Detecting Depression in Social Media using Fine-Grained Emotions

Conference Paper - June 2019

4 authors:

Mario Ezra Aragon
Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE)

Adrián Pastor López-Monroy
University of Houston

Luis Carlos Gonzalez Gurrola
Autonomous University of Chihuahua

Manuel Montes
Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE)

Some of the authors of this publication are also working on these related projects:

- Algorithms for Anomaly Detection View project
- NLPinSE View project
Detecting Depression in Social Media using Fine-Grained Emotions

Mario Ezra Aragón*, A. Pastor López-Monroy†, Luis C. González-Gurrola‡ and Manuel Montes-y-Gómez*

* Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Mexico.
† Centro de Investigación en Matemáticas (CIMAT), Mexico.
‡ Facultad de Ingeniería, UACh, Mexico.
mearagon@inaoep.mx, pastor.lopez@cimat.mx, lcgonzalez@uach.mx, mmontesg@inaoep.mx

Abstract

Nowadays social media platforms are the most popular way for people to share information, from work issues to personal matters. For example, people with health disorders tend to share their concerns for advice, support or simply to relieve suffering. This provides a great opportunity to proactively detect these users and refer them as soon as possible to professional help. We propose a new representation called Bag of Sub-Emotions (BoSE), which represents social media documents by a set of fine-grained emotions automatically generated using a lexical resource of emotions and subword embeddings. The proposed representation is evaluated in the task of depression detection. The results are encouraging: the usage of fine-grained emotions improved the results from a representation based on the core emotions and obtained competitive results in comparison to state of the art approaches.

1 Introduction

Mental Disorders affect millions of people around the world. Out of these disorders, depression has been ranked among the most common, even with a high incidence in mortality rates (Kessler et al., 2017; Mathers and Loncar, 2006). It is imperative then, to come with effective approaches to detect depression before it causes irreparable damage to mere individuals that suffer it and their loved ones. In a connected world where we live, it is very normal to share personal information, matters and concerns in social media platforms. This fact poses an opportunity, since the understanding of depression through the analysis of social media documents increases the chances to detect people that present signs of depression and could lead to provide them professional help (Guntuku et al., 2017; Pestian et al., 2010).

Several works in literature have explored how to use linguistic and sentiment analysis to detect depression (Xue et al., 2013). For example, in (Huang et al., 2014) the authors applied sentiment analysis (SA) to assign polarity to tweets. They count the number of positive, negative, neutral words, and the ratio of the negative and positive words, and found that depressed users post longer emotional tweets. The work of Wang et al. (2013) enriched SA with features derived from psychological research like the use of first person pronouns, user social interaction and user behaviors in microblogs. An interesting finding is that the time of the posts is useful to detect people with high risk of committing suicide. In a recent work, Chen et al. (2018) proposed to use emotions with the aim to identify depression on Twitter users. That study openly exposed the potential of using discrete emotions as features, instead of only using linguistic features, and broad categories to represent them. To further investigate this latter point, in this study we propose to model emotions in a fine-grained way and use them to build a new representation to tackle the problem of detecting depression in users of social media. We construct these fine-grained emotions using lexical information extracted from emotions combined with subword embeddings. The leading hypothesis of our study is that emotions could be better, and more flexible, represented at a lower level, instead of only using broad categories such as “anger”, “joy”, “negative” or “positive”.

2 The Bag of Sub-Emotions (BoSE) Representation

Figure 1 depicts our proposed approach. In a first step, we compute a set of fine-grained emotions for each broad emotion described in the lexical resource by Mohammad and Turney (2013). Then,
we use the obtained fine-grained emotions to mask the texts, eventually representing them by a histogram of their frequencies. Accordingly, we named this new representation \textbf{BoSE}, for Bag of Sub-Emotions. In the following sections we detail each step of our proposed approach.

2.1 Generating Fine-Grained Emotions

To generate the fine-grained emotions we use a lexical resource based on eight recognized emotions, e.g., Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise and Trust (Ekman and Davidson, 1994), and two main sentiments\(^1\), Positive and Negative. We represent this as \(E = \{E_1, E_2, ..., E_{10}\}\), where \(E\) is the set of emotions in the lexical resources and \(E_i = \{w_1, ..., w_n\}\) is the set of words associated to the emotion \(E_i\). We compute a word vector for each word using pre-trained Wikipedia sub-word embeddings from FastText (Bojanowski et al., 2016) of size 300, and then we create subgroups of words by emotion using the \textit{Affinity Propagation} (AP) clustering algorithm (Thavikulwat, 2008). This AP clustering algorithm has several appealing characteristics, e.g. it does not employ artificial elements (centroids) to create clusters and it does not require to specify the number of groups before running the algorithm. To have an idea of how the vocabulary was distributed among emotions and the number of generated clusters after applying AP to the lexical resource we present Table 1.

