PEIMEX at eRisk2018: Emphasizing personal information for depression and anorexia detection

Rosa M. Ortega-Mendoza[1], A. Pastor López-Monroy[2], Anilu Franco-Arcega[3], and Manuel Montes-y-Gómez[4]

[1] Instituto Tecnológico Superior del Oriente del Estado de Hidalgo, Hidalgo, Mexico
[2] Houston of University, Texas, USA
[3] Universidad Autónoma del Estado de Hidalgo, Hidalgo, Mexico
[4] Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Puebla, Mexico
mortega@itesa.edu.mx, alopezmonroy@uh.edu, afranco@uaeh.edu.mx, mmontesg@inaoe.mx

Abstract. The early detection of risks behaviors can significantly contribute to prevent serious psychological and health consequences. This paper reports our participation at eRisk 2018 in the tracks of early detection of anorexia and depression in social media. Our approach considers that the sentences where users refer to themselves contain terms that better expose their interests and habits and, therefore, are able to reveal characteristics of their personality, and social and psychological states. The main idea is to emphasize the value of these terms by using novel feature selection and term weighting techniques. The approach achieved a competitive performance, it obtained results higher than average values of the competition. These results evidence that terms related to personal information are important for risk detection.

Keywords: Early Text Classification · Anorexia Detection · Depression Detection · Personal Information.

1 Introduction

Recently, the prediction of risks on the Internet has emerged as a relevant and challenging research area. When the risks are related to health it is very important to predict them as soon as possible (early prediction), because the impact of these problems can be lethal for health and integrity of people. The early prediction provides opportunities for assistance that helps to mitigate or minimize these problems. The eRisk 2018 Challenge [8] evaluates solutions for tackling two crucial disorders that may cause serious problems in social relations of people or even to suicide, in specific: early risk detection of depression and anorexia. In this paper, we describe our approach submitted for participating on these tasks.

The proposed method is based on the relevance of personal phrases -sentences with a first person pronoun[5]- to characterize social media users. In previous works [12, 13], supported in psychological findings [14], we demonstrated that terms located in this

[5] Namely: *I, me, mine, myself, my,* as well as *im*, which is very common in social media.

type of phrases have a special value for discriminating among different profiles. We hypothesized that in these phrases users better expose their interests, preferences, habits, fears and routines, which help to easily characterize and discriminate the different profiles.

These previous findings have inspired the hypothesis of this research work, which states that personal phrases are not only relevant for author profiling but also to detect behaviors or mental states (disorders), for example, risks of anorexia and depression. We believe that people suffering these disorders make clear their psychological state when they talk about themselves in social media. Thus, in personal phrases users expose personal information as thematic interests that may be similar among people with the same disorder and different from healthy people, as a result, the discriminating among those people type can be better achieved. For example, the following phrases were obtained from users diagnosed with depression[6]: "*I am currently prescribed cymbalta and wellbutrin*" and "*I had a dream last night that included thousands of spiders so the dream I had last night is still kinda freaking me out, even though i don't have arachnophobia*"; that could suggest that terms about medicines and phobias could be signs of the presence of this disorder. Therefore, this work is aimed to detect users with these disorders by emphasizing the value of personal information by means of special methods for feature selection and term weighting schemes.

This paper is organized as follows. Section 2 exposes the related work. Section 3 describes the proposed method. Section 4 presents the experiments. Finally, conclusions and future work are drawn in section 5.

2 Related work

Although the early depression detection on the Internet is a emerging area, several automatic methods have been proposed to face this challenge. For example, in the eRisk 2017, 30 methods from 8 different institutions were presented [7]. The contributions enclose: machine learning approaches on various features [10] including linguistic meta information [19] or features based on a depression lexicon [17]; graph models [20]; Bag of Words models and semantic representation of documents as Paragraph Vector, Latent Semantic Analysis, and Recurrent Neural Networks using Long Short Term Memory [21]. Some other works have used combinations of supervised learning and information retrieval approaches [1]. Finally, one of the most outstanding approaches has considered the variation of the vocabulary along the different time steps as a concept space for document representation [3].

