before fatal decisions could be taken. Moreover, with the rise of social media, some issues regarding how to identify patterns of depressed users in a timely fashion is an important task that is attracting the attention from the NLP community. Accordingly, the CLEF 2017 launched a challenge to identify depressed users in *reddit* forums. Our proposal to attend this task was based on a two-step classification procedure, where we first look at the post level to create basic features that were then applied at user level, to build a profile for each user. To evaluate this methodology, we applied a 10-fold cross validation on the training data, obtaining an F-score of 0.73 in the identification of depressed users. For the CLEF challenge, our team reached the fourth place, obtaining an F-score of 0.48.

Keywords: natural language processing, depression, social media

1 Introduction

According to the National Institute of Mental Health (NIMH), 1 in 10 adults will exhibit some kind of depression at some point in life [15]. People affected with depression suffer from feelings of hopelessness, loss of motivation, have sleep and eating disorders, and feel disconnected from other people. If untreated, depression can lead to suicide. Psychological criteria for detection of depression is presented in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), and consists of long-term (two or more years) behavioral analysis performed by a professional. This requires individuals to get diagnosed by a psychologist, something that many people are not willing or capable to do.

Online social platforms are quickly arising in popularity, becoming nearly ubiquitous in countries where internet is readily-available. These platforms allow people to share and express their thoughts and feelings freely and publicly with other people. As could be supposed, this information is a rich source for Natural Language Processing tasks such as: Opinion Mining [11], Sentiment Analysis [8]

or inferring mental health issues [6]. Following this last route, in this manuscript we approach the problem of predicting depressive users based on the analysis of posts that they share. Motivation for this task is that if depressive symptoms are accurately identified in a timely fashion, then, professionals could intervene before depression progresses. The approach that we introduce in this manuscript was submitted to the CLEF eRisk 2017 Challenge (Early Risk Prediction On The Internet). For this task, we were provided with a dataset [14] built from *reddit* posts, grouped into anonymized users and ordered chronologically. *reddit* is an online discussion forum that covers a wide variety of topics. Content is organized into areas of interest called *"subreddits"*, and it is there that threads are created and commented.

The proposed approach is based on a two-step classification procedure; the first step focuses on the analysis of individual posts by their content, whereas the second step considers the classification of users by their behaviors expressed by their kind of posts. The main idea behind our approach consists in modeling each user post, instead of directly modeling each user, under the assumption that each post is a window into the mind of a person in a particular point in time. Then, if we collect enough of these posts, we can observe how that person's mind changes - or do not - through time, in order to improve the model of each user. In our first set of experiments, we applied a 10-fold cross validation on the training data, obtaining a F1 score of 0.73 for predicting depressed individuals and above 0.90 for non-depressed individuals. For the challenge task, we ended up in the fourth place (out of eight teams) with an F-score of 0.48. Since our participation in the challenge was late, we only submitted predictions for the last part of the competition, so we would like to further analyze our model simulating their application at different levels of evidence for each user.

The organization of this paper is as follows. Section 2 presents some related work. In section 3 we describe our approach. Experimental Settings are presented in section 4, while results are presented in section 5. Finally, Conclusion is presented in section 6.

2 Related Work

Social media platforms have become an attractive place to infer health related issues for entire communities or individual users. A study presented in [5] compared online depression communities against control online communities, reporting that using topics and style markers as features showed good predictive power to differentiate between users. Two of the most common online platforms that have been used to mine this kind of data are *Facebook* and *Twitter*. In *Facebook* some efforts have been made to identify different contexts of depression. In [13], a study presents the case of predicting Postpartum Depression from shared data. While on [12], the study focused on evaluating depressive symptoms through an application developed for *Facebook*. In [4], the authors attempt to answer the question: *what public health information could be learned from Twitter?* Being Depression one of the most common ailments that could be associated to a cer-

tain group of words. One of the former works to characterize depressed users in Twitter is [3], where authors proposed some attributes to measure depressive behavior such as social engagement, language or emotion. The authors built a Support Vector Machine that achieved an score of 70% accuracy to predict, even prior to onset, depressive users. In [9], Minsu et al. followed a qualitative study to investigate perception differences between depressed and non-depressed twitter users. They found that depressive users were more prone to perceive Twitter as a tool for social awareness and emotional interactions. More recently, Nadeem et al. [7] compiled over 2.5M tweets to identify Major Depressive Disorder (MDD), with a reported accuracy of 81%. The authors claim that their method could even help estimate the risk of an individual being depressed. With a good number of studies focused on detecting depression, new works suggest the possibility of detecting other type of social phenomena like suicidal ideation [10]. To the best of our knowledge, no works have been done using *reddit* as a platform, being the dataset created by Losada and Crestani [14] the first attempt to fill this gap. We will use this dataset for experimentation.

