Artificial Intelligence In Medicine 123 (2022) 102202

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

Revealing traces of depression through personal statements analysis in social media

Rosa María Ortega-Mendoza [a,b], Delia Irazú Hernández-Farías [a,c,*], Manuel Montes-y-Gómez [a], Luis Villaseñor-Pineda [a]

[a] Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Luis Enrique Erro No. 1, Sta. Ma. Tonantzintla, C.P. 72840 Puebla, Mexico
[b] Universidad Politécnica de Tulancingo (UPT), Ingenierías No. 100 Col. Huapalcalco, C.P. 43629 Tulancingo, Hidalgo, Mexico
[c] División de Ciencias e Ingenierías Campus León, Universidad de Guanajuato Lomas del Bosque No. 103, Lomas del Campestre, C.P. 37150 León, Guanajuato, Mexico

ARTICLE INFO

Keywords:
Depression detection
Personal information
Personal pronouns
DPP-EXPEI

ABSTRACT

Depression is a common and very important health issue with serious effects in the daily life of people. Recently, several researchers have explored the analysis of user-generated data in social media to detect and diagnose signs of this mental disorder in individuals. In this regard, we tackled the depression detection task in social media considering the idea that terms located in phrases exposing personal statements (i.e., phrases characterized by the use of singular first person pronouns) have a special value for revealing signs of depression. First, we assessed the value of the personal statements for depression detection in social media. Second, we adapted an automatic approach that emphasizes the personal statements by means of a feature selection method and a term weighting scheme. Finally, we addressed the task in hand as an early detection problem, where the aim is to detect traces of depression with as much anticipation as possible. For evaluating these ideas, benchmark Reddit data for depression detection was used. The obtained results indicate that the personal statements have high relevance for revealing traces of depression. Furthermore, the results on early scenarios demonstrated that the proposed approach achieves high competitiveness compared with state-of-the-art methods, while maintaining its simplicity and interpretability.

1. Introduction

Depression is one of the most common mental health illnesses that affects seriously the daily life of people. It can produce a variety of emotional and physical problems that can diminish the activities of an individual provoking negative effects on her/his surrounding personal context, work [1] or school [2], and even basic human needs such as sleeping [3] and eating [4]. Severe cases can lead to self-harm and suicide [5]. Although fortunately, depression is a treatable disorder, it is often undetected due to several reasons such as the patient's own inability to recognise the problem or because of the social stigma associated with mental illnesses. Recently, the rise in the use of social media has opened new opportunities for detecting depression [6–8]. In these platforms, people freely share and express their thoughts and feelings. Furthermore, often these media are used by depressed users to gather information about their illness or to discuss about their problems and symptoms.

Based on the idea that language is a powerful indicator of personality, social or emotional status, as well as mental health [9], several research works have leveraged the content generated by social media users as a rich source of knowledge to study, infer, and track users with mental illnesses such as depression. Particularly, it has been demonstrated that individuals having depression manifest changes in their language and behaviour (e.g., greater negative emotion and high self-attentional focus) [10]. In this regard, the development of methods for automatic depression detection on social media has gained a special interest in the computational linguistics research community. Such a challenging task has been generally tackled as a text classification problem, considering a wide variety of text representations and classification models, and concluding that the thematic and stylistic aspects are different between depressed and mental healthy people [11]. However, little has been studied about the relationship of information

* Corresponding author at: Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Luis Enrique Erro No. 1, Sta. Ma. Tonantzintla, C.P. 72840 Puebla, Mexico.

E-mail addresses: rosa.ortega@upt.edu.mx (R.M. Ortega-Mendoza), di.hernandez@ugto.mx (D.I. Hernández-Farías), mmontesg@inaoep.mx (M. Montes-y-Gómez), villasen@inaoep.mx (L. Villaseñor-Pineda).

https://doi.org/10.1016/j.artmed.2021.102202
Received 22 April 2020; Received in revised form 16 September 2021; Accepted 29 October 2021
Available online 10 November 2021

exposed by the use of self-references with the expression of depression.

From the psychological perspective, the relative frequency of singular first-person pronouns in naturally spoken language helps to predict depressive symptoms. For example, some studies indicate that self-references are more frequent in depressed people than in those never-depressed [12–14]. Some theories also suggest that high self-focused attention can reveal signs of the depression disorder [15,16]. Besides, in the literature, some computational methods have observed a high frequency of first-person pronouns in texts generated by depressed people [17–20]. However, to the best of our knowledge, no other author has studied the value of the pronouns' contexts for the depression detection. In this regard, we are interested in investigating whether or not the personal statements, i.e., information present in phrases with a singular first-person pronoun (we also refer to it as *personal information*), concentrates keywords revealing the users' mental status.

Supported on these psychological findings and on some previous works showing that personal information helps to determine users traits such as their gender and age [21,22], we hypothesize that when users share personal information they tend to use words that could become potential clues to reveal mental health disorders such as depression. Particularly, we believe that in personal statements similar thematic interests are shared by people with the same disorder, being different from healthy people. For example, the next phrase, a personal statement, written by a user diagnosed with depression concentrates terms highly related to this domain: *"I am* ***prescribed medication*** *to treat my* ***depression****. I cannot calm my* ***anxiety****"*. On the other hand, we believe that emphasizing the value of these words by feature selection and term weighting techniques could help to distinguish adequately depressed from non-depressed people. Furthermore, with these ideas in mind, we also faced the *early depression detection* problem, where the goal is to identify, as soon as possible, users who show signs of depression. The early detection increases the chances for appropriate treatment as quickly as possible. In this regard, we consider that, unlike techniques based on the frequency of terms, the proposed approach allows us to pay special attention to frequent terms in the personal contexts of the authors, favoring the early detection of this disorder.

In summary, the contributions of this paper are as follows. First, the assessment of the relevance of the personal statements for depression detection. Second, the adaption of a feature selection method and a term weighting scheme that automatically extract and emphasize this kind of information for the task of depression detection. It is worth noting that, unlike most previous methods dealing with this task, the proposed approach allows pointing out the most salient terms found in texts written by depressed users as well as interpreting the obtained results in terms of classification. These interpretability qualities are crucial considering that one of the purposes of developing such kind of approaches is to bring the specialists with tools for supporting their diagnosis. Third, the definition of a setting for tackling early depression detection. Supported on a set of words automatically extracted from the personal statements of social media users, we propose a heuristic to identify those suffering from depression as early as possible.

The rest of this paper is organized as follows. Section 2 provides a review of related work on depression detection on social media and early risk detection. Section 3 describes a study about the relevance of personal statements in depression detection. Section 4 presents the proposed approach for depression detection. Section 5 reports the experiments and results. Finally, Section 6 points out our conclusions and future work.

2. Related work

Data gathered from social media have been used for analyzing topics related to public health. Several health conditions (such as allergies, migraine, obesity, among others) have been analyzed by considering Twitter data [23]. Computational linguistics methods have been also applied to study mental illnesses such as *attention deficit hyperactivity disorder*, *generalized anxiety disorder* [24], *schizophrenia* [25], *post-traumatic stress disorder* [11], *postpartum depression* [26], *risks of suicide* [27], as well as *depression*.

Research on topics related to depression has been carried out by different areas such as psychology, medicine, linguistics, etc. All of them have tried to identify causes, symptoms and, of course, how to correctly diagnose a patient having depression. In such a way, user-generated content in social media has become an interesting subject of study of depression-related topics [28,29].

It has been proven that social media data capture aspects related to the mental and emotional condition of people [30]. Therefore, data coming from different social media platforms such as *Twitter* [18,31–34], *Facebook* [35,36], *Reddit* [6,37], and *LiveJournal* [38] have been used for investigating depression in user-generated content.

