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 micro blogs. 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 as "anger", "joy", "negative" or "positive".

2 The Bag of Sub-Emotions (BoSE) Representation

Figure 1 depicts ourproposed 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,

Proceedings of NAACL-HLT 2019, pages 1481–1486
Minneapolis, Minnesota, June 2 - June 7, 2019.

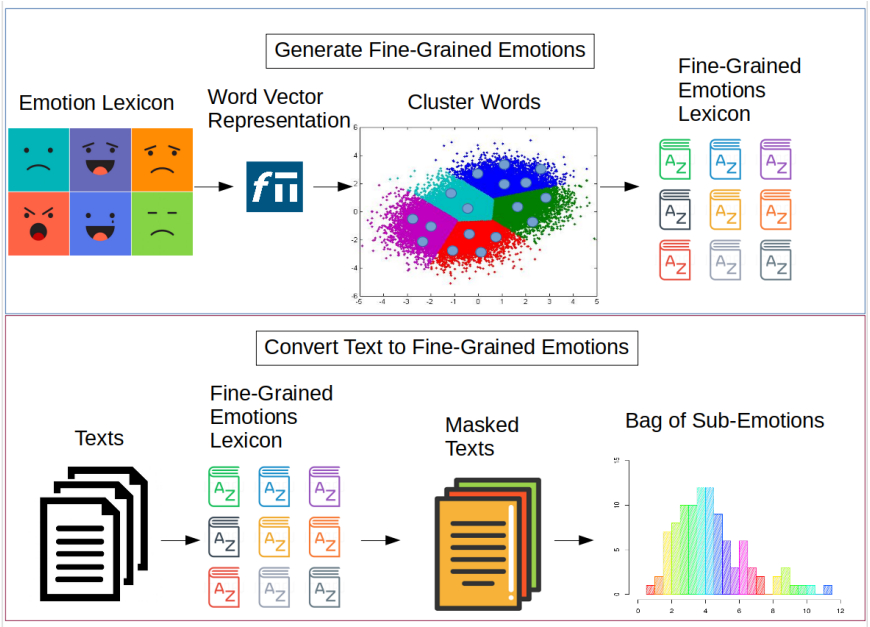

Figure 1: Diagram that represents the creation of the Bag of Sub-Emotions (**BoSE**) representation. First, Fine-Grained Emotions are generated from a given Emotion Lexicon; then, texts are masked using these fine-grained emotions and their histogram is build as final representation.

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 **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 *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. We can appreciate that words with similar context tend to group together, as shown in each column. We can also notice that words corresponding to the same broad emotion consider very different topics, for example, in the Anger emotion, the group *anger3* is related to fighting and battles, whereas the group *anger2* is about growls or loud noises. Another interesting example are the groups from the Surprise emotion, where groups express different kinds of surprises like art and museums (*surprise2*), accidents and disasters (*surprise1*), as well as magic and illusions (*surprise3*).

[1]In the rest of the paper we refer to these sentiments as emotions as well.

Anger			Joy		
anger1	***anger2***	***anger3***	***joy1***	***joy2***	***joy3***
abomination	growl	battle	accomplish	bounty	charity
fiend	growling	combat	achieve	cash	foundation
inhuman	thundering	fight	gain	money	trust
abominable	snarl	battler	reach	reward	humanitarian
unholy	snort	fists	goal	wealth	charitable
Surprise			**Disgust**		
surprise1	***surprise2***	***surprise3***	***disgust1***	***disgust2***	***disgust3***
accident	art	magician	accusation	criminal	cholera
crash	museum	wizard	suspicion	homicide	epidemic
disaster	artwork	magician	complaint	delinquency	malaria
incident	gallery	illusionist	accuse	crime	aids
collision	visual	sorcerer	slander	enforcement	polio

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

Emotion	Vocabulary	Clusters
anger	6035	444
anticipation	5837	395
disgust	5285	367
fear	7178	488
joy	4357	318
sadness	5837	395
surprise	3711	274
trust	5481	383
positive	11021	740
negative	12508	818

Table 1: Size of the vocabulary for each emotion presented in the lexical resources, and number of generated clusters.

2.2 Converting Text to Fine-Grained Emotions

Text masking: In this step documents are masked by replacing each word with the label of its closest fine-grained emotion. To this end, we compute the vector representation of each word using sub-word embeddings from FastText, then we measure the distance of each word vector against the centroid vectors from all fine-grained emotions by means of the cosine similarity, and, finally, we substitute each word by the label of its closest fine-grained emotion. To illustrate this process consider the text "*Leave no stone unturned*", which will be masked as "*fear2 negative8 anger10 anticipation3*".

Text representation: Based on the masked documents, we build their **BoSE** representations computing a frequency histogram of their fine-grained emotions. To build these representations we follow two different approaches: *i*) similar to the Bag-of-Words representation we create a histogram counting the number of occurrences of each fine-grained emotion in the text, we refer to this representation as **BoSE-unigrams**, and *ii*) we create a histogram counting the number of occurrences of fine-grained emotion sequences in the text, we refer to this representation as **BoSE-ngrams**. For the latter representation, we tested different sizes and combinations of sequences; using unigrams and bigrams we obtained the best performance for this task.

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 (Walck, 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).

Data set	Training		Testing	
	Dep	ND	Dep	ND
eRisk'17	83	403	52	349
eRisk'18	135	752	79	741

Table 2: Depression data sets used for experimentation. Each data set have two classes (Depressed = Dep, Non-Depressed = ND).

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 to 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

Method	Dep'17	Dep'18
BoW-unigrams	0.59	0.58
BoE-unigrams	0.57	0.60
BoSE8-unigrams	0.56	0.60
BoSE-unigrams	**0.61**	**0.61**
BOW-ngrams	0.58	0.60
BoE-ngrams	0.61	0.58
BoSE8-ngrams	0.57	0.59
BoSE-ngrams	**0.64**	**0.63**

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

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.