The anorexia detection is a problem mostly studied from the medical (psychiatry and psychology) perspective [5, 9, 22, 23], and recently explored from the automatic analysis of the writing using social media texts. For example, in [15] the authors applied a method based on regular expressions and machine-learning classifiers to identify a set of tweets that show the presence of certain diseases or health states as eating disorders. In [18], the authors compared written sentences by patients with anorexia versus healthy patients, finding a set of negative emotions to characterize the presence

[6] Examples of personal phrases from the positive class in the depression corpus of eRisk 2018.

of this alimentary disorder. Therefore, they found a relation between the mental state of the people and the characteristics of their language .

Personal pronouns have shown to be very useful features to characterize authors or their mental state. For example, a study among depressed, formerly-depressed, and never-depressed students found that depressed individuals used more frequently the pronoun *I* [14, 16]. Similarly, an analysis of Twitter messages has shown that users suffering from depression used the words *my* and *me* much more frequently than others [11]. Even though we studied the terms around of a first person pronoun rather than the pronouns as attributes, these works suggest that the use of self-references are strongly related to the expression of people's feelings, concerns and opinions.

3 The proposed method

Our work is inspired by the ideas proposed in [12, 13] about the relevance of personal information as the essence of documents in order to model profiles of users. As mentioned before, in this research we hypothesize that personal information has high relevance for modeling behaviors or mental disorders. Therefore, we use the DPP-EXPEI approach that emphasizes personal information in the building of the text representations. The first part of this section describes that approach. Afterwards, the adjustment for tackling early risk tasks is presented.

3.1 The DPP-EXPEI approach

The DPP-EXPEI approach was introduced in a previous work for author profiling task [13]. It is based on a supervised classification framework using a standard BoW representation. The approach is aimed to emphasize the value of terms located in personal phrases by means of two processes implicated in the construction of the text representations: a feature selection stage based on a novel technique called *discriminative personal purity (DPP)*, and a term weighting scheme named *exponential reward of personal information (EXPEI)*.

Feature selection using DPP. The goal of feature selection is to detect the subset of most relevant terms for the classification task. DPP is a feature selection technique that considers that the terms more relevant for profiling are those expressed inside personal phrases. Technically, DPP selects terms according to a score about the distribution of terms across the categories as well as the kind of phrases they appear in. The DPP scheme, described in the Formula 1, takes into account the level of occurrence of terms in personal phrases (their purity PP), in combination with an estimation of the distribution of terms across the categories by means of the Gini function. More details about this formula are provided in [13].

$$DPP(t_i) = \max_{k=1}^{|C|} \{PP_k(t_i)\} \cdot gini(t_i) \tag{1}$$

Term weighting using EXPEI. In a BoW, the term weighting schemes assign a weight, w_{ij}, for each term t_i of the vocabulary. The assigned weight represents how the term is contributing to the description of the document d_j. The EXPEI term weighting scheme proposes an exponential rewarding to the weight of terms that mainly occur in personal phrases. Technically, the EXPEI scheme considers a traditional term weighting such as the normalized frequency (TF) and then it rewards their occurrence in personal phrases according to the PEI value as shown in Formula 2 (more details are presented in [13]). The PEI measure handles the reward because estimates the quantity of personal information revealed by a term.

$$w_{ij} = \left(\sqrt{TF(t_i, d_j)}\right)^{1-PEI(t_i,d_j)} \tag{2}$$

3.2 Early risk detection based on DPP-EXPEI

Each post was represented as vector formed by a combination of content and style features. Particularly, it includes content words, punctuation marks, slang words and out of-dictionary terms like emoticons. The proposed method considers that personal phrases have terms highly discriminative for such disorders. Then, the representation is formed by vectors considering the 1,000 terms more discriminative according to the DPP technique. We also added the stopwords into the representation. Each attribute inside the representation was weighted with the EXPEI scheme. The text representation was then feed to a machine learning algorithm. Particularly, we used the linear Support Vector Machine (SVM) from the LIBLINEAR library considering default parameters [4]. For taking a decision about early detection we designed some criteria, which are described below.