According to [39], Twitter posts can be considered as a trustworthy source for analyzing depressive content. They identified that users tend to share information about their depression in a broad sense, including depressive feelings, treatment, and so on. In [33], the authors compare depressed against non-depressed Twitter users in terms of several aspects. Depressed users show lower social activity, tend to use more negative expressions, and reflect more medical concerns and religious thoughts. Nadeem et al. [30] retrieved all the messages posted on Twitter during a year by users who manifested having been diagnosed with depression. They experimented with bag-of-words together with standard classifiers. A set of Twitter data retrieved from a mental health campaign in Canada[1] was used by [40]. The authors experimented with n-grams, lexicon-based features as well as with the frequency of personal pronouns.

Deep learning based methods have also been applied for depression detection. In [18], the authors experimented with Convolutional Neural Networks and Recurrent Neural Networks obtaining competitive results on the detection of Twitter users with depression. Written texts reflecting potential clues of depression are more likely to be very subjective. Therefore, some authors have decided to take advantage of affective content for identifying depressed users. In [34], the authors exploited sentiment analysis in order to determine the polarity of Twitter posts. They found that longer emotional tweets are posted by users with depression. The role of emotions in depression-content posts has also been used for identifying depression on Twitter [31] and on Reddit posts [6]. More recently, a one-class classification approach for dealing with depression detection on Reddit was proposed by [41]. Some other researches have considered the contacts' network structure in order to analyze the presence of depression [42,43]. In [33], the authors developed a Twitter dataset composed by posts shared by users diagnosed with clinical depression. They experimented with a probabilistic model considering a wide range of features such as emotional expression, linguistic style, and social network properties.

2.1. Shared tasks on depression detection

Interest on investigating mental health issues from the perspective of Natural Language Processing has led into the organization of shared tasks dedicated to promote the research on these topics. In the framework of the 2015 Computational Linguistics and Clinical Psychology (CLPsych),[2] it was organized the CLPsych-2015 shared task [32]. Its aim was to identify users having depression from Twitter data. Participating systems addressed this task by applying topic models, bag-of-words, rule-based approaches, character language models, among others.

Being able to detect depression involves an additional challenge, it is needed to identify any clue of this mental illness as soon as possible in order to avoid non-desirable further consequences. Therefore, it is crucial to consider such a challenging task as an *early risk detection*

[1] The hashtag #BellLetsTalk was used for collecting such data.
[2] http://clpsych.org/

(henceforth ERD) problem. Research on ERD related issues attempts to develop systems taking care of the time as a fundamental factor for solving such tasks [37]. The first dataset comprising a set of chronological social media posts regarding depressive and non-depressive users was developed by Losada et al. [37]. Those data serve not only for researching differences on the use of the language among depressed and non-depressed users, but also for studying the evolution of language of depressed users.

In 2017, the first shared task on early risk detection of depression (eRisk) was organized [7]. The aim of the task was to identify early traces of *depression* by considering a given sequence of writings in chronological order. In this sense, the *Early Risk Detection Error* (ERDE) measure was proposed[3] for considering simultaneously the accuracy of the classifiers and the delay in making a prediction. Participating systems at the shared task addressed the challenge by taking advantage of bag-of-words, n-grams, lexicon-based features, Latent and Concise Semantic Analysis, as well as, standard classifiers and neural networks, among others. The second edition of the eRisk task was organized in 2018 [8]. Approaches exploiting different machine learning algorithms, as well as Convolutional Neural Networks, bag-of-words representations, word embeddings, and domain-specific vocabularies were proposed for dealing with depression detection. In the last two editions [44,45], a new task aimed at calculating the level of severity of depression of a given user was organized.

3. Relevance of the personal statements for depression detection

From both psychological and computational linguistics perspectives, it has been established that the words we use reveal who we are, our thoughts, feelings, ideology, behaviors, and even our personality [12,46]. In this regard, the personal pronouns have been characterized as discriminative markers of personal traits [47]. Based on these findings, previous works have shown that phrases with first-person pronouns contain terms particularly useful to predict user traits such as gender and age [21,22]. Following these ideas, in this paper we hypothesize that in personal statements depressed users expose terms related to their symptoms, medications, fears, etc. Hence, terms from personal statements could be very helpful to detect signs of depression. To illustrate this assumption some examples of personal statements are shown in Table 1. In these phrases, the users pointed out personal experiences sharing clues that suggest a depressive status such as medications related to treatment of depression (e.g., *cymbalta*) or feelings (e.g., *anxiety, awful, anger, hopelessness,* etc.).

Naturally, users categorized as non-depressed also employ personal statements in the information they share, but their content is most of the time not related to depression. Besides, people can also use personal statements for sharing other experiences, where depressed behaviour is not evident. Beyond these difficulties, our hypothesis indicates that clues of depression are highly concentrated in personal statements. Attempting to validate it, we carried out the following first exploratory experiment to assess the value of personal statements.

Table 1
Examples of personal statements corresponding to users diagnosed with depression from collections of the eRisk shared tasks[a].

- *I practice relaxation techniques which slightly calm my fears but my depression is so hard to overcome...*
- *I was prescribed with duloxetine and bupropion...*
- *Very often, I feel a mixture of happiness, depression and fury. I really need to someone to share my feelings...*
- *I feel so bad every time i do something wrong...*
- *Last week I was crying a lot. Sometimes, I feel good but my eyes are full of tears. I just want to stay all day in my bed...*

[a] We slightly modified the original phrases to protect the privacy of the users.

3.1. Assessing the value of personal statements for depression detection

Traditionally, the task of detecting social media users suffering from depression has been faced as a binary text classification problem; its goal is to assign predefined classes (*depressed* and *non-depressed*) to a given set of users, each one represented by a single document that comprises all his/her writings. Considering this framework, we were interested in assessing the relevance of personal statements for this task, comparing the detection performance when using only this type phrases for training the classifiers against the performance achieved when considering the rest of the documents' information. For doing so, we first divided the phrases from each document into five subsets associated to different personal pronouns families: *I*, *you*, *he/she*, *we*, and *they*.[4] Then, we performed the classification of depressed and non-depressed users considering each of these groups to train some traditional classifiers.

3.1.1. Settings

3.1.1.1. Datasets. For experimental purposes, we took advantage of the corpora used in the eRisk shared tasks of the 2017 and 2018 editions [7,8]. Both collections contain writings from English-speaking users of reddit.com.[5] Particularly, the training set from the 2018 edition corresponds to the union of the training and testing partitions of the 2017 edition. Some statistics of these datasets are described in Table 2. In both collections, there are two categories of users: *depressed* and *non-depressed*. Users considered as depressed are those that explicitly expressed that they were *diagnosed* with depression. Also, Table 2 shows one of the main challenges of these corpora, a high imbalance towards the non-depressed class.

3.1.1.2. Filtering process. We split each document into phrases, and then divided them into groups according to the aforementioned pronouns families. For example, the set of personal statements is formed by the phrases containing pronouns from the *I*-family: *I*, *me*, *mine*, *my*, *myself*, *Im*.[6] Similarly we created the sets of phrases corresponding to the You-, He/She-, We- and They-families. Therefore, for each user, we kept only those phrases containing the selected personal pronoun family, reducing the amount of data per user employed to train the classifiers. Henceforth, we refer to these sets of phrases as *filtered subcorpora*.

3.1.1.3. Classification process. We perform a binary classification task (*depressed* vs. *non-depressed*; classes commonly also known as positive and negative, respectively) by considering for each user only the information filtered out according to a given pronoun family at time. In this

Table 2
Distribution of the train and test partitions of eRisk 2017 and 2018 datasets. Statistics represent the number of documents (i.e., users) in train and test partitions, respectively.

Class	2017		2018	
	Train	Test	Train	Test
depressed	83	52	135	79
non-depressed	403	349	752	741
Total	486	401	887	820

[3] Further details on ERDE are described on Section 5.2.1

[4] It is important to mention that the pronoun *it* was not considered because it refers to objects.