3 The two-step classification approach

This section presents an overview of the proposed two-step classification approach. The first step considers the analysis of individual posts by their content. Its aim is to discriminate messages produced by depressed and non-depressed users. Then, a second step analyzes this information with the aim of modeling the respective users' behaviors based on their category. The following explains in detail these two steps.

3.1 Post Level Analysis

The main goal of the first step is to classify each post as likely produced by a *depressed* or *non-depressed* person. For this, we train a classifier using user posts considering that all posts inherit the label from their respective author. More formally, let $\mathcal{D} = \{(D_1, y_1), \ldots, (D_n, y_n)\}$ be a training set of labeled user-documents, that is, $\mathcal{D}$ is a collection of $n-$tuples of user-documents, i.e. posts, (D_i) and category-labels (y_i), where $y_i \in \{\text{depressed}, \text{non-depressed}\}$. Also let $D_i = \{(P_1, y_i), \ldots, (P_n, y_i)\}$ be the set of posts from user D_i. We represented each post P_j by a feature vector $\mathbf{P_j}$ that combines three different kinds of features: unigrams, bigrams and trigrams. For each individual feature space we considered the 10k top frequent items and then we selected all the features with information gain greater than zero. Additionally, we included two extra time attributes: the hour of the post and a binary attribute for whether or not it was posted on a weekend. Therefore, the final post representation $\mathbf{P_j}$ is expressed as follows:

$$\mathbf{P_j} = \langle \mathbf{P_j^{1gram}}, \mathbf{P_j^{2gram}}, \mathbf{P_j^{3gram}}, \mathrm{P_j^{time}}, \mathrm{P_j^{weekend}} \rangle \quad (1)$$

3.2 User Level Analysis

The intuitive idea behind analyzing posts individually is that users can be understood in terms of their behavior, rather than by their vocabulary. Hence, in this second step we build the user representation by considering some statistics obtained from the first classification step.

Given a document D_i, we first classify all its posts using the previous classifier (refer to Section 3.1). Then, we model D_i as the sequence of predicted labels for each post $P_j \in D_i$, that is, $D'_i = (y_1, \ldots, y_h)$, where y_i corresponds to the predicted label of post P_i. From this sequence of labels, we represent each user by a 12-dimensional vector $\mathbf{D_i}$ including the features described below. For simplicity we will refer to a post as being "depressive" (or "non-depressive") depending if it was generated by a depressed or non-depressed user.

1. Percentage of posts that were classified as "depressive".
2. Percentage of posts that were classified as "non-depressive".
3. Percentage of times that a "non-depressive" post is followed by another "non-depressive" post.
4. Percentage of times that a "non-depressive" post is followed by a "depressive" post.
5. Percentage of times that a "depressive" post is followed by a "non-depressive" post.
6. Percentage of times that a "depressive" post is followed by another "depressive" post.
7. Percentage of "depressive" posts written in the morning.
8. Percentage of "non-depressive" posts written in the morning.
9. Percentage of "depressive" posts written in the evening.
10. Percentage of "non-depressive" posts written in the evening.
11. Percentage of "depressive" posts written in the night.
12. Percentage of "non-depressive" posts written in the night.

Time periods were considered after analyzing heatmaps of users' activity, See Figure 1. Accordingly, the periods that we consider were morning [6am-2pm], ii) evening [2pm-10pm], and iii) night [10pm-6am].

4 Experimental Settings

Dataset. For the evaluation we use the dataset presented in [14]. This dataset is in XML format, where each user had a list of posts, and each post contained the following fields: title, date, info, and text. The info field simply contained the string "reddit post" and was discarded for not providing any relevant information. Title and text fields both contained useful information; they were merged into a "text" field.

Preprocessing. Uppercase letters were turned to lowercase, numbers were removed, and negations were joined to the following word (i.e., "don't want" returns a single token "don't_want").