Early decision, Wait or Classify? Our approach is prepared to detect risks by analyzing the personal phrases. The early detection axis is tackled by external criteria related to review the classifier decisions. We used two different criteria in order to decide whether to submit a decision for a subject or wait for more chunks. The criteria are:

- C1: to assign the positive decisions taken by the classifier in the current chunk.
- C2: to submit a positive decision only if the current chunk as well as one previous chunk were classified as positive.

The criterion C1 takes a decision analyzing the current tags assigned by the classifier, meanwhile C2 is more strict because it takes a decision analyzing previous classifications. In both C1 and C2 cases, negative decisions are taken at the end, more specific, until the submission the decisions from the chunk-10. Taking into account these criteria, we submitted results obtained from different combinations.

4 Experiments

4.1 Datasets

The datasets presented in the CLEF 2018 eRisk forum consist of writings (posts or comments) from a set of social media users. There are two categories of users for each task: depressed and non-depressed, and, with anorexia and non-anorexia respectively. For each user, his collection of writings has been divided into 10 chunks. The first chunk contains the oldest 10% of the messages, the second chunk contains the second oldest 10%, and so forth. In summary, the number of users in training and test datasets are shown in Table 1. More details about the datasets are described in [8].

Table 1: Number of users by category of the datasets

Subjects	Training	Test
Task 1: Depression		
depressed	135	79
non-depressed	752	741
Total	807	820
Task 2: Anorexia		
with anorexia	20	41
non-anorexia	132	279
Total	152	320

4.2 Runs configuration

We evaluated five different configurations which are combinations of: *i*) feature selection techniques, Information Gain (IG) and DPP; *ii*) term weighting schemes, the normalized frequency (TF) and EXPEI; and *iii*) the two criteria to decide about the early classification.

4.3 Evaluation

The evaluation took into account not only the correctness of the (binary) decision (i.e. whether or not the user has a risk behavior) by means of the F_1 measure, but also the delay taken by the approach to make the decision through an early risk detection error measure called $ERDE_o$, with cutoff parameter o set to 5 and 50 posts ($ERDE_5$ and $ERDE_{50}$). This measure, introduced in [6], aimed to reward early alerts (positive decisions) taking into account the number of distinct textual items seen before giving the answer. On the other hand, it associates a cost to the delay in the detection of true positives, which increases according the number of distinct textual items seen before taking a decision.

4.4 Results

According to the eRisk2018, the tasks were divided into a training stage and a test stage. At the beginning of the tasks, the training sets were provided with the whole writings (all chunks from all training users) and their respective categories. On the other hand, the test sets were provided gradually; each test chunk was released one week after from the previous test chunk. Therefore, the participants provided weekly the runs (predictions) on the test datasets. This section presents the results from these two stages.

Training stage Given that we had access to all chunks from the training users, we trained and evaluated our model using all information by means of a 10CFV strategy. In this case, the early criteria were not take into account because we were mainly interested building models that consider all the information from the users. The results are showed in Figure 1.

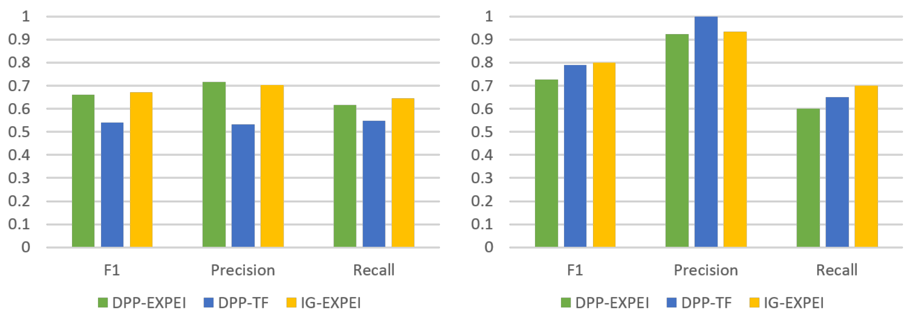
For depression detection task, the results indicate that DPP-EXPEI is better than DPP-TF, suggesting the relevance of giving more weight (or emphasizing) to the terms from personal phrases by the EXPEI term weighting scheme. Also, it is noticed that the combination IG-EXPEI has a similar performance than DPP-EXPEI, indicating that the terms from personal phrases, which were selected by DPP, have also good discriminative power.

