

INAOE-CIMAT at eRisk 2019: Detecting Signs of Anorexia using Fine-Grained Emotions

Mario Ezra Aragón¹, A. Pastor López-Monroy², and Manuel Montes-y-Gómez¹

¹ Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Mexico
`{mearagon,mmontesg}@inaoep.mx`

² Centro de Investigación en Matemáticas (CIMAT), Mexico
`pastor.lopez@cimat.mx`

Abstract. In this paper, we present our approach to the detection of anorexia at eRisk 2019. The main objective of this shared task is to identify as soon as possible if a user shows signs of anorexia by using their posts on Reddit. For this, we evaluate a representation called Bag of Sub-Emotions (BoSE), a new technique that represents user posts by building a set of fine-grained emotions. At the beginning, emotions are defined according to categories in a given lexical resource, then fine-grained emotions are discovered by clustering word vectors in each category. For our participation, we chose to evaluate different strategies based on the temporal stability that a user presents and perform early predictions using this representation. The proposed approach shows better performance than the average results of other participants; in addition, due to its interpretability and simplicity, it offers an excellent opportunity for the analysis and detection of mental disorders in social media.

Keywords: eRisk 2019 · Anorexia Detection · Bag of Sub-Emotions.

1 Introduction

Anorexia nervosa is an eating disorder that affects many adolescents and young adults these days. It is a desire to lose weight through excessive restriction of the number of calories and the types of food people eat. Anorexia is characterized by difficulties of maintaining appropriate body weight, and in many people presents a distorted body image. Anorexia can affect people of all ages and genders. The 2019 Early Risk Prediction on the Internet (eRisk@CLEF) shared task 1 has the objective of dealing with this problem by using Natural Language Processing (NLP) techniques and machine learning. The main goal is to identify if a user presents signs of anorexia as soon as possible, by processing their post history as pieces of evidence. Posts are processed in the order they were created, applying sequentially monitoring of the user's interactions in their social media platforms.

Copyright © 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CLEF 2019, 9-12 September 2019, Lugano, Switzerland.

In this work, we described the joint participation of INAOE-CIMAT, two research centers from Mexico, at this forum using a new representation that we have called Bag of Sub-Emotions (BoSE), an interpretable and straightforward approach, based on the usage of fine-grained emotions to capture specific emotions that the users present on their post. This representation is created by using a clustering algorithm over a lexical resource of emotions and then mask the post of the users to generate a histogram of these new emotions. We evaluate our representation using five different strategies for the early prediction.

The remainder of this paper is as follows: Section 2 presents some related work for the anorexia detection task and early predictions. Section 3 describes our new text representation based on fine-grained emotions. Section 4 and Section 5 presents the experimental settings as well as the obtained results. Lastly, Section 6 depicts our conclusions.

2 Related Work

In this Section, we present a review of the different works related to the detection of anorexia in social media. Anorexia is the most common Eating Disorder (ED) related to a mental disorder, and consists of an unusual habit of eating or abnormal attitudes towards food [13]. Several works in the literature have focused on analyzing user-generated content from their social media platforms to identify signs of anorexia. Some of these works have proposed to analyze the user posts to generate syntactic and semantic features [2–6], where they explore the words that are often used by people with anorexia signs. Another well-known strategy is the employment of words or dictionaries that are related to anorexia, and then create a representation by using the occurrence or frequency of such words [5]. Other examples that had been explored are the Deep Learning techniques, which also are getting competitive results [2, 6, 8]. Last but not least, a traditional type of strategy is to exploit sentiment analysis to create emotional characteristics to represent each user post [5, 7]; inspired in this last approach we explore the usefulness of a representation based on a set of automatically-learned fine-grained emotions, which help to model the emotional profile of users in a more specific way.

3 Representation

In this section, we describe the representation that was used to participate in the shared task. Our approach is inspired by the hypothesis that emotions are better represented at a finer level, instead of only using general concepts as "anger" or "joy".

Figure 1 illustrates the general steps of our proposed approach. The first part explains the generation of the fine-grained emotions given an emotion lexicon. The second part depicts a masking process used to have the fine-grained emotions as tokens, and then the creation of their histogram as final representation.