Classification. For post and user classification we used a Naïve-Bayes Classifier (NBC). For validation we applied a stratified 10 fold-cross validation.

5 Results

To evaluate the appropriateness of the proposed approach to identify depressed users from online forums, we compared its performance against a traditional text classification approach using a BOW representation. The first part of Table 1 shows the results from this comparison; they indicate that the two-step classification approach outperformed by more than 40% the baseline results. In order to evaluate the usefulness of some components of the proposed approach we carried out some experiments disabling some of them. In particular, we ran an experiment without considering the time information for the post representation and other using only word unigrams. The last two rows from Table 1 present the results from these two experiments; they clearly show that word sequences and time information are key elements in the classification of depressed posts and therefore on the identification of depressed users.

Table 1. F_1 results: the proposed method, traditional text classification approach, and some variations of the method.

	F-Score
Proposed approach	0.9141
Traditional TC approach	0.7608
Proposed approach (without time information)	0.8508
Proposed approach (using only unigrams)	0.8316

An analysis on the relevance of the features showed us that the most discriminative n-grams were those related to the health condition of the users such as depression, diagnosis and therapy; this result was somehow expected due to the nature of the corpus. However, other n-grams related to feelings, news and interpersonal relationships were also highly discriminative for both kind of users. Table 2 shows the 40 most discriminative word n-grams.

Table 2. The 40 most discriminative n-grams

depression, pope, to_talk_to, my_depression, miserable, conservative
depression_, global, depresesion_and, therapist, gop, obama
depression_., anxiety, suicidal, meds, $_billion, not_alone
diagnosed, my_life, myself, depressed, when_i_was, conspiracy
diagnosed_with, because_i_have, boyfriend, feel_like_i
hollywood, iraq, like_i, report, president, cnn
makes_me_feel, of_my_life, isis, me_feel, vote, helped_me

We also analyze the posting behavior of the users. The heatmaps in Figure 1 show that there is a clear difference in the posting behaviors of both kinds of users, particularly on weekends. We expected depressed users to be more active during night, as people with depression also suffers from insomnia or has bad sleeping habits overall, but evidence showed us that it was the contrary, not depressed users were more active overall, except on weekends.

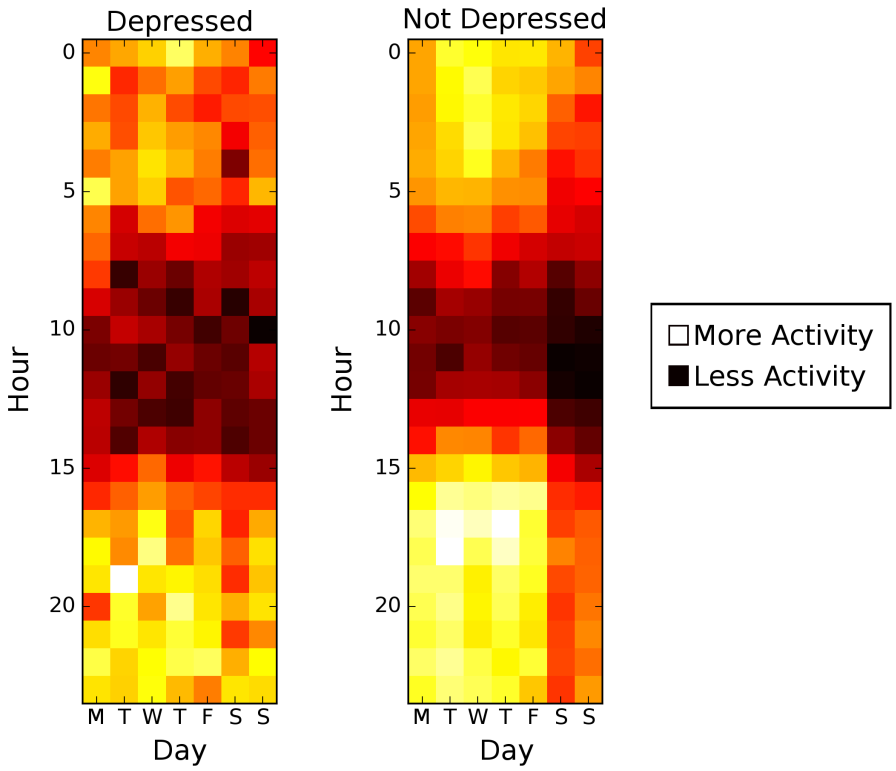

Fig. 1. Activity of both kinds of users during the week. Each square represents an hour.