[5] https://www.reddit.com/

[6] This string is frequently used in documents from social media

experiment, a Bag of Words (BoW) with the traditional TF term weighting (normalized frequency) was used. Lexical unigrams representing content and style features were considered including: content words, slang words, punctuation marks and, out-of-dictionary terms such as emoticons. These features were extracted from the top 10,000 frequent terms in each collection. For the classification phase, a Support Vector Machine (SVM) with L2 normalization was used by means of the LinearSVC implementation of Python Sklearn with a grid search strategy (ranging from 0.125 to 16) for the parameter C optimizing the accuracy. We decided using a SVM classifier motivated by its results in the eRisk shared tasks, where it has achieved the best outcomes among the traditional classifiers, and comparable results against approaches based on deep learning techniques. Indeed, several participants have argued that the small size of the available datasets may be the reason for the SVM's competitiveness with respect to deep learning techniques.

Similar to most of the work in the state-of-the-art, we used F1 of the depressed class as the main evaluation measure, which remarks the importance in the positive class.

3.1.2. Results of the analysis

The results obtained with the application of the described methodology in both collections are shown in Fig. 1. Having in mind our goal of assessing the usefulness of the personal statements to depression detection, we compare the obtained results with each filtered subcorpora against the use of the whole information (represented as "All" in the figure). In each subfigure, the bar columns represent the F1 values of the positive class (Y-axis in the left) and the black dots show the percentage of phrases in the corresponding filtered subcorpora (Y-axis in the right). The variation of results according to the used pronoun-family suggests that the information expressed or shared by people depends on the interaction with the others (i.e., depends on who we are talking about). Interestingly, the performance achieved by the subset of personal statements (phrases corresponding to the I-family (pink bar) is very similar to the one obtained by the use of all information in the texts (blue bar). It is worth noting that the results obtained with the filtered data, in all cases, correspond to less than 20% of the information in the original corpus. Although moderate results were obtained using the subcorpora related to the second-person pronoun (*you*), the value of the phrases with the singular first-person pronoun (*I*) is very remarkable. In contrast, results in the negative class show that the non-depressed people use the personal pronouns in a similar way without peculiarities in their use for distinguishing this group of people. Therefore, these results support our hypothesis that personal statements contain special information revealing potential clues of depression. Additionally, these results confirm the findings of our previous works about the relevance of personal statements to characterize the author of a given text.

It is natural to wonder whether or not the phrases without personal pronouns are important for the task at hand. In this regard, in [21] was concluded that these kinds of phrases can also contain important information, although in a less grade than personal statements. Hence, a subsequent idea is to use all information, personal and non-personal but emphasizing the personal one [22] by suitable schemes for this purpose (refer to Section 4).

3.1.3. Why are the personal statements helpful for depression detection?

The previous experiment showed that personal statements capture different concepts depending on the family of pronouns used in the filtering process. Therefore, it is to be expected that the most relevant information for depression detection varies according to the pronoun used. In this section, we deepen the analysis showing the more relevant terms from each filtered subcorpus. Particularly, information gain was calculated on each subcorpus. Table 3 shows the top 40 words with the highest information gain for detecting depression in each subcorpus in the eRisk 2018 collection.[7]

There is a very noticeable difference in the type of terms in each row. The lists of words related to the *I* and *you* pronoun families belong to a vocabulary strongly associated with the depression domain, whereas the words related to the presence of the rest of the pronouns mostly express information of salient news topics at the time of the dataset creation, for example in the political domain:*Obama*, *government*, *country*, *Russia*, among others. These results suggest that phrases with *plural first-person pronouns* are not relevant for the task. On the other hand, this analysis confirms that phrases with *singular first-person pronouns* are relevant for depression detection. Interestingly, phrases with the *second person pronouns* also expose differences between depressed and non-depressed people. This means that personal statements with the pronouns *I* and *you* are indeed relevant for detecting depression.

3.2. Second-person narratives: a voice for expressing depression

It was to be expected that self-references involve clues for depression detection; however, in the previous section the phrases with *second person pronouns* stood out for their contribution to the task at hand. We deepen on this finding to understand the role of this type of sentences for detecting depression. Table 4 shows some examples of phrases filtered by using the *second person pronoun*. We observed that these phrases correspond to what is known as *second-person narration* [48], a common way to connect with the reader by a narrative in the second person. Second-person narration is commonly used to do narratives about depression as portraying the disease [49]. Narrators frequently use it in order to help the reader to understand the feelings associated with depression. This suggests that words in phrases belonging to the *you* pronoun family can also concentrate a special value for depression detection. Therefore, special attention should be placed to phrases with *I* and *you* family pronouns, hereinafter referred to as *personal statements*. Particularly, we take advantage of this type of phrases by emphasizing its value by means of an adaptation of the DPP-EXPEI approach [22].

4. Adapting DPP-EXPEI to the depression detection task

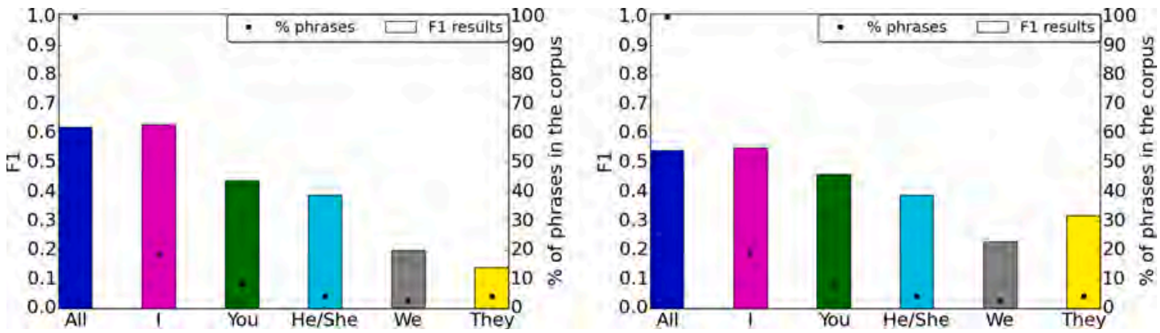
The DPP-EXPEI approach is aimed to emphasize the value of terms located in personal statements. It was introduced in a previous work for the author profiling task achieving remarkable results [22]. Technically, it is based on a supervised classification framework using a standard BoW representation. In line with the ideas supporting this approach and considering its simplicity as well as its good performance, we propose an adaption, referenced as DPP$_{fd}$-EXPEI, to face the depression detection task in social media.

The approach is based on a measure called *Personal Expression Intensity* (hereafter denoted as PEI), which is aimed to quantify how much each term is revealing personal information about the profile of an author or social media user. PEI assumes that the more frequent is a term in the personal statements of a document, and the less frequent in its non-personal statements, the more revealing is the term about the profile of the document's author. Formally, the function PEI for a term t_i in a document d_j is defined in the Eq. (1) estimating in a single value the combination (balance) of the two measures (*personal precision* and *personal coverage*, ρ and τ, respectively) that analyze the occurrences of the terms in personal statements:

$$PEI\left(t_i, d_j\right) = 2\frac{\rho\left(t_i, d_j\right)\cdot\tau\left(t_i, d_j\right)}{\rho\left(t_i, d_j\right) + \tau\left(t_i, d_j\right)} \tag{1}$$

where $\rho(t_i, d_j)$ estimates the concentration of personal information

[7] We showed results only in eRisk 2018 collection because it has the largest size. Moreover, this includes data of the 2017 collection.



(a) eRisk 2017

(b) eRisk 2018

Fig. 1. Results of the positive class using different filtered subcorpora. Bar columns represent the F1 values (primary axis) of the positive class. The black dots show the percentage of personal statements (secondary axis) in the subcorpora. Similar results are obtained when all the information is used than when the filtered subcorpora related to I pronoun were employed, however, the latter have much less information.

Table 3
The top 40 more discriminative words for each filtered subcorpus from the eRisk 2018 collection. The terms were selected by means of information gain for the depression detection problem.