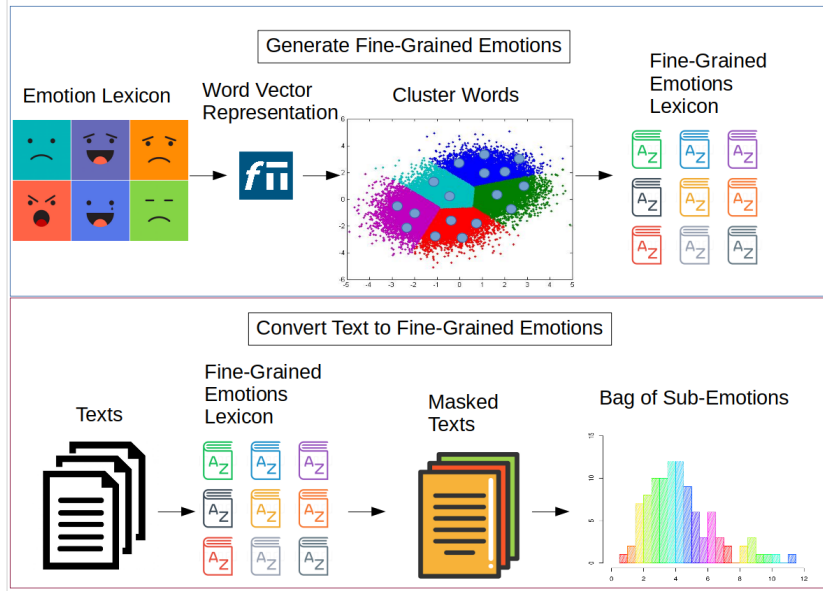


Fig. 1. This diagram represents the creation of our representation called Bag of Sub-Emotions (**BoSE**) [1].

Generate Fine-Grained Emotions: We use a lexical resource to compute a set of fine-grained emotions based on eight recognized main emotions (Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise and Trust) [12] and two main sentiments (Positive and Negative). In this stage, we compute a word vector from FastText for each word presented in the lexical resource. Then, we create subgroups of words separated by emotions employing the Affinity Propagation clustering algorithm and use their centroids (prototypes) as a new vocabulary for the fine-grained emotions.

Convert Text to Fine-Grained Emotions: Once we calculate the fine-grained emotions, we utilize them to mask the text by measuring the cosine distance between the words in the documents and the fine-grained emotions. Then, we represent the posts of the users creating a histogram of the frequencies of fine-grained emotions. We named this representation **BoSE**, for Bag of Sub-Emotions (see [1] for more details).

4 Experimental Settings

This shared task is a continuation of eRisk 2018 T2 task [10], which consists in detecting traces of anorexia in users of Reddit as soon as possible. The latter is done by sequentially processing the users' posts. This year, organizers modified the way the data is released, which was variable-chunk-length³ based in 2017

³ e.g., users that wrote more, would have more information per chunk

and 2018, but now is an item-by-item version. The latter means that a server iteratively provides user writings in chronological order, using a token identifier for each team. For each writing that the server offers, we need to respond with a prediction to continue with the next round of posts; otherwise, the server will be still waiting.

Our objective for the shared Task 1 is to decide if a user presents signs of anorexia applying every five posts a preprocessing and a classification procedure to make the labels for each user. Lastly, we used five different strategies to sent the predictions. We explained the whole process below.

Preprocessing: For the experiments, the posts are normalized by removing special characters and lowercasing all the words. After these processes, texts are masked using the fine-grained emotions previously computed.

Classification: Once we built the BoSE representation, we selected the most relevant features (sequences of fine-grained emotions) by using the χ^2 distribution X_k^2 [9]. To classify the users, we used a Support Vector Machine (SVM) with a linear kernel and $C = 1$.

Prediction making: For each post that the server provides, we need to make a prediction to tell if the user presents signs of anorexia or not, and the main idea is to make a correct detection as soon as possible. We tackled the task by using the following five strategies: *i*) we considered the label obtained directly from the classifier; *ii*) we used the probability of the label, assigned as positive if the chance is higher than 60% of belonging to that class; *iii*) similar to the first strategy, we considered the label obtained directly from the classifier, but only assigned the label 1 if the user is detected as positive in the previous prediction as well; *iv*) the user is classified as positive if the probability of the classifier is higher than 60% in the actual and previous predictions; *v*) similar to the fourth strategy but the classification probability needs to be higher than 70%.