After this process, now each (broad) emotion is represented by a set of fine-grained emotions, \(E_i = \{F_{i1}, ..., F_{ij}\}\), where each \(F_{ij}\) is a subset of the words from \(E_i\) and is represented by the average vector of their respective embeddings. These subgroups of words allow separating each broad emotion in different topics that help identify and capture the fine-grained emotions expressed by users in their posts. Figure 2 presents some examples of groups of fine-grained emotions that were automatically computed by our approach.

---

\(^1\)In the rest of the paper we refer to these sentiments as emotions as well.
**Table 1:** Size of the vocabulary for each emotion presented in the lexical resources, and number of generated clusters.

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Vocabulary</th>
<th>Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>anger</td>
<td>6035</td>
<td>444</td>
</tr>
<tr>
<td>anticipation</td>
<td>5837</td>
<td>395</td>
</tr>
<tr>
<td>disgust</td>
<td>5285</td>
<td>367</td>
</tr>
<tr>
<td>fear</td>
<td>7178</td>
<td>488</td>
</tr>
<tr>
<td>joy</td>
<td>4357</td>
<td>318</td>
</tr>
<tr>
<td>sadness</td>
<td>5837</td>
<td>395</td>
</tr>
<tr>
<td>surprise</td>
<td>3711</td>
<td>274</td>
</tr>
<tr>
<td>trust</td>
<td>5481</td>
<td>383</td>
</tr>
<tr>
<td>positive</td>
<td>11021</td>
<td>740</td>
</tr>
<tr>
<td>negative</td>
<td>12508</td>
<td>818</td>
</tr>
</tbody>
</table>

Figure 2: Examples of words grouped by Fine-Grained Emotions.

3 Experimental Settings

**Preprocessing:** For our experiments, we normalized the texts by removing special characters and lowercasing all the words. After preprocessing we masked the texts using the fine-grained emotions.

**Classification:** Once built the BoSE representation, we selected the more relevant features (i.e., sequences of fine-grained emotions) using the \( \chi^2 \) distribution \( X_k^2 \) (Walcik, 2007). Then, we used a Support Vector Machine (SVM) with a linear kernel and \( C = 1 \) to classify the documents.

**Baselines:** To properly evaluate the relevance of using fine-grained emotions in the detection of depression, we considered a representation based on the occurrences of broad emotions and the words that do not have an associated emotion. We named this approach Bag-of-Emotions (BoE). We also compared our results against a Bag-of-Words representation based on word unigrams and n-grams, since they are the common baseline approaches for text classification. Additionally, we compared our results against the \( f_1 \) results from the participants of the eRisk 2017 and 2018 evaluation tasks (Losada et al., 2017, 2018).
Table 2: Depression data sets used for experimentation. Each data set have two classes (Depressed = Dep, Non-Depressed = ND).

<table>
<thead>
<tr>
<th>Data set</th>
<th>Training</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dep</td>
<td>ND</td>
</tr>
<tr>
<td>eRisk’17</td>
<td>83</td>
<td>403</td>
</tr>
<tr>
<td>eRisk’18</td>
<td>135</td>
<td>752</td>
</tr>
</tbody>
</table>

**Data Collections:** We evaluated our approach in the task of depression detection, using the data sets from the eRisk 2017 and 2018 evaluation tasks (Losada et al., 2017, 2018). These data sets contain Reddit posts for several users. The users which explicitly mentioned that were diagnosed with depression were automatically labeled as positive. Vague expressions like "I think I have depression" or "I’m depressed" were discarded, the rest of them were labeled as negative. Table 2 shows some numbers from these data sets; (Losada and Crestani, 2016) describes these collections in more detail.

4 Experimental Results

The goal of our first experiment was to evaluate the appropriateness of the BoSE representation to identify depressed users. To accomplish this, we compared its performance against the results from a traditional BOW representation as well as a representation considering only the broad emotions. Table 3 shows the $f_1$ performance over the positive class for the BOW, BoE and BoSE approaches. It can be noticed that the BoSE representation outperforms both baseline results, particularly when sequences of fine-grained emotions were considered. To better characterize the BoSE representation, we evaluated it without considering the clusters associated to the positive and negative sentiments. We refer to these experiments as BoSE8. Results from this variant show a drop in performance, confirming that sentiment information is relevant to the identification of depressed users.