Pronoun family	Words
I	*depression, feel, depressed, medication, diagnosed, suicidal, anxiety, therapist, life, feeling, meds, therapy, really, depressive, boyfriend, person, feelings, alone, family, relationship, suicide, cope, friends, disorder, trying, talk, anxious, mental, helped, want, things, parents, nervous, exhausted, anymore, help, bed, hard, relationships, emotionally.*
You	*depression, talk, feel, thank, help, medication, feeling, happy, things, makeup, relationship, country, market, alone, meds, experiences, attracted, suicidal, car, disorder, watch, positive, really, depressed, friends, com, attractive, content, family, feelings, google, political, coffee, sex, advice, therapist, policy, boyfriend, girlfriend.*
He/She	*michael, wants, campaign, george, created, market, law, obama, shows, state, prison, bill, public, clinton, interview, american, written, election, wrote, republican, former, adorable, specifically, federal, sanders, party, president, innocent, political, york, relationship, career, behind, private, interesting, til, candidate, gun.*
We	*things, sex, news, field, tax, relationship, quickly, government, depression, cut, security, reality, state, president, new, obama, politics, europe, together, communicate, system, land, military, using, used, mental, federal, states, disagree, country, russia, power, information, free, org, says, talked, police, citizens, program.*
They	*depression, public, people, states, government, companies, scientists, com, bill, based, american, former, country, costs, money, police, https, report, power, america, system, failed, greater, calls, new, security, court, admit, gorgeous, types, big, state, private, opportunity, president, across, claim, vote, lost, economic.*

Table 4
Examples of personal statements using the second-person narration from the collections of eRisk shared tasks[a].

- *think that you are not alone, there are many others in the same situation*
- *you cannot get out of bed, breathe, stop crying...*
- *it does not matter your background and behaviour, there is always someone who will hate you.*
- *you are so disillusioned with your life and with yourself that you have created someone who does not exist*
- *when you have feelings of hurting yourself, it's best to talk to someone you trust*

[a] We slightly modified the original phrases to protect the privacy of the users.

revealed by the term t_i, that is, the percentage of personal statements in the subset of phrases containing the term t_i, and $\tau(t_i, d_j)$ indicates the portion of the personal statements from document d_j covered by term t_i; in other words, the probability of occurrence of the term t_i in the subset of personal statements from d_j. Therefore, the values of PEI are ranged in an [0,1] interval.

The proposed approach takes advantage of PEI by means of two processes implicated in the building of the text representation: a feature selection method and a term weighting scheme. The adaption regards to the original DPP-EXPEI is mainly focused in the first process.

4.1. Feature selection using DPP_{fd}

The goal of feature selection is to detect the subset of most relevant terms for a given classification task. We believe that suitable feature selection methods should be applied according to the characteristics of the task at hand. Specifically, the depression detection task has been broadly tackled as a binary classification problem with an emphasis on the depressed category as the class of interest. Moreover, the task generally presents a high imbalanced degree towards the non-depressed class. In this regard, we designed the DPP_{fd} measure (*Discriminative Personal Purity for Depression)* as a feature selection technique characterized for giving preference to the terms commonly used by depressed users in their personal statements. Formally, the DPP_{fd} function for a term t_i is described in Eq. (2):

$$DPP_{fd}(t_i) = \max_{k=1}^{|C|}\{PP_k(t_i)\}\cdot wdif(t_i) \tag{2}$$

It consists of two components: a descriptive factor called *personal purity (PP)* that captures how a term describes the personal information from the documents of a category c_k, and a discriminative component referenced as *weighted difference (wdif)* that estimates how a term is distributed among the documents of the categories (in this case depressed and non-depressed). In the following paragraphs, we described these two components.

4.1.1. Personal purity

Based on the idea that writings with personal information better describe their authors, this factor estimates the capability of a term to describe personal information of authors (users) belonging to a category c_k as follows:

$$PP_k(t_i) = log_2\left(2 + \frac{1}{2}\sum_{d_j \in c_k}\frac{PEI(t_i, d_j) + 1}{NEI(t_i, d_j) + 1}\right) \tag{3}$$

where the function NEI (*Non-personal Expression Intensity*) is an analogous concept to PEI, but it transfers the focus to the occurrences of a term t_i in the set of phrases without first-person pronouns in a document d_j.

4.1.2. Weighted difference

This factor is a modification of the original DPP-EXPEI approach where the Gini factor was used instead. As it was mentioned before, we propose to emphasize the value of terms belonging to the depressed category. For achieving it, we designed a biased difference of the distribution of terms in the two categories, depressed and non-depressed. This difference is showed in Eq. (4):

$$wdif(t_i) = F_d(t_i)^{1/m} - F_n(t_i), \tag{4}$$

where $F_d(t_i)$ and $F_n(t_i)$ represent the relative frequency (ranging from 0 to 1 values) of the number of documents containing the term t_i in the depressed and non-depressed categories, respectively. In the equation, the most important component is the $1/m$ exponent.[8] It allows assigning more importance to terms belonging to instances of the depressed category, and, therefore, to face the class imbalance problem.

In essence, the adaption DPP_{fd} allows selecting terms strongly related to personal information from depressed users. Hence, the terms selected by DPP_{fd} could be considered as a kind of *lexicon* related to depression.

4.2. Term weighting using EXPEI

When a BoW approach is used, it is necessary to use a term weighting schema to associate a weight w_{ij} to each term t_i at each document d_j. EXPEI (*EXponential rewarding of PErsonal Information*) is a term weighting schema that proposes to give an exponential reward to the weight of those terms that tend to occur more frequently in personal statements. EXPEI is based on the normalized frequency, but it rewards the occurrence of the terms in personal statements by means of their *PEI* value and compensates low frequencies by means of the exponent $1/m$. Eq. (5) shows the EXPEI term weight. Further information can be found in [22].

$$w_{ij} = \left(F\left(t_i, d_j\right)^{1/m}\right)^{1-PEI\left(t_i,d_j\right)} \tag{5}$$

5. Results and discussion

This section presents the results of DPP$_{fd}$-EXPEI on the depression detection task. It is organized in three main subsections. The experiment in Section 5.1 evaluates the relevance of the approach to this task when it is used to emphasize the phrases containing *I* and *you* personal pronouns. The second experiment, at Section 5.2, aims to assess the applicability of DPP$_{fd}$-EXPEI in an early detection scenario. Finally, Section 5.3 presents an error analysis of the obtained results.

In the following experiments, we used the same experimental configuration as described in Section 3.1.1. However, in this case, we selected the most relevant terms by means of DPP_{fd}, then, using these terms, we built a standard BoW representation considering the *EXPEI* term weighting scheme.

5.1. Experiment 1: Emphasizing personal information with DPP$_{fd}$-EXPEI

As previously mentioned, DPP$_{fd}$-EXPEI allows selecting and weighting terms according to their ability to capture personal statements from the authors. Thus, the purpose of this experiment is to evaluate the relevance of the DPP$_{fd}$-EXPEI approach for detecting social media users suffering from depression. As part of this evaluation, we compared the performance obtained by DPP$_{fd}$-EXPEI and the traditional Information Gain-Term Frequency (IG-TF) approach. It is known that attributes with negative Information Gain (IG) do not help to the classification. Considering this aspect, we tried to carry out a fair assessment using the same number of attributes in each representation. This number was estimated by the number of attributes with positive IG: 1000 and 1500 attributes for the eRisk 2017 and 2018 collections, respectively. Fig. 2 shows the obtained results. We noted that *i*) the proposed adaptation DPP$_{fd}$-EXPEI obtained better results than the original approach DPP-EXPEI, suggesting that the strategy followed by DPP_{fd} succeeds in capturing terms strongly related to the depression domain; *ii*) DPP$_{fd}$-EXPEI considerably outperformed the traditional IG-TF approach, confirming the relevance and usefulness of personal statements for the task in hand, and suggesting that the terms in personal statements are valuable even when they are not very frequent in a global way.