5 Experimental Results

To determine the parameters for the model before the prediction in the server, we first evaluated our model with the dataset provided in 2018. For that corpus, there are two categories of users: with anorexia and control. We measured the F1 over the predictions using the whole post history of the users. In Table 1 we present the obtained results over the training dataset; we compare our approach with traditional representations like Bag of Words (BoW) using unigrams and n-grams and a representation based on the core emotions that we named Bag of Emotions (BoE).

For the test dataset, we trained our model using all the users in the training dataset and then we determined if the users show or not show traces of anorexia using the five different strategies mentioned in Section 4. Table 2 show the results obtained by the five strategies over the test dataset. Note that on these results: run1 did not work on the server, and we still do not know the reason for this, therefore, their results are not included in the table. The strategy that obtained

| Method | Unigrams | Ngrams |
|--------|-------------|-------------|
| BoW | 0.69 | 0.69 |
| BoE | 0.50 | 0.58 |
| BoSE | 0.82 | 0.81 |

Table 1. F1 results over the positive class against baseline methods in the Training Dataset

the best results is the fourth (marked as run3); it consists in classifying the user as positive if the probability is higher than 60% in the actual and previous prediction, which involves the temporal stability obtained by the classifier where we get two consecutive positives predictions over the user.

| Method | F1 | ERDE ₅ | ERDE ₅₀ | latency-weighted F1 |
|--------|-------------|-------------------|--------------------|---------------------|
| run 0 | 0.66 | 0.09 | 0.04 | 0.62 |
| run 2 | 0.66 | 0.09 | 0.09 | 0.50 |
| run 3 | 0.68 | 0.09 | 0.05 | 0.63 |
| run 4 | 0.66 | 0.09 | 0.05 | 0.61 |

Table 2. Results over the positive class in the Testing Dataset

To a further analysis of our results in the first part of Figure 2, we present a boxplot of all the results obtained for F1 measure and Latency-weighted F1, the green **X** mark represents the position of our results. We can appreciate that our results for both evaluation metrics are in the highest quartile, indicating the good results obtained for this task.

In the second part of Figure 2, we present the boxplots of the results of all participants in accordance to the ERDE5 and ERDE50 evaluation metrics. In these results, our approach is placed in the middle quartile. This performance was somehow expected since our approach does not focus on fast prediction, but more on the temporal stability of the predictions. [11] presents the overall results of the task as well as a complete analysis of every team approach.

The interpretability of our method allows us to offer more analysis of what is captured by the fine-grained emotions, we selected some of the most relevant fine-grained emotions for the detection according to the χ^2 distribution. In Table 3, we present some of these fine-grained emotions as well as some words that correspond to them. We can observe that most of the emotions are related to psychical or mental harms like bruising, breakdown, abandoned; or body parts near the stomach or intestine, which are topics that people commonly associated to anorexia.

6 Conclusions

In this paper, we present our approach to decide if a user presents signs of anorexia by using the post history in chronological order and make a predic-

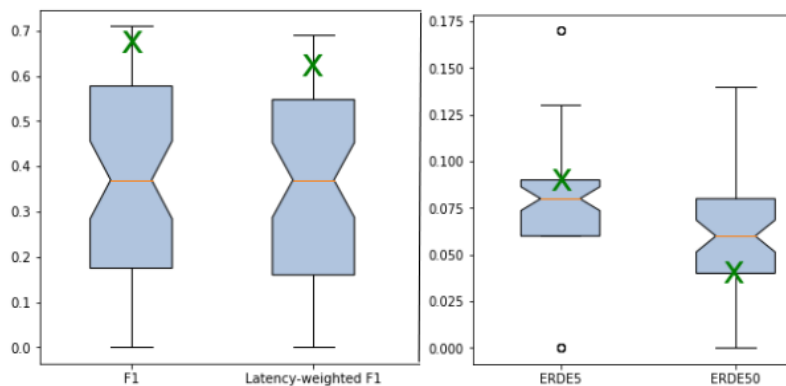


Fig. 2. Boxplot for the results in F1, Latency-weighted F1, ERDE5, and ERDE50, where the green X mark represents our obtained results.