For the anorexia detection task, the results show that DPP-TF is working better than DPP-EXPEI suggesting that DPP takes advantage of traditional term weighting schemes (e.g. TF) for this task. We also noted that IG-EXPEI has a better performance than the DPP combinations. This suggests that terms in personal phrases are revealing the risk in a lower level than terms with higher IG values (which tend to be very frequent).

In summary, in this stage, we observed DPP-EXPEI is working better for depression and DPP-TF for anorexia detection. Specifically, we observed that the feature selection technique emphasizing personal information (DPP) is better for depression detection task than for anorexia. Also, it was noticed that the term weighting scheme EXPEI is working well for both tasks. Moreover, we observed that for the anorexia task our approach is very precise, but it shows low recall. That is, there were very few mistakes in positive prediction, but only few positives cases were detected, which affected the global performance. On the other hand, the depression detection results to be a more complex task, due to the lower performance than anorexia detection, with small differences between precision and recall values.

Testing stage. The final submissions to the CLEF 2018 early risk detection tasks were scored using the $ERDE_5$, $ERDE_{50}$ as well as F_1 scores. The official results are showed in Table 2. In both depression and anorexia detection tasks, our best results considering the $ERDE$ (5 or 50) measure were obtained using configurations that included the DPP technique. This means that our approach is faster (it required fewer writings to make the alert) when DPP is used. On the other hand, results indicate the EXPEI technique is working for both tasks, suggesting that the occurrence of terms in personal phrases is an important discriminative factor.

Regarding the depression detection, the criterion C1 obtained the best results. On the other hand, the prediction of the positive class (alert) was better when the representation is formed by terms located in personal phrases than when the terms are selected with IG (please refer to the F_1 results).



(a) Depression (b) Anorexia

Fig. 1: Results in the training datasets for the two tasks. The results corresponds exclusively to the positive class.

Table 2: Official results for the eRisk 2018 challenge

Task	Name submission	Name method	Criterion	$ERDE_5$	$ERDE_{50}$	F_1	P	R
Depression	PEIMEXA	DPP-EXPEI	C1	10.30%	**7.22%**	0.38	0.28	**0.62**
	PEIMEXB	DPP-EXPEI	C2	10.30%	7.61%	**0.45**	**0.37**	0.57
	PEIMEXC	DPP-TF	C1	**10.07%**	7.35%	0.37	0.29	0.51
	PEIMEXD	IG-EXPEI	C1	10.11%	7.70%	0.39	0.35	0.44
	PEIMEXE	IG-PEXPEI	C2	10.77%	7.32%	0.35	0.25	0.57
Anorexia	PEIMEXA	DPP-EXPEI	C1	12.70%	9.25%	0.46	0.39	0.56
	PEIMEXB	DPP-EXPEI	C2	**12.41%**	**7.79%**	0.64	0.57	**0.73**
	PEIMEXC	DPP-TF	C1	13.42%	10.50%	0.43	0.37	0.51
	PEIMEXD	IG-EXPEI	C1	12.94%	9.86%	**0.67**	**0.61**	**0.73**
	PEIMEXE	IG-EXPEI	C2	12.84%	10.82%	0.31	0.28	0.34

In the anorexia detection task, the approach was favored when previous classifications were considered (i.e., when the C2 criterion was used). However, the more discriminative terms were obtained with IG (refer again to the F_1 values). These observations agree with the results from the train stage, suggesting that DPP is better for the depression detection task than for the anorexia detection. However, more deeply studies are needed to conclude about the relation of the content of personal phrases and the language use by people with anorexia.