To deeply understand the performance of the proposed approach, we compared the relevant attributes extracted by DPP_{fd} and IG. Fig. 3 shows the results of this comparison. In the figure, the red color and vertical orientation represent the words identified as relevant attributes for both IG and DPP_{fd} schemes. On the other hand, the green color and horizontal orientation show the words distinguished by only DPP_{fd}. The scheme DPP_{fd} captures the relevance of words by their context and not by their frequency. It is worth noting that, words relevant only to DPP_{fd} are highly and intuitively related to the depression domain, for example: *suicide*, *scared*, *upset*, *nervous*, *crying*, *uncomfortable*, *asleep*, and *mental*. Although the DPP_{fd} focuses on the discriminative terms of the positive class, it is possible that some terms associated with the negative class rank in the top positions. For example, the next words are probably closely related to negative class: *happiness* and *birthday*. This makes evident that DPP_{fd} enriches the selection of terms since it captures terms maybe that are no globally frequent, but relevant in personal contexts. Therefore, this shows that depressed users manifest significant signs of depression in personal statements which are not detected by schemes based on the global frequency such as IG.

5.2. Experiment 2: Early depression detection using DPP$_{fd}$-EXPEI

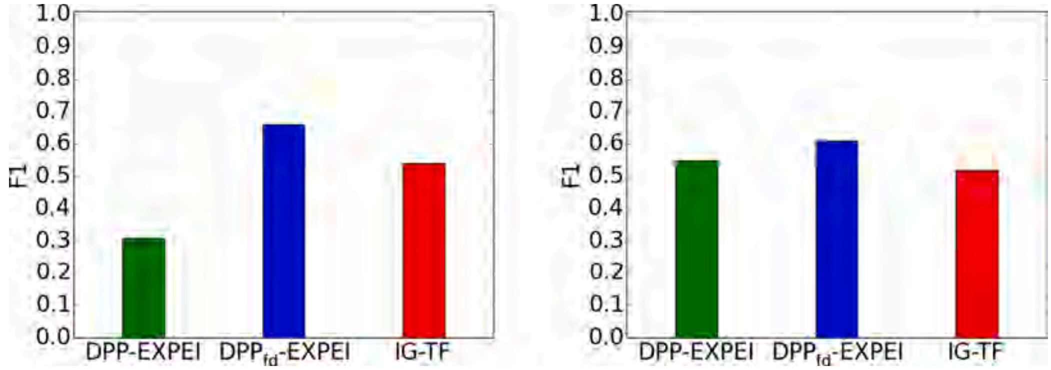
The purpose of early detection is to identify users with depression as soon as possible, that is, to generate an alert on their mental health status considering as few information as possible. This section presents the evaluation of DPP$_{fd}$-EXPEI under this scenario considering the evaluation framework used at the 2017 and 2018 eRisk shared tasks. In the following subsections, we describe the eRisk evaluation framework, the way we configured and tuned our approach for emitting early decisions, and the performance of DPP$_{fd}$-EXPEI in comparison to state-of-the-art results.

5.2.1. Evaluation framework of the 2017 and 2018 eRisk shared tasks

The early detection challenge in the eRisk forum consists on processing chronologically writings of users and detecting signs of depression as soon as possible. To simulate an early detection scenario, the sequence of users' writings in the test partitions (described in Section 3.1.1) were divided into 10 sequential chunks. The first chunk contains the oldest 10% of the messages, the second chunk contains the second oldest 10%, and so forth.

The task was organized into training and test stages. In the training stage the whole documents from the set of training users was released. Whereas, in the test stage, the 10 sequential releases of data (chunks) from the test users were provided. After the release of each data chunk, and before the next release, the systems must emit a decision about each user. Two choices are possible: to emit a decision (depressed or non-depressed) or to make no decision, which means waiting for more chunks. Once a decision is emitted for a user, it is immutable in later chunks. For evaluating the participating methods, two aspects are considered: *i*) the correctness of the output, and *ii*) the delay to emit a decision, based on the number of chunks needed to emit the decision. The first aspect is evaluated by means of the F_1 measure over the positive class. The second aspect is considered by a new measure, $ERDE_o$ (*Early Risk Detection Error*), which penalizes decisions after reading o writings (i.e., posts), and rewards early alerts. This measure was introduced in [37], where they proposed to use the average $ERDE_5$ and $ERDE_{50}$ as performance scores for this task.

[8] Experimentally, several values for m were evaluated (e.g., 2, 3, 4, and 5) finding that $m = 2$ (square root) is suitable for the purpose.



(a) eRisk 2017 (b) eRisk 2018

Fig. 2. Results (F1 values of the depressed class) of the DPP$_{fd}$-EXPEI approach in depression detection on the eRisk 2017 and 2018 collections. A comparison was realized against both the original DPP-EXPEI and the traditional IG-TF approaches. We evaluated on both corpora using all writings of users in the test partition.

Fig. 3. Top-50 terms selected by DPP_{fd} in the eRisk 2018 corpus. The vertical orientation and the red color represent words selected by both *IG* and DPP_{fd}. The horizontal orientation and the green color indicate the words disregarded by *IG* but relevant for DPP_{fd}. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

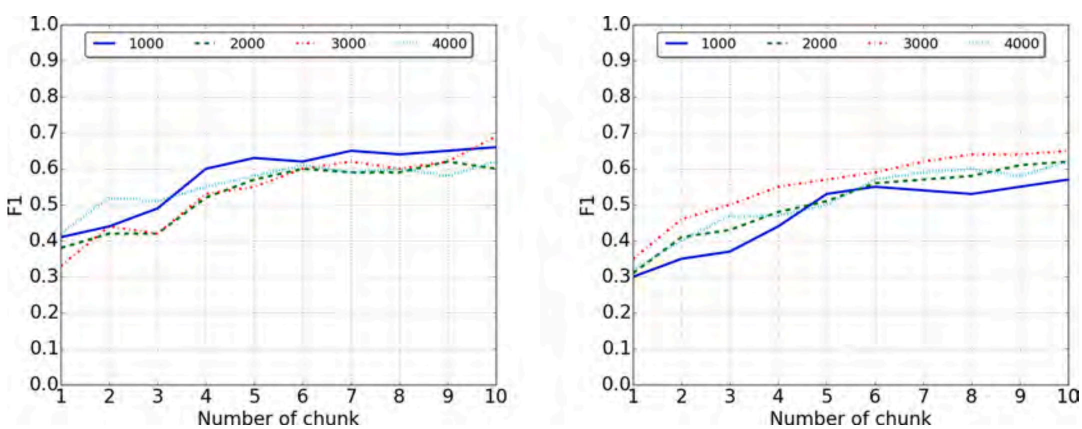
5.2.2. Tuning the approach for early decisions

This section shows the tuning of the approach for early detection. Two aspects were considered, the number of features (terms) used in the representation and the criteria for emitting early decisions.

5.2.2.1. Size of the representation. To determine an adequate number of features for early detection, we evaluated the detection performance considering the top *n* most relevant terms as selected by DPP$_{fd}$, with n = 1000,2000,…4000. Results are presented in Fig. 4. Fig. 4(a) shows that for the eRisk 2017 collection, the proposed approach takes advantage of 1000 terms in most chunks, whereas, in the eRisk 2018 collection, the use of 3000 terms is predominantly better than the other configurations. These results indicate that the size of representation (number of terms) is related to the number of instances of the training set, i.e., the larger the number of users, the larger the number of depression-related terms that DPP could find.

5.2.2.2. Criterion for producing early decisions. As previously explained, under the early detection scenario, classifiers must be able to decide whether to produce an alert (classify a user as depressive) or wait to read a next chunk of information. To achieve this behaviour in our approach, we considered the following decision heuristic: if after reading *n* chunks of information our approach classifies a user as depressive, then his/her post history is checked for the presence of *some keywords* corresponding to the depression domain, in the case it contains three or more of them, then the decision is confirmed and an alert is generated, otherwise the decision is omitted and postponed until the reading of the next chunk.

For the experiments, we built the domain-specific vocabulary using the 50-top terms selected by DPP$_{fd}$. It comprises the set of words that reveal the most quantity of personal information from the depressed users of the training sets. Table 5 shows some words included in such vocabulary (listed in alphabetical order); as it can be noticed, most of these terms are highly associated with the depression domain.



(a) eRisk 2017 dataset (b) eRisk 2018 dataset

Fig. 4. F1 values of the depressed class obtained by tuning the proposed approach using different numbers of attributes (1000, 2000, 3000, and 4000) on the test datasets according the chunks released chronologically in both competitions.