Table 3. Examples of fine-grained emotions relevant for the detection of anorexia

| | |
|------------|---|
| anger4 | bruising, contusion, bleeding, fracture |
| disgust32 | breakdown, fight, crushed, abandoned |
| disgust21 | stomach, intestinal, bile, esophagus |
| negative65 | bathroom, toilet, washroom |
| anticip10 | hurting, refused, anxious, afraid |
| anticip12 | ashamed, embarrass, upset, disgust |
| fear19 | food, eating, eat, consume |

tion as soon as possible. We proposed a new representation that automatically creates fine-grained emotions using a lexical resource of emotions and FastText sub-word embeddings. The main idea of using these fine-grained emotions is that our representation can capture more specific emotions and topics that the users express through their posts and help to detect potential users that have anorexia. Over the training dataset, our representation obtained better results than most of the best previous eRisk participant’s methods. The simplicity and interpretability of our representation are worth mentioning, which differs with other methods that are more difficult and complex, in particular those that used a lot of different features and different models from traditional to deep. For the testing dataset, our representation also obtains good results in comparison with most of this year participants, proving evidence about the usefulness of capturing the specific emotional content of users that have anorexia. Our results represent an opportunity to use BoSE in other health tasks such as Depression or Post-Traumatic Stress Disorder (PTSD).

Acknowledgments

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

References

1. Aragón, ME., López-Monroy, AP., González-Gurrola, LC., Montes-y-Gómez, M.: Detecting Depression in Social Media using Fine-Grained Emotions. 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). (2019)
2. Trotzke, M., Koitka, S., Friedrich, CM.: 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. (2018)
3. Ramiandrisoa, F., Mothe, J., Farah, B., Moriceau, V.: IRIT at e-Risk 2018. Proceedings of the 9th International Conference of the CLEF Association, CLEF 2018, Avignon, France. (2018)
4. Ortega-Mendoza, RM., Lopez-Monroy, AP., Franco-Arcega, A., Montes-Y-Gómez, M.: PEIMEX at eRisk2018: Emphasizing Personal Information for Depression and Anorexia Detection. Proceedings of the 9th International Conference of the CLEF Association, CLEF 2018, Avignon, France. (2018)
5. Ramírez-Cifuentes, D., Freire, A.: UPF's Participation at the CLEF eRisk 2018: Early Risk Prediction on the Internet. Proceedings of the 9th International Conference of the CLEF Association, CLEF 2018, Avignon, France. (2018)
6. Liu, N., Zhou, Z., Xin, K., Ren, F.: TUA1 at eRisk 2018. Proceedings of the 9th International Conference of the CLEF Association, CLEF 2018, Avignon, France. (2018)
7. Ragheb, W., Moulahi, B., Aze, J., Bringay, S., Servajean, M.: Temporal Mood Variation: at the CLEF eRisk-2018 Tasks for Early Risk Detection on The Internet. Proceedings of the 9th International Conference of the CLEF Association, CLEF 2018, Avignon, France. (2018)
8. Wang, YT., Huang, HH., Chen, HH.: A Neural Network Approach to Early Risk Detection of Depression and Anorexia on Social Media Text. Proceedings of the 9th International Conference of the CLEF Association, CLEF 2018, Avignon, France. (2018)
9. Walck, C.: Hand-book on Statistical Distributions for experimentalists. University of Stockholm, Internal Report SUFPFY/9601. (2007)
10. Losada, DE., Crestani, F., Parapar, J.: 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. (2018)
11. Losada, DE., Crestani, F., Parapar, J.: Overview of eRisk 2019: Early Risk Prediction on the Internet. Experimental IR Meets Multilinguality, Multimodality, and Interaction. 10th International Conference of the CLEF Association, CLEF 2019, Lugano, Switzerland. (2019)
12. Mohammad, S.M., Turney, P.D.: Crowdsourcing a Word-Emotion Association Lexicon. Computational Intelligence. (2013)
13. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders. Fourth Edition. Washington, DC: American Psychiatric Press. (1994)