6 Conclusion

In this paper we presented our approach to identify depression through online publications, which is based on a two step analysis. The analysis of individual posts to characterize users' behavior was very useful in a second step to analyze

the users from higher point of view. In this work, the general behavior of each user is captured by means of observing the posting history, especially the posting behavior and the posting period of time.

We presented experimental results of our method that shows to be useful compared with other methodologies. For this, we study the user-documents from different perspectives, including: post, and analysis of period of time. We found that both kinds of users can be differentiated by exploiting the time of publication, but more importantly, the changes in their behavior related to depressed and non-depressed posts. It was also found that both users have different activity during the weekend and during the day, which is valuable for future research paths.

Future work includes developing an early risk assessment strategy. We could not participate until late for this contest, and thus we did not develop an strategy for early risk assessment.

Acknowledgements. This work was supported by CONACYT under scholarship 735623/599448 and project grant PDCPN-2014-01-247870.

References

1. Clarke, F., Ekeland, I.: Nonlinear oscillations and boundary-value problems for Hamiltonian systems. Arch. Rat. Mech. Anal. 78, 315–333 (1982)
2. Clarke, F., Ekeland, I.: Solutions périodiques, du période donnée, des équations hamiltoniennes. Note CRAS Paris 287, 1013–1015 (1978)
3. Choudhury, M. , Gamon, M. , Counts, S. and Horvitz, E: Predicting depression via social media. Proceedings of the 7th International AAAI Conference on Weblogs and Social Media, Boston, MA, July 8-11 (2013)
4. Paul MJ, Dredze M.: You are what you Tweet: Analyzing Twitter for public health. ICWSM. 2011 Jul 17;20:265-72.
5. Nguyen, T.,Phung, D., Dao, B and Venkatesh, S and Berk, M: Affective and content analysis of online depression communities. IEEE Transactions on Affective Computing, 5(3):217226. (2014)
6. Coppersmith, G., Dredze, M., Harman, C., and Hollingshead, K. (2015). From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses. NAACL HLT 2015, 1.
7. Nadeem M, Horn M, Coppersmith G (2016) Identifying depression on Twitter. arXiv:1607.07384 [cs,Stat]. Available at http://arxiv.org/abs/1607.07384. Accessed on May 23 2017.
8. Kouloumpis, E., Wilson, T., and Moore, J. D. (2011). Twitter sentiment analysis: The good the bad and the OMG!. ICWSM, 11(538-541), 164.
9. Park M, McDoald DW, Cha M. Perception differences between the depressed and non-depressed users in Twitter. Seventh International AAAI Conference on Weblogs and Social Media; July 8-11, 2013; Massachusetts, USA. 2013.
10. G. B. Colombo, P. Burnap, A. Hodorog, and J. Scourfield. Analysing the connectivity and communication of suicidal users on twitter. Computer Communications, 73:291300, 2016.
11. Hu, M., and Liu, B. (2004, July). Mining opinion features in customer reviews. In AAAI (Vol. 4, No. 4, pp. 755-760).

12. Park S., Lee S. W., Kwak J., Cha M., Jeong B. (2013). Activities on Facebook reveal the depressive state of users. J. Med. Int. Res. 15:e217 10.2196/jmir.2718
13. De Choudhury M, Counts S, Horvitz EJ, Hoff A. Characterizing and predicting postpartum depression from shared Facebook data. 2014 Presented at: ACM Computer Supported Collaborative Work 2014; February 15-19, 2014; Baltimore, MD p. 626-638.
14. Losada, David E., and Fabio Crestani. "A Test Collection for Research on Depression and Language Use." CLEF. 2016.
15. National Institute of Mental Health. Depression Research, 1999.
16. Michalek, R., Tarantello, G.: Subharmonic solutions with prescribed minimal period for nonautonomous Hamiltonian systems. J. Diff. Eq. 72, 28–55 (1988)
17. Tarantello, G.: Subharmonic solutions for Hamiltonian systems via a $\mathbb{Z}_p$ pseudoindex theory. Annali di Matematica Pura (to appear)
18. Rabinowitz, P.: On subharmonic solutions of a Hamiltonian system. Comm. Pure Appl. Math. 33, 609–633 (1980)