Table 5
Keywords of the depression domain used in the criterion of early alerts.

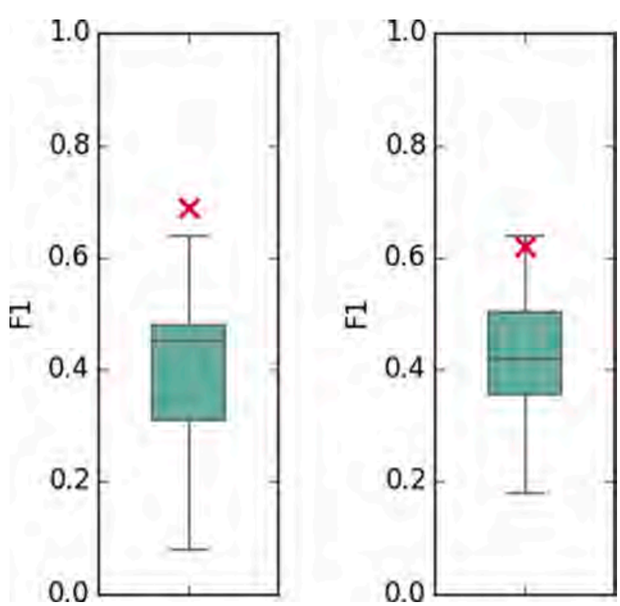
anxiety, asleep, attractive, birthday, boyfriend, cry, crying, dating, depressed, depression, diagnosed, disorder, emotional, emotionally, feeling, feelings, happiness, healthy, helped, helpful, lately, lonely, medication, meds, mental, mood, nervous, panic, personality, relate, relationship, relationships, scared, sleeping, suicidal, suicide, talked, therapist, therapy, uncomfortable, upset, woke,...

5.2.3. *Results of the approach in early depression detection*

The goal of this experiment was to evaluate the proposed approach in the early detection scenario, and to compare it against state-of-the-art results in both shared tasks, eRisk 2017 and 2018.

Fig. 5 shows the distribution of the official results in the shared tasks according to the F_1 metric. In the figure, the red cross indicates the results of the proposed approach. From this figure, it is possible to observe competitive results with respect to participating teams. It should be noted that the DPP$_{fd}$-EXPEI approach significantly outperforms the winner from the eRisk 2017 shared task. In the eRisk 2018 dataset, the performance was slightly lower than the winner; this result would place our approach in the second position. Coincidentally, in the 2017 and 2018 editions of eRisk, the best F_1 score was achieved by the same team [9,50]. They proposed a robust system based on a broad set of features, ranging from hand-crafted features (for example, specific terms related to antidepressants) to automatically learned features extracted by some NNs, in particular some LSTMs arrangements.

In contrast to the winner approach, ours is very simple and intuitive. It is based on a BOW representation, which uses terms associated to the personal communications of depressed users as features. Furthermore, it offers the possibility to interpret the results by tracking the reasons of the classifier decisions. For example, the words that are more relevant (personal and related to depression) in a given post could be identified. For instance, in the sentences in Fig. 6, which correspond to a specific test user, we depict the estimated relevance of the words according to their EXPEI values (the greatest the intensity, the greatest their relevance). It is possible to observe that these words (such as *anger, anxiety, diagnosed broke, crying, mood, anger*, among others) are highly associated to the depression-sickness.



(a) eRisk 2017 (b) eRisk 2018

Fig. 5. Distribution of F1 results (of the depressed class) in eRisk 2017 and eRisk 2018 shared tasks. The red crosses represent the results obtained by the DPP$_{fd}$-EXPEI approach using the same methodology and datasets. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

From a different perspective, Fig. 7 shows the distribution of the official results according to the ERDE measure. ERDE is an error measure, meaning that the lower the error, the better the performance of the system. Although the proposed approach did not improve the best results in both collections, it obtained results very close to the winners. Similar to the general result trend, the approach performed better in ERDE$_{50}$ than in ERDE$_5$. This suggests that the approach requires a minimum quantity of information to be able to emit a positive decision.

Specifically, based on ERDE$_{50}$ the approach would rank in third place in both 2017 and 2018 collections. According to ERDE$_5$, the approach would rank in the fifth place in 2017 and the fourteenth place (with only six teams above) in 2018. Although we achieved our worst result with ERDE$_5$, it is not a bad result since it is the average performance of all the participants (it is almost in the limit of quartile one). These results show the pertinence of the DPP$_{fd}$-EXPEI approach for early detection.

5.3. *Results' analysis*

5.3.1. *Assessing the robustness of DPP$_{fd}$-EXPEI*

In order to evaluate the robustness of the DPP$_{fd}$-EXPEI approach, we carried out some additional experiments using various classification algorithms in addition to the SVM. In these experiments we also considered the IG-TF baseline approach. In particular, we used the following classifiers: k-Nearest Neighbors (denoted as kNN with $k = 3, 5$, and 7), Random Forest (RF), and Naïve Bayes (NB). In all cases we employed their Sklearn implementations and considered the original distributions of the collections, as introduced in Section.

Table 6 shows the obtained results for the eRisk 2017 and 2018 data sets. With three of the four classifiers used, DPP$_{fd}$-EXPEI performed better than IG-TF, confirming its suitability for the task at hand. In both collections, the best performance measured over the *depressed* category was obtained with the SVM classifier in combination with DPP$_{fd}$-EXPEI. On the contrary, in both collections, and with both approaches, the worst results were obtained with Random Forest, showing that it does not correctly handle very dispersed data, as is the case of social media texts related to depression.

5.3.2. *A closer look at the SVM results*

This section presents a detailed analysis of the results achieved by the SVM classifier. Table 7 shows the precision (P), recall (R), and F_1 scores for the DPP$_{fd}$-EXPEI and IG-TF approaches. It also includes the confusion matrix for each data set.

From Table 7 a clear pattern emerged, the proposed DPP$_{fd}$-EXPEI approach is more strict to emit a positive decision than the traditional IG-TF method, thus showing a trend for lower recall levels but higher precision rates. This can be clearly noticed in the two confusion matrices when comparing the number of false positives and false negatives of both approaches. The difference in false positives is very noticeable, ours has only 15 and 40 for the 2017 and 2018 collections respectively, while IG-TF has 41 and 58. On the other hand, for what it concerns to false negatives, the rates between both methods and collections are more similar; our approach has 19 and 27, while IG-TF has 18 and 31.

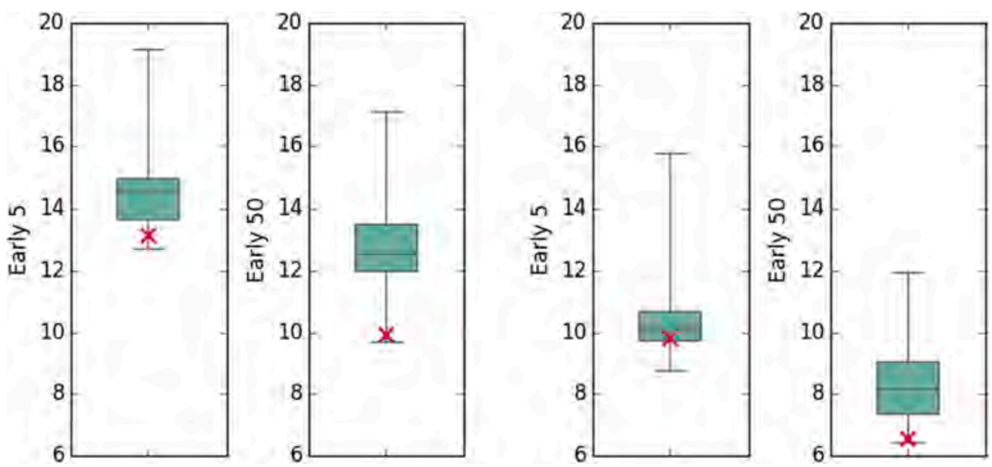
5.3.3. *A qualitative analysis of the classification errors*

To deeply understand the results of the proposed approach, we conducted a qualitative analysis of the false positives and negatives reported in the previous section.

