Eating Disorders Detection by means of Deep Learning*

Xabier Larrayoz[1,*,†], Nuria Lebeña[1], Arantza Casillas[1] and Alicia Pérez[1]

[1]HiTZ Center - Ixa, University of the Basque Country (UPV/EHU), 20080 Donostia, Spain

Abstract

Our approach for the IberLEF 2023 MentalRiskES Workshop Task 1 (Eating disorders detection) is presented in this paper. The objective of the task is to examine social media users' posts in search of early indicators of eating disorders. A challenge inherent to the task rests on the highly skewed data-set. In order to successfully address the issue of class imbalance, our approach calls for the use of a neural network that incorporates a unique loss function. Our method gives you the freedom to modify the penalty for false positives and false negatives in accordance with particular needs. Our method also presents a variety of potential modifications that need more research. Our approach is, however, preliminary, since the model employed was not adjusted for Spanish and, thus, there is room for improvement.

Keywords

Early risk prediction, Natural Language Processing, Class imbalance, Deep learning, Mental health

1. Introduction

Despite the fact that mental health is crucial to our lives, it is still stigmatized in our culture. Many patients and their families are left to suffer in silence because they are unable to get the help and understanding they need. Social media is extremely important when society isn't actually involved. More and more people view social media platforms as venues where they can express their problems and experiences, especially those pertaining to their mental health. It is crucial to remember that texts posted on social media might be a hint of possible problems because many people who suffer from mental disorders start exhibiting their signs and symptoms there.

However, given the volume of data produced each day, it is impractical to process this data the old-fashioned way. In this regard, the various editions of the MentalRiskES workshop serve as a meeting place where methodologies and useful techniques for early detection of various types of health risks, such as eating disorders, unidentified disorders, or depression, through the textual analysis of posts and messages from social media users, have been developed.

To fulfill Task 1 of the MentalRiskES Workshop 2023 [1]: Eating disorders detection, we describe our system in this work. The strategy relies on creating vector representations of user

IberLEF 2023, September 2023, Jaén, Spain

[*] Corresponding author.

[†] These authors contributed equally.

xlarrayoz001@ikasle.ehu.eus (X. Larrayoz); nuria.lebena@ehu.eus (N. Lebeña); arantza.casillas@ehu.eus (A. Casillas); alicia.perez@ehu.eus (A. Pérez)

CEUR Workshop Proceedings (CEUR-WS.org)

communications by sentence embedding first, and then utilizing deep learning-based techniques to identify encouraging remarks. In order to address unbalanced classes or situations where false negatives have a large influence, an original loss function is also introduced.

2. Related work

One in eight people worldwide suffer from a mental condition according to a research recently conducted by World Health Organization. Given that NLP has shown itself as a potent tool for the diagnosis of these problems, several competitions have recently begun to focus on applying machine learning techniques to detect various mental disorders. SMM4H [2], CLEF and CLPsych [3] are examples of competitions that focus on mental health in english. IberLEF [4], was started in 2019 as part of the annual conference of the Spanish Society for Natural Language Processing (SEPLN). It is a shared task for NLP systems in Spanish and other Iberian languages, this year it first includes MentalRiskES an innovative set of tasks on early risk detection of mental diseases using comments from Telegram users.

On previous years similar tasks have been addressed but not in the mental health field. Exist task consists on sexism identification and categorization of tweets both in Spanish and English. On 2022, the best team achieved a F1 score of 0.79 in sexism identification and 0.51 on sexism categorization [5]. They based their work on exploring different transformer-based solutions. For classifying Spanish tweets they explored different transformers structures fine tunned with Spanish data: BETO [6], BERTIN [7], MarIAbase [8] and RoberTUITO [9].

In Da-Vincis task, participants are asked to detect and classify tweets in Spanish that report violence. Also transformer based solutions got the best results: the best team proposed a Multi-Task Learning (MTL) approach based on BERT [10] transformer [11], the second best team also based their solution on BERT transformer [12].

The early detection of pathological gambling has been one of the tasks to be tackled in eRisk@CLEF for a number of years. It has many similarities with the MentalRiskES task as the aim is to detect the first signs of pathological gambling as early as possible (a binary classification problem labeling each user with 0 or 1) and provide an estimation of users risk level to be pathological gamblers. In the previous edition, the SINAI [13] group proposed a feed-forward neural network (FFNN) [14] model fed by a vector containing lexical features of the text.

Given that in similar Spanish task BERT based methods have proven to be efficient, we suggest a similar strategy where we feed the FFNN with vectors encoded using Sentence-BERT (SBERT) [15] and Dynamic Averaging Network (DAN) [16], a variation of the Universal Sentence encoder (USE) [17].