Method	Dep'17	Dep'18
first place	**0.64**	**0.64**
second place	0.59	0.60
third place	0.53	0.58
BoSE-ngrams	**0.64**	0.63

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.

Examples of relevant sequences	
”anger1” ”anger11-anticipation10” ”disgust16-anger11” ”disgust11-fear17”	
anger1	abandoned, deserted, unattended
anger11	unsociable, crowd, mischievous
anticip10	disappointed, inequality, infidelity
disgust16	unsatisfactory, dilution, influence
disgust11	insecurity, desolation, incursion
fear17	hysterical, immaturity, injury

Table 5: Examples of words that create the fine-grained emotions

3. Our approach obtained comparable results to the best reported approaches in both data sets. It is important to highlight that the participants of these tasks tested different complex models with a wide range of features and sophisticated approaches based on traditional and deep learning representations of texts, whereas ours only relies on the use of fine-grained emotions as features.

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 seems 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 emotional-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

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2016. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, pages 135–146.

Xuetong Chen, Martin D. Sykora, Thomas W. Jackson, and Suzanne Elayan. 2018. What about mood swings? identifying depression on twitter with temporal measures of emotions. *Companion Proceedings of the The Web Conference 2018, International World Wide Web Conferences Steering Committee*, pages 1653–1660.

Paul Ed Ekman and Richard J Davidson. 1994. The nature of emotion: Fundamental questions. *New York, NY, US: Oxford University Press*.

Dario G. Funez, Ma. Jos Garciarena Ucelay, Ma. Paula Villegas, Sergio G. Burdisso, Leticia C. Cagnina, Manuel Montes-y Gmez, and Marcelo L. Errecalde. 2018. Unsls participation at erisk 2018 lab. *Proceedings of the 9th International Conference of the CLEF Association, CLEF 2018, Avignon, France*.

Sharath Chandra Guntuku, David B Yaden, Margaret L Kern, Lyle H Ungar, and Johannes C Eichstaedt. 2017. Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences*, pages 43–49.

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

ical lexicons. *2014 IEEE 11th Intl Conf on Ubiquitous Intelligence and Computing and 2014 IEEE 11th Intl Conf on Autonomic and Trusted Computing and 2014 IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops*, pages 844–849.

Ronald C Kessler, Evelyn J Bromet, Victoria Shahly Peter de Jonge, and Marsha. 2017. The burden of depressive illness. *Public Health Perspectives on Depressive Disorders*.

David E. Losada and Fabio Crestani. 2016. A test collection for research on depression and language use. *Proceedings of the 7th International Conference of the CLEF Association, CLEF 2016, Evora, Portugal*.

David E Losada, Fabio Crestani, and Javier Parapar. 2017. erisk 2017: Clef lab on early risk prediction on the internet: Experimental foundations. *Proceedings of the 8th International Conference of the CLEF Association, CLEF 2017, Dublin, Ireland*.

David E Losada, Fabio Crestani, and Javier Parapar. 2018. Overview of erisk 2018: Early risk prediction on the internet (extended lab overview). *Proceedings of the 9th International Conference of the CLEF Association, CLEF 2018, Avignon, France*.

Colin D Mathers and Dejan Loncar. 2006. Projections of global mortality and burden of disease from 2002 to 2030. *PLOS Medicine, Public Library of Science*, pages 1–20.

Saif M. Mohammad and Peter D. Turney. 2013. Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, pages 436–465.

John P. Pestian, Henry Nasrallah, Pawel Matykiewicz, Aurora Bennett, and Antoon A. Leenaars. 2010. Suicide note classification using natural language processing: A content analysislin heidelberg. *Biomed Inform Insights*, page BII.S4706.

Precha Thavikulwat. 2008. Affinity propagation: A clustering algorithm for computer-assisted business simulation and experimental exercises. *Developments in Business Simulation and Experiential Learning*.

Marcel Trotzek, Sven Koitka, and Christoph M. Friedrich. 2017. Linguistic metadata augmented classifiers at the clef 2017 task for early detection of depression. *Proceedings of the 8th International Conference of the CLEF Association, CLEF 2017, Dublin, Ireland*.

Marcel Trotzek, Sven Koitka, and Christoph M. Friedrich. 2018. Word embeddings and linguistic metadata at the clef 2018 tasks for early detection of depression and anorexia. *Proceedings of the 9th International Conference of the CLEF Association, CLEF 2018, Avignon, France*.

Ma. Paula Villegas, Dario G. Funez, Ma. Jos Gariarena Uelay, Letiia C. Cagnina, and Marcelo L. Errealde. 2017. Lidic - unsl's partiipation at erisk 2017: Pilot task on early detetion of depression notebook for erisk at clef 2017. *Proceedings of the 8th International Conference of the CLEF Association, CLEF 2017, Dublin, Ireland*.

C. Walck. 2007. Hand-book on statistical distributions for experimentalists. *University of Stockholm, Internal Report SUFPFY/9601*, pages 36–44.

Xinyu Wang, Chunhong Zhang, Yang Ji, Li Sun, Leijia Wu, and Zhana Bao. 2013. A depression detection model based on sentiment analysis in micro-blog social network. *Trends and Applications in Knowledge Discovery and Data Mining, Springer Berlin Heidelberg*, pages 201–213.

Yuanyuan Xue, Qi Li, Li Jin, Ling Feng, David A. Clifton, and Gari D. Clifford. 2013. Detecting adolescent psychological pressures from micro-blog. *Health Information Science, Springer International Publishing*, pages 83–94.