

Detecting Mental Disorders in Social Media Using a Multichannel Representation

Mario Ezra Aragón Saenzpardo¹,
Adrián Pastor López Monroy², Manuel Montes y Gómez¹

¹ Instituto Nacional de Astrofísica, Óptica y Electrónica,
Mexico

² Centro de Investigación en Matemáticas,
Mexico

`pastor.lopez@cimat.mx`, `{mearagon, mmontesg}@inaoep.mx`

Abstract. Common mental disorders such as depression, anorexia, dementia, post-traumatic stress disorder (PTSD) and schizophrenia affect millions of people around the world. In this work, to detect mental disorders in social media, we propose: 1) different representations from the information shared by the users. For example, topic information, phonetic or writing style, and emotion information. 2) A model that automatically creates a representation combining the previous representations. With these, the model can learn to represent social media documents (a.k.a. posts) by using the combination of these different types of information. The generated representations (individual and combined) will be evaluated in different tasks related to mental disorders, for example, depression detection, anorexia detection and post-traumatic stress disorder (PTSD). As preliminary results; we design a new representation considering emotions as information called Bag of Sub-Emotion(BoSE), which represents social media documents by a set of fine-grained emotions automatically generated using a lexical resource of emotions and sub-word embeddings. We evaluated this first representation in depression and anorexia detection. The results are encouraging; the usage of fine-grained emotions improved the results from traditional representations and a representation based on the core emotions and obtained competitive results in comparison to state of the art approaches. We also present results from a representation inspired by the emotional changes of a user, this representation combined with BoSE obtain better results than using them separately.

Keywords: Mental disorders, natural language processing, machine learning, deep learning.

1 Introduccion

Motivation: Currently, millions of people around the world are affected by different mental disorders that interfere in their thinking and behavior, damaging their daily life [1, 2]. Timely detection of mental disorders is important to help people before the illness gets worse, minimizing disabilities and returning them to their normal life. The stigma related to mental disorders creates barriers to improve the resources that help the detection of these problems. The most popular way for people to share information is using social media platforms, and people tend to share topics related to work issues and personal matters. People with mental disorders tend to share more about their concerns looking for some advice, support or just because they want to relieve suffering. This creates an excellent opportunity to automatically detect users that have a mental disorder and refer them as soon as possible to seek professional help.

Previous Works in the area: To accomplish the detection of mental disorders, automated analysis of social media is made using predictive models that uses features or variables that are extracted from the data posted by the users in their social media accounts. For example, one of the most commonly used features are word frequencies [6]. Other works used the LIWC dictionary [5], to represent the posts by means of a histogram of psycho-linguistic categories [7], or consider a representation based on the polarity information of the posts [8].

Hypothesis: People that present some mental disorder tend to express differently than healthy people; their topics of interests, writing style, relation with others and even their activity hours have different behavior. The hypothesis is that learning to combine different channels of information, could give a broader view that helps to detect signs of mental disorders and obtain better classification results that using single information.

Objective: Design a method applying traditional NLP techniques combined with deep learning techniques to automatically learn a multichannel representation using the information generated by the users in social media platforms. Then use this representation for the detection of mental disorders and improve the results obtained by traditional and state of the art approaches.

Specific Objectives:

1. Design methods that learn new representations of the different channels in the post history of the users: the context, the style of the author, the emotions used and phonetic information.
2. Design a model that automatically combines the different information channels and focuses on the critical parts of the data.
3. Develop a method to incorporate the importance of temporal information presented in the sequences of the posts.
4. Evaluate the utility of our proposed method in different tasks related to mental disorders.

2 Methodology

This section briefly describes the methodology to reach the proposed objectives.

1. **Identify and Obtaining datasets related to mental disorders.** We plan to obtain datasets like depression detection, anorexia detection, PTSD detection. For example, datasets from **CLEF eRisk** or **CLPSYCH**.¹
2. **Develop methods that extract information in different channels.** In this step, it is necessary the analysis of different kinds of information presented in the posts to extract and create separate channels. For example, could be the topics that are contained, the emotions presented or the style of the author for writing or expressing
3. **Develop a model to create a representation that combines the different channels automatically.** This step involves the development of a model that automatically combines the different channels obtained in the step before, and creates a new representation. To overcome this problem using models inspired in Deep Neural Networks that learn to combine and or give importance to a different type of information.
4. **Design an approach that effectively incorporates sequential information in the representation.** Due to the nature of the information that is created involving the sequencing of actions, where a user writes a post one after another. For example, deep learning models like Recurrent Neural Networks that take time and sequence into account.