3. Materials and methods

The training set provided for this edition consists of 74 users suffering from anorexia or bulimia, and 101 control users. Each user has a series of messages published on a social network. The small size of the dataset stands out, which poses a challenge for any attempt to train a model. The test set officially employed to assess the challenge, is sent to the participants iteratively through a connection to a server. The total number of users is 150, of which 64 are users who suffer from anorexia or bulimia.

The proposed approach involves an encoder that transforms the plain text of the messages into a numerical vector, which can serve as the input for a feed-forward neural network (FFNN). Firstly, each message undergoes a pre-processing stage, which includes converting the text to lowercase, cleaning special characters, and removing stop-words. Subsequently, each message is passed through an encoder that generates the input for the FFNN. The FFNN then produces a real value, where positive values indicate that the model assigns a positive label to the initial post, while any other value results in assigning a negative label.

The utilized encoders include Dynamic Aggregation of Networks (DAN) and Sentence-BERT (SBERT). It is important to note that both models have been exclusively trained on an English corpus, which presents an initial limitation for their application to a task involving Spanish data. Nevertheless, the models have surpassed our initial expectations by delivering superior performance. We use the encoders in order to obtain the numeric representation of the posts that serve as input to the FFNN. DAN encoder generates vectors of size 512, while SBERT operates with vectors of size 384.

We needed to have the posts tagged for the training because gold-labels are provided at the user level and we trained our FFNN using posts. That is, during the training phase, it is necessary to compare the projected post-label confidence to an expected or desired confidence. The post-level confidence is not provided, which is the root cause of the problem. In this research, we investigated a user-based message labeling as the reference post-label confidence: consists of labeling each post with the user-label assigned to the author of the post. That is, all of a user's posts will be classified as positive if the subject was labeled as positive. Alternative reference assignment procedures might be used to address future tasks.

In our approach, we propose an original loss function designed to handle imbalanced data or situations where a false positive is preferred over a false negative. During the training of the neural network, we apply a cross-entropy-based loss function with a weighted system. For a given sequence of posts from the same user, we make predictions and calculate an initial loss value using cross-entropy. Additionally, considering all independent post-level predictions, we generate a user-level prediction by assigning a positive value if at least one post has a positive label. Subsequently, we compare this new user-level prediction with the actual user label and assign a weight based on the specific scenario. When both labels match, the weight is set to 1; however, different weights are assigned for discordant labels, enabling us to penalize false positives and false negatives differently. This newly assigned weight is then multiplied by the initial loss to obtain an updated value that contributes to the adjustment of the network parameters.

4. Results

Given the size of the dataset, preliminary tests were conducted using the provided training and evaluation sets. However, due to the limited number of users in the development set (dev-set), consisting of only five individuals, a comprehensive analysis of the sensitivity of different parameters could not be carried out. For the same reason, we didn't conduct a hyper-parameter optimization. The network was trained with a learning rate of $5x10-5$ together with 5 epochs. Drop-out technique with a value of 0.2 and AdamW optimizer was also applied. In accordance with the competition rules, three variations were allowed to be submitted, one of which utilized SBERT, while the remaining variants employed the DAN model. Additionally, the first two configurations used weight values of 4 and 2, respectively for false negatives and false positives, while the last configuration used a weight ratio of 2 and 2.

Run	**DAN vs SBERT**
0	SBERT
1	DAN
2	DAN

Table 1
Submitted Runs: Description of the configurations explored. The second column refers to the encoding strategy.

In Table 2, the results obtained in the initial section of Task 1 are presented. It is evident that the models employing the DAN architecture have yielded superior outcomes compared to the SBERT version, thereby demonstrating the efficacy of the generated vector representations. Recently, this particular system has exhibited remarkable performance in the eRisk 2023 competition, securing one of the top positions. Nevertheless, the limited size of the available training dataset has hindered the optimal learning of the model. Furthermore, the utilization of an encoder trained on English corpora, distinct from the Spanish texts presented in this competition, elucidates the performance disparity relative to the winning team. Notwithstanding these limitations, considering that the encoder has exclusively been trained on English texts, it has showcased exceptional performance.

Team	Run	Accuracy	Macro-P	Macro-R	Macro-F1
CIMAT-NLP-GTO	0	**0.967**	**0.964**	**0.969**	**0.966**
Xabi_EHU	1	0.733	0.746	0.747	0.733
Xabi_EHU	2	0.740	0.773	0.707	0.709
BaseLine - Roberta Base	2	0.700	0.783	0.736	0.694
Xabi_EHU	0	0.693	0.688	0.691	0.689

Table 2
Decision-based evaluation for Task 1. Our Team is denoted as Xabi_EHU.