Regarding the official results, the proposed method had a competitive performance in both tasks. The results are considerably and consistently better than the average methods' performances in terms of ERDE_5, ERDE_{50} and F_1 values. In specific, using each of the measures the approach was ranked among the top 20 and 12 positions for depression and anorexia tasks respectively. Specifically, the proposed method was better ranked using the ERDE_{50} than the ERDE_5; in the former case it was located in eighth and fourth positions of the competition at depression and anorexia detection tasks respectively. These performances indicate the approach tends to produce better decisions when more information is considered.

5 Conclusions and future work

In this paper we presented a novel approach for early depression and anorexia detection that was evaluated at the eRisk 2018 challenge. The approach is inspired by a previous work about the relevance of personal phrases (sentences with a first person pronoun) to expose interests, preferences, habits and routines our social media users. In this research we hypothesize that terms located in personal phrases help to highlight information that reveals behavior or mental state of people. The idea is exploiting terms occurring in personal phrases by means of term selection and weighting methods: DPP and EXPEI respectively.

Preliminary results in the eRisk 2018 datasets showed the feasibility of the approach. We achieved results higher than the average values of performance at the competition. Our results indicate that there is relevant information in social media texts to detect health risks of users. The DPP scheme is worked better for depression than for anorexia detection. The EXPEI schemes enriched both tasks. We think that the relevance of personal information in the tasks can be exploited in different ways, for example, in combination with word embeddings.

We used a machine learning algorithm with default parameters, worthwhile further investigations are needed to refine the parameters of classifiers or to probe new classifiers to improve the performance. In addition, we are interested in analyzing the use of phrases containing other kind pronouns, such as third person pronouns, since we have the intuition that many depression problems have their roots in interpersonal relations.

References

1. Almeida, H., Briand, A., Meurs, M.: Detecting Early Risk of Depression from Social Media User-generated Content. In: Cappellato et al. [2]