3 State of the Research

This section presents the preliminary work that has been done until now that supports our hypothesis and research proposal.

1. Our first experimental approach consists of the usage of the emotions channel (part of the second step in the methodology); it is proposed a new representation called Bag of Sub-Emotions (BoSE). This channel represents social media documents using a set of fine-grained emotions that are automatically generated using lexical resources based on emotions and sub-word embeddings. To evaluate this representation, two different tasks were used: depression and anorexia detection. The results are promising; the usage of these fine-grained emotions improved the results from a representation based on traditional methods and based on the core emotions. The results obtained are also competitive in comparison to state of the art approaches (Table 1).
2. Temporal analysis for the emotion channel. A first exploration of the temporal information that is presented in the emotion channel. With this analysis we can appreciate the behaviour of the emotions through time of people with a mental disorder.

¹ <https://early.irlab.org/> and <http://clpsych.org/>

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

Method	Dep'17	Dep'18	Anor'18
BoW-unigrams	0.60	0.58	0.69
BoE-unigrams	0.57	0.60	0.50
BoSE-unigrams	0.61	0.61	0.82
BoW-ngrams	0.59	0.60	0.69
BoE-ngrams	0.61	0.58	0.58
BoSE-ngrams	0.64	0.63	0.81
delta-BoSE	-	0.53	0.79
Late Fusion	-	0.64	0.84

3. An early and late fusion of the temporal features with the original BoSE (part of the third step in the methodology). This approach obtains a little increase in the results that using the information separated.
4. An approach inspired in the modeling of fine-grained emotions expressed by the users and deep learning architectures with attention mechanisms for the detection of depression and anorexia.

4 Conclusions

In this document we present the research proposal during the Ph.D. program and part of the work that has been made. The main interest of this research is the detection of mental disorders in users through the post in different social media platforms. The work will focus on the detection of these users improving the state of the art results, using a new multichannel representation that exploits traditional natural language process methods combined with deep learning methods. The emotional channel present useful information that helps the detection of mental disorders. BoSE obtained better results than the proposed baselines and also improved the results of only using broad emotions. Incorporating temporal analysis over the emotion channel and combine it with the previous representation demonstrate that helps the detection of users that presents signs of mental disorders. It is worth mentioning the simplicity and interpretability of the representation, creates a more straightforward analysis of the results.

References

1. Kessler, R., Bromet, E., Jonge, P., Shahly, V., Marsha.: The burden of depressive illness. Public Health Perspectives on Depressive Disorders (2017)
2. Mathers, C., Loncar, D.: Projections of global mortality and burden of disease from 2002 to 2030. PLOS Medicine, Public Library of Science (2006)

3. Aragón, M.E., López-Monroy, A.P., González-Gurrola, L.C., 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, 1 (2019)
4. Aragón, M.E., López-Monroy, A.P., Montes-y-Gómez, M.: INAOE-CIMAT at eRisk 2019: Detecting Signs of Anorexia using Fine-Grained Emotions. Proceedings of the 10th International Conference of the CLEF Association, CLEF 2019, Lugano, Switzerland. (September, 2019).
5. Tausczik, YR., Pennebaker, JW.: The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology* (2010)
6. Tsugawa, S., Kikuchi, Y., Kishino, F., Nakajima, K., Itoh, Y., Ohsaki, H.: Recognizing depression from twitter activity. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (2015)
7. Schwartz, HA., Eichstaedt, J., Kern, M., Park, G., Sap, M., Stillwell, D., Kosinski, M., Ungar, L.: Towards assessing changes in degree of depression through facebook. In: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality (2014)
8. Coopersmith, G., Dredze, M., Harman, C.: Quantifying mental health signals in Twitter. In: Workshop on Computational Linguistics and Clinical Psychology (2014)