Among the main causes of *false positive failures* we found the following:

- Users talking about other people's depression. There are several users sharing the experiences of a family member suffering from depression.

Fig. 6. Example of the relevance for the words in a given sentence according to the EXPEI term weighting scheme. The intensity of the red color is related to EXPEI values in a range [0-1]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



(a) eRisk 2017 (b) eRisk 2018

Fig. 7. Distribution of results according to ERDE values in eRisk 2017 and 2018 shared tasks. The red crosses indicate the values obtained by the proposed approach (DPP$_{fd}$-EXPEI). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 6
F-score rates over the positive class using different classification algorithms, when DPP$_{fd}$-EXPEI and IG-TF are used for building the text representations.

Dataset	Approach	3NN	5NN	7NN	RF	NB	SVM
eRisk 2017	DPP$_{fd}$-EXPEI	0.53	0.49	0.53	0.19	0.54	**0.66**
	IG-TF	0.43	0.41	0.36	0.31	0.12	0.54
eRisk 2018	DPP$_{fd}$-EXPEI	0.50	0.52	0.49	0.25	0.52	**0.61**
	IG-TF	0.34	0.37	0.40	0.34	0.26	0.52

- Users writing about other diseases with similar clinical conditions. Some users describe symptoms very similar to depression but that correspond to other kinds of mental disorders such as schizophrenia, bipolar disorder, among others.
- Users commenting on the topic but not suffering from depression. We read many documents presumably written by psychologists or students researching or explaining about this disorder.

All previous cases share a common explanation, the content of these documents has a large intersection with the vocabulary associated with the depression disorder. This situation highlights a problem of our approach, that it is based on a BoW representation, and, consequently, it does not allow us to distinguish among different usage contexts. However, we consider that the approach works fine and it could be very useful in real scenarios, since these three cases could be in fact considered as alerts.

Table 7
Results obtained with the SVM. On the left, the Precision, Recall, and F_1 scores for the two collections. On the right, the confusion matrices for both experiments, where rows represent the real classes and columns the predicted ones. The depressed and non-depressed classes are denoted as *dep* and *non-dep*, respectively.

Dataset	Method	Class	Metrics			Confusion Matrix	
			P	R	F	*dep*	*non-dep*
eRisk 2017	DPP$_{fd}$-EXPEI	*dep*	0.69	0.63	0.66	33	19
		non-dep	0.95	0.96	0.95	15	334
	IG-TF	*dep*	0.45	0.65	0.54	34	18
		non-dep	0.94	0.88	0.91	41	308
eRisk 2018	DPP$_{fd}$-EXPEI	*dep*	0.57	0.66	0.61	52	27
		non-dep	0.96	0.95	0.95	40	701
	IG-TF	*dep*	0.45	0.61	0.52	48	31
		non-dep	0.96	0.92	0.94	58	683

Regarding *false negative failures*, an opposite, general and common cause was observed: most of the documents show little intersection with the vocabulary associated with the depression domain. For example, several depressed users mentioned signs of their depression in a very few personal statements, but, in the general, they shared content about very diverse topics. Perhaps, this behaviour indicates a single episode of depression, and suggests that the individual has overcome the illness.

6. Conclusions and future work

This paper explored the relevance of personal statements in depression detection using texts from social media. Our work was inspired in two ideas from a psychological perspective: *i*) there is a relationship between the language use and the mental state, and *ii*) there is a link of self-focus with depression. Particularly, we hypothesized that people talking about themselves tend to expose valuable information that can reveal their mental state, especially, traces of a depressive disorder. To assess this idea, we proposed and evaluated the DPP$_{fd}$-EXPEI approach, which emphasizes the value of the personal information in the representation construction. Our results showed that: *i*) phrases with pronouns from the *I* and *you* families represent an important source of discriminative traces of depression; *ii*) the use of methods that emphasize this type of information enhance the detection task; and *iii*) the DPP$_{fd}$-EXPEI approach is simple but effective for early depression detection, and at the same time it is more transparent (with interpretable results) than state-of-the-art approaches. Particularly, DPP$_{fd}$-EXPEI allows us to capture relevant clues for depression detection in self-references, where people tend to reflect their interests, opinions, and problems. Then, throughout the new term weighting approach it is possible to pay particular attention on words related to the depression domain (like suicide, scared, nervous, etc.) which could be easily interpreted by mental health specialists in contrast to the more abstract representations used by other current detection approaches.

As future work, we are interested in confirming the obtained results using other datasets, not only in English but also in other languages. Also, we would like to analyze the impact of the datasets' sizes and class imbalance in the performance of our approach. Besides, we are planning to explore the evolution of the words' relevance according to their weights in each data chunk in early contexts. The obtained results encourage us to explore new distributed representations for better taking advantage of personal information. Furthermore, we are interested in adapting machine learning algorithms (even deep learning architectures) to emphasize this type of information in the classification phase.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was funded by CONACYT (postdoctoral fellowship CVU-174410 and project FC 2016-2410).