To further evaluate the relevance of the BoSE representation, Table 4 compares its results against those from the first three places at the eRisk 2017 and 2018 evaluation tasks (Losada et al., 2017, 2018). To contextualize this comparison, consider that the first place in both years (Trotzek et al., 2017, 2018) defined multiple strategies and considered a wide range of features to build their models, e.g., they extract readability features, LIWC features, user-level linguistic metadata, neural word embeddings, specific terms related to depression, and used models based on LSTM neural networks and convolutional neural networks, using four machine learning models in an ensemble model. Other top performers (Villegas et al., 2017) combined semantic representation considering partial information and temporal variation features. In (Funez et al., 2018) they implemented two models; one based on flexible temporal variation of terms and a second model based on sequential incremental classification.

$$
\begin{array}{|c|c|c|}
\hline
\text{Method} & \text{Dep’17} & \text{Dep’18} \\
\hline
\text{BoW-unigrams} & 0.59 & 0.58 \\
\text{BoE-unigrams} & 0.57 & 0.60 \\
\text{BoSE8-unigrams} & 0.56 & 0.60 \\
\text{BoSE-unigrams} & 0.61 & 0.61 \\
\hline
\text{BOW-ngrams} & 0.58 & 0.60 \\
\text{BoE-ngrams} & 0.61 & 0.58 \\
\text{BoSE8-ngrams} & 0.57 & 0.59 \\
\text{BoSE-ngrams} & 0.64 & 0.63 \\
\hline
\end{array}
$$

Table 3: F1 results over the positive class against baseline methods

$$
\begin{array}{|c|c|c|}
\hline
\text{Method} & \text{Dep’17} & \text{Dep’18} \\
\hline
\text{first place} & 0.64 & 0.64 \\
\text{second place} & 0.59 & 0.60 \\
\text{third place} & 0.53 & 0.58 \\
\text{BoSE-ngrams} & 0.64 & 0.63 \\
\hline
\end{array}
$$

Table 4: F1 results over the positive class against top performers at eRisk

From the obtained results we highlight the following observations:

1. Our approach outperformed the traditional BOW representation in both data sets, indicating that considering emotional information is quite relevant for the detection of depression in online communications.

2. The use of fine-grained emotions as features helps improving the results from a representation that only considers broad emotions. This result confirms our hypothesis that depressive users tend to express their emotions in a different way than non depressive users.
4.1 Analysis of the Fine-Grained Emotions
To offer a glimpse of what fine-grained emotions actually capture, we selected the most relevant sequences for the detection of depression according to the $\chi^2$ distribution. Table 5 shows some relevant sequences of fine-grained emotions as well as some examples of the words that correspond to these sequences.

Most of the fine-grained emotions that present high relevance for the detection of depression are related to negative topics, for example, the anger emotion is associated to the feeling of abandonment or unsociable, and the disgust emotion is related to dilution, insecurity and desolation. These fine-grained emotions seem to capture the way a depressed user expresses about himself or his environment.

5 Conclusions and Future Work
In this study we proposed a new representation that creates fine-grained emotions that were automatically generated using a lexical resource of emotions and sub-word embeddings from FastText. Using these fine-grained emotions our approach can automatically capture more specific topics and emotions that are expressed in the documents by users that have depression. BoSE obtained better results than the proposed baselines and also improved the results of only using broad emotions. It is worth mentioning the simplicity and interpretability of our approach, which contrasts with the best previous eRisk competition methods that are much more complex and difficult to interpret (most of the participants used plenty of different features and a vast range of models, including deep). Our results encourage to attempt this approach based on fine-grained emotions in other relevant health and safety tasks such as the detection of anorexia and self-harm. In addition, we also plan to explore the learning of emotion-based representations by means of a deep neural network from which we could exploit local invariance properties to model fine-grained emotions.

Acknowledgments
This research was supported by CONACyT-Mexico (Scholarship 654803 and Project FC-2410).

References


Xiaolei Huang, Lei Zhang, Tianli Liu, David Chiu, Tingshao Zhu, and Xin Li. 2014. Detecting suicidal ideation in chinese microblogs with psycholog-