2. Cappellato, L., Ferro, N., Goeuriot, L., Mandl, T. (eds.): Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum, Dublin, Ireland, September 11-14, 2017, CEUR Workshop Proceedings, vol. 1866. CEUR-WS.org (2017)
3. Errecalde, M.L., Villegas, M.P., Funez, D.G., Ucelay, M.J.G., Cagnina, L.C.: Temporal Variation of Terms as Concept Space for Early Risk Prediction. In: Cappellato et al. [2]
4. Fan, R.E., Chang, K.W., Hsieh, C.J., Wang, X.R., Lin, C.J.: LIBLINEAR: A Library for Large Linear Classification. Journal of Machine Learning Research **9**, 1871–1874 (2008)
5. Higgins, A., Cahn, S.: Detection of anorexia nervosa in primary care. Eating Disorders **26**(3), 213–228 (2018)
6. Losada, D.E., Crestani, F.: A Test Collection for Research on Depression and Language Use. In: Fuhr, N., Quaresma, P., Gonçalves, T., Larsen, B., Balog, K., Macdonald, C., Cappellato, L., Ferro, N. (eds.) Experimental IR Meets Multilinguality, Multimodality, and Interaction. pp. 28–39. Springer International Publishing, Cham (2016)
7. Losada, D.E., Crestani, F., Parapar, J.: eRISK 2017: CLEF Lab on Early Risk Prediction on the Internet: Experimental Foundations. In: Jones, G.J.F., Lawless, S., Gonzalo, J., Kelly, L., Goeuriot, L., Mandl, T., Cappellato, L., Ferro, N. (eds.) Experimental IR Meets Multilinguality, Multimodality, and Interaction - 8th International Conference of the CLEF Association, CLEF 2017, Dublin, Ireland, September 11-14, 2017, Proceedings. Lecture Notes in Computer Science, vol. 10456, pp. 346–360. Springer (2017)
8. Losada, D.E., Crestani, F., Parapar, J.: Overview of eRisk – Early Risk Prediction on the Internet. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Ninth International Conference of the CLEF Association (CLEF 2018). Avignon, France (2018)
9. Lyons, E.J., Mehl, M.R., Pennebaker, J.W.: Pro-anorexics and recovering anorexics differ in their linguistic Internet self-presentation. Journal of Psychosomatic Research **60**(3), 253 – 256 (2006)
10. Malam, I.A., Arziki, M., Bellazrak, M.N., Benamara, F., Kaidi, A.E., Es-Saghir, B., He, Z., Housni, M., Moriceau, V., Mothe, J., Ramiandrisoa, F.: IRIT at e-Risk. In: Cappellato et al. [2]
11. Nadeem, M.: Identifying Depression on Twitter. CoRR **abs/1607.07384** (2016)
12. Ortega-Mendoza, R.M., Franco-Arcega, A., López-Monroy, A.P., Montes-y-Gómez, M.: I, Me, Mine: The Role of Personal Phrases in Author Profiling. In: Fuhr, N., Quaresma, P., Gonçalves, T., Larsen, B., Balog, K., Macdonald, C., Cappellato, L., Ferro, N. (eds.) Experimental IR Meets Multilinguality, Multimodality, and Interaction: 7th International Conference of the CLEF Association, CLEF 2016, Évora, Portugal, September 5-8, 2016, Proceedings, pp. 110–122. Springer International Publishing, Cham (2016)
13. Ortega-Mendoza, R.M., López-Monroy, A.P., Franco-Arcega, A., Montes-y-Gómez, M.: Emphasizing personal information for Author Profiling: New approaches for term selection and weighting. Knowledge-Based Systems **145**, 169 – 181 (2018)
14. Pennebaker, J.: The Secret Life of Pronouns: What Our Words Say About Us. Bloomsbury USA (2011)
15. Prieto, V.M., Matos, S., Álvarez, M., Cacheda, F., Oliveira, J.L.: Twitter: A Good Place to Detect Health Conditions. PLOS ONE **9**(1), 1–11 (01 2014)
16. Rude, S., Gortner, E.M., Pennebaker, J.: Language Use of Depressed and Depression-vulnerable College Students. Cognition & Emotion **18**(8), 1121–1133 (2004)
17. Sadeque, F., Xu, D., Bethard, S.: UArizona at the CLEF eRisk 2017 Pilot Task: Linear and Recurrent Models for Early Depression Detection. In: Cappellato et al. [2]
18. Spinczyk, D., Nabrdalik, K., Rojewska, K.: Computer aided sentiment analysis of anorexia nervosa patients' vocabulary. BioMedical Engineering OnLine **17**(1), 19 (2018)
19. Trotzek, M., Koitka, S., Friedrich, C.M.: Linguistic Metadata Augmented Classifiers at the CLEF 2017 Task for Early Detection of Depression. In: Cappellato et al. [2]

20. Villatoro-Tello, E., Ramírez-de-la-Rosa, G., Jiménez-Salazar, H.: UAM's Participation at CLEF eRisk 2017 task: Towards Modelling Depressed Blogers. In: Cappellato et al. [2]
21. Villegas, M.P., Funez, D.G., Garciarena, M.J., Cagnina, L.C., Errecalde, M.L.: LIDIC - UNSL's Participation at eRisk 2017: Pilot Task on Early Detection of Depression. In: Cappellato et al. [2]
22. Walsh, J.M.E., Wheat, M.E., Freund, K.: Detection, evaluation, and treatment of eating disorders. Journal of General Internal Medicine **15**(8), 577–590 (Aug 2000)
23. Weeda-Mannak, W.L.: Aspects of Early Detection of Anorexia Nervosa. In: Krakowski, A.J., Kimball, C.P. (eds.) Psychosomatic Medicine: Theoretical, Clinical, and Transcultural Aspects, pp. 119–124. Springer US, Boston, MA (1983)