References

[1] Conti D, Burton W. The economic impact of depression in a workplace. J Occup Med 1994;36(9):983–8.
[2] Hysenbegasi A, Hass SL, Rowland CR. The impact of depression on the academic productivity of university students. J Mental Health Policy Econ 2005;8:145–51.
[3] Luca A, Luca M, Calandra C. Sleep disorders and depression: brief review of the literature, case report, and nonpharmacologic interventions for depression. In: Clinical interventions in aging; 2013.
[4] Casper RC. Depression and eating disorders. Depression and Anxiety 1998;8(S1): 96–104. https://doi.org/10.1002/(SICI)1520-6394(1998)8:1+<96::AID-DA15>3.0.CO;2-4.
[5] Zubrick SR, Hafekost J, Johnson SE, Sawyer MG, Patton G, Lawrence D. The continuity and duration of depression and its relationship to non-suicidal self-harm and suicidal ideation and behavior in adolescents 12-17. J Affect Disord 2017;220: 49–56. https://doi.org/10.1016/j.jad.2017.05.050.
[6] Aragón ME, López-Monroy AP, González-Gurrola LC, Montes-y-Gómez M. Detecting depression in social media using fine-grained emotions. In: Proceedings of the 2019 conference of the North American chapter of the Association for Computational Linguistics: human language technologies, volume 1 (long and short papers). Minneapolis, Minnesota: Association for Computational Linguistics; 2019. p. 1481–6.
[7] Losada DE, Crestani F, Parapar J. eRISK 2017: CLEF lab on early risk prediction on the internet: experimental foundations. In: Jones GJF, Lawless S, Gonzalo J, Kelly L, Goeuriot L, Mandl T, et al., editors. Experimental IR meets multilinguality, multimodality, and interaction - 8th International Conference of the CLEF Association, CLEF 2017, Dublin, Ireland, September 11–14, 2017, Proceedings, Vol. 10456 of Lecture Notes in Computer Science. Springer; 2017. p. 346–60.
[8] Losada DE, 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] Trotzek M, Koitka S, Friedrich C. Word embeddings and linguistic metadata at the clef 2018 tasks for early detection of depression and anorexia. In: Proceedings of the Ninth International Conference of the CLEF Association (CLEF 2018). France: Avignon; 2018.
[10] De Choudhury M, De S. Mental health discourse on reddit: self-disclosure, social support, and anonymity. Proc Eight Int AAAI Conf Weblogs Social Media 2014: 71–80.
[11] 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, Maryland, USA: Association for Computational Linguistics; 2014. p. 51–60.
[12] Pennebaker J. The secret life of pronouns: what our words say about us. USA: Bloomsbury; 2011.
[13] Pennebaker J, Mehl M, Niederhoffer K. Psychological aspects of natural language use: our words, our selves. Annu Rev Psychol 2003;54:547–77. https://doi.org/10.1146/annurev.psych.54.101601.145041.
[14] Rude S, Gortner E-M, Pennebaker J. Language use of depressed and depression-vulnerable college students. Cognition and Emotion 2004;18(8):1121–33.
[15] Brockmeyer T, Zimmermann J, Kulessa D, Hautzinger M, Bents H, Friederich H-C, et al. Me, myself, and i: self-referent word use as an indicator of self-focused attention in relation to depression and anxiety. Frontiers in Psychology 2015;6: 1564. https://doi.org/10.3389/fpsyg.2015.01564.
[16] Zimmermann J, Brockmeyer T, Hunn M, Schauenburg H, Wolf M. First-person pronoun use in spoken language as a predictor of future depressive symptoms: preliminary evidence from a clinical sample of depressed patients. Clin Psychol Psychother 2017;24(2):384–91. https://doi.org/10.1002/cpp.2006.
[17] De Choudhury M, Counts S, Horvitz E. Social media as a measurement tool of depression in populations. In: Proceedings of the 5th Annual ACM Web Science Conference, WebSci '13. New York, NY, USA: ACM; 2013. p. 47–56.
[18] Husseini Orabi A, Buddhitha P, Husseini Orabi M, Inkpen D. Deep learning for depression detection of twitter users. In: Proceedings of the fifth workshop on computational linguistics and clinical psychology: from keyboard to clinic. New Orleans, LA: Association for Computational Linguistics; 2018. p. 88–97.
[19] Ramírez-Esparza N, Chung C, Kacewicz E, Pennebaker J. The psychology of word use in depression forums in english and in spanish: Testing two text analytic approaches. In: ICWSM 2008 - proceedings of the 2nd international conference on weblogs and social media; 2008.
[20] Shen T, Jia J, Shen G, Feng F, He X, Luan H, et al. Cross-domain depression detection via harvesting social media. In: Proceedings of the twenty-seventh International Joint Conference on Artificial Intelligence, IJCAI-18. International Joint Conferences on Artificial Intelligence Organization; 2018. p. 1611–7. https://doi.org/10.24963/ijcai.2018/223.
[21] Ortega-Mendoza RM, Franco-Arcega A, López-Monroy AP, 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, et al., editors. Experimental IR meets multilinguality, multimodality, and interaction: 7th international conference of the CLEF Association, CLEF 2016, Évora, Portugal, September 5-8, 2016, Proceedings. Cham: Springer International Publishing; 2016. p. 110–22.
[22] Ortega-Mendoza RM, López-Monroy AP, Franco-Arcega A, Montes-y-Gómez M. Emphasizing personal information for author profiling: new approaches for term selection and weighting. Knowl-Based Syst 2018;145:169–81.
[23] Paul M, Dredze M. You are what you tweet: analyzing twitter for public health. In: International AAAI conference on web and social media; 2011.
[24] 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. Denver, Colorado: Association for Computational Linguistics; 2015. p. 1–10.
[25] Mitchell M, Hollingshead K, Coppersmith G. Quantifying the language of schizophrenia in social media. In: Proceedings of the workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality. Denver, Colorado: Association for Computational Linguistics; 2015.
[26] De Choudhury M, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. In: Proceedings of the SIGCHI conference on human factors in computing systems, CHI '13. New York, NY, USA: ACM; 2013. p. 3267–76.
[27] Cash SJ, Thelwall M, Peck SN, Ferrell JZ, Bridge JA. Adolescent suicide statements on myspace. Cyberpsychol Behav Soc Netw 2013;16(3):166–74.
[28] Li A, Jiao D, Zhu T. Detecting depression stigma on social media: a linguistic analysis. J Affect Disord 2018;232:358–62.
[29] Martínez-Castaño R, Pichel JC, Losada DE. A big data platform for real time analysis of signs of depression in social media. Int J Environ Res Public Health 2020;17(13):4752.

[30] M. Nadeem, M. Horn, G. Coppersmith, D. S. Sen, Identifying depression on twitter, CoRR abs/1607.07384.
[31] Chen X, Sykora MD, Jackson TW, Elayan S. What about mood swings: identifying depression on twitter with temporal measures of emotions. In: Companion proceedings of the the web conference 2018, WWW '18, international world wide web conferences steering committee, Republic and Canton of Geneva, Switzerland; 2018. p. 1653–60.
[32] 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, Colorado: Association for Computational Linguistics; 2015. p. 31–9.
[33] De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. AAAI; 2013.
[34] Huang X, Zhang L, Chiu D, Liu T, Li X, Zhu T. Detecting suicidal ideation in chinese microblogs with psychological lexicons. In: 2014 IEEE 11th Intl Conf on Ubiquitous Intelligence & Computing and 2014 IEEE 11th Intl Conf on Autonomic & Trusted Computing and 2014 IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops; 2014. p. 844–9.
[35] De Choudhury M, Counts S, Horvitz EJ, Hoff A. Characterizing and predicting postpartum depression from shared facebook data. In: Proceedings of the 17th ACM conference on computer supported cooperative work & social computing, CSCW '14. New York, NY, USA: ACM; 2014. p. 626–38.
[36] Moreno M, Jelenchick L, Egan K, Cox E, Young HN, Gannon KE, et al. Feeling bad on Facebook: depression disclosures by college students on a social networking site. Depress Anxiety 2011;28(6):447–55.
[37] Losada DE, Crestani F. A test collection for research on depression and language use CLEF 2016, Évora (Portugal), experimental IR meets multilinguality, multimodality, and interaction. 2016. p. 28–9.
[38] Nguyen T, Phung D, Dao B, Venkatesh S, Berk M. Affective and content analysis of online depression communities. IEEE Trans Affect Comput 2014;5(3):217–26. https://doi.org/10.1109/TAFFC.2014.2315623.
[39] Park S, Lee SW, Kwak J, Cha M, Jeong B. Activities on facebook reveal the depressive state of users. J Med Internet Res 2013;15(10):e217.
[40] Jamil Z, Inkpen D, Buddhitha P, White K. Monitoring tweets for depression to detect at-risk users. In: Proceedings of the fourth workshop on computational linguistics and clinical psychology — from linguistic signal to clinical reality. Vancouver, BC: Association for Computational Linguistics; 2017. p. 32–40.
[41] J. Aguilera, D. I. Hernández Farías, R. Ortega-Mendoza, M. Montes, Depression and anorexia detection in social media as a one-class classification problem, Appl Intell doi:https://doi.org/10.1007/s10489-020-02131-2.
[42] Kawachi I, Berkman LF. Social ties and mental health. J Urban Health 2001;78(3):458–67.
[43] Rosenquist J, Fowler J, Christakis N. Social network determinants of depression. Mol Psychiatry 2010;16:273–81.
[44] Losada DE, Crestani F, Parapar J. Overview of eRisk 2019. Early risk prediction on the internet. In: 10th international conference of the CLEF association. Springer; 2019. p. 340–57.
[45] Losada DE, Crestani F, Parapar J. Overview of erisk at CLEF 2020: early risk prediction on the internet (extended overview). In: Conference and labs of the evaluation forum, CEUR workshop proceedings; 2020.
[46] Schwartz HA, Ungar LH. Data-driven content analysis of social media: a systematic overview of automated methods. The Annals of the American Academy of Political and Social Science 2015;659(1):78–94.
[47] Chung C, Pennebaker JW. The psychological function of function words. In: Social communication: frontiers of social psychology. Psychology Press; 2007. p. 343–59.
[48] Morrissette B. Narrative "you" in contemporary literature. Comp Lit Stud 1965;2(1):1–24.
[49] Phillips N. Can "you" help me understand?: second-person narratives and the depiction of depression in a.m. jenkins's damage and chris lynch's freewill. In: The ALAN review. 33; 2005. p. 23–9. https://doi.org/10.21061/alan.v33i1.a.5.
[50] Trotzek M, Koitka S, Friedrich CM. Linguistic metadata augmented classifiers at the CLEF 2017 task for early detection of depression. In: Working notes of CLEF 2017 - conference and labs of the evaluation forum, Dublin, Ireland, September 11-14; 2017. p. 2017.