5. Conclusions

Our involvement is concentrated on the identification of eating problems. The objective is to estimate the subject-level label (either control or suffer), given a sequence of messages, using the fewest number of messages possible. We search for robust models that could punish false negatives because the Suffer label is the minority group in the very tiny dataset and it is. As a result, we developed a particular loss function. We have also been unable to use ML-based architectures due to the data imbalance, as all efforts to generalize the data and reach acceptable performance have been ineffective. Overfitting was not a goal when we developed our method. The provided runs essentially used an FFNN with a DAN or SBERT encoder. Adding a user-defined loss function improved training. This technique required us to heuristically find the reference of the post-level label in order to train the system because the user-level label is estimated using a posterior-level label. The model produced superior outcomes for the majority class (control users), which was a difficulty throughout the development phase. The performance on the classification task is inferior because the model's development has been concentrated on the primary binary classification task. Of course, the suggested strategy can be made better. We are encouraged to keep experimenting with different approaches by making use of all the information at each stage and redefining the model in order to enhance the estimation of the user-level label. Other fundamental questions worth investigating include defining a reference for the posterior confidence level. In any case, the suggested strategy is flexible. Using the corresponding encoder, the same approach can be applied to literature in other languages and various mental diseases. The weights of the modified loss function will change based on the class balance, but the model may still be competitive with a standard loss function. In order to further enhance the system's performance, it is advisable to consider the utilization of an encoder specifically trained in the Spanish language. This would facilitate a superior adaptation to the linguistic characteristics and nuances of the dataset. Additionally, it is crucial to explore methodologies for augmenting the dataset size. By increasing the quantity of available data, a more comprehensive understanding of the task at hand can be attained, leading to better optimization of the system's configuration.

Acknowledgments

This work was elaborated within the framework of LOTU (TED2021-130398B-C22) funded by MCIN/AEI/10.13039/501100011033, European Comission (FEDER), and by the European Union "NextGenerationEU"/PRTR. Besides, this work was partially funded by the Spanish Ministry of Science and Innovation (DOTT-HEALTH/PAT-MED PID2019-106942RB-C31); by the Basque Government (IXA IT-1570-22, Predoctoral Grant PRE-2022-1-0069) and Ikasiker grants published in the 11/07/2022 BOPV; and by EXTEPA within Misiones Euskampus 2.0.

References

[1] A. M. Mármol-Romero, A. Moreno-Muñoz, F. M. Plaza-del-Arco, M. D. Molina-González, M. T. Martín-Valdivia, L. A. Ureña-López, A. Montejo-Ráez, Overview of MentalriskES at IberLEF 2023: Early Detection of Mental Disorders Risk in Spanish, Procesamiento del Lenguaje Natural 71 (2023).

[2] D. Weissenbacher, J. Banda, V. Davydova, D. Estrada Zavala, L. Gasco Sánchez, Y. Ge, Y. Guo, A. Klein, M. Krallinger, M. Leddin, A. Magge, R. Rodriguez-Esteban, A. Sarker, L. Schmidt, E. Tutubalina, G. Gonzalez-Hernandez, Overview of the seventh social media mining for health applications (#SMM4H) shared tasks at COLING 2022, in: Proceedings

of The Seventh Workshop on Social Media Mining for Health Applications, Workshop & Shared Task, Association for Computational Linguistics, Gyeongju, Republic of Korea, 2022, pp. 221–241. URL: https://aclanthology.org/2022.smm4h-1.54.

[3] A. Tsakalidis, J. Chim, I. Bilal, A. Zirikly, D. Atzil Slonim, F. Nanni, P. Resnik, M. Gaur, K. Roy, B. Inkster, J. Leintz, M. Liakata, Overview of the clpsych 2022 shared task: Capturing moments of change in longitudinal user posts, in: Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology, Association for Computational Linguistics, Seattle, USA, 2022, pp. 184–198. doi:10.18653/v1/2022.clpsych-1.16.

[4] J. Gonzalo, M. Montes-y Gómez, F. Rangel, Overview of iberlef 2022: Natural language processing challenges for spanish and other iberian languages (2022).

[5] A. Vaca-Serrano, Detecting and classifying sexism by ensembling transformers models, language 2 (2022) 1.

[6] J. Cañete, G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, J. Pérez, Spanish pre-trained bert model and evaluation data, Pml4dc at iclr 2020 (2020) 1–10.

[7] J. de la Rosa, E. G. Ponferrada, P. Villegas, P. G. de Prado Salas, M. Romero, M. Grandury, Bertin: Efficient pre-training of a spanish language model using perplexity sampling, 2022. arXiv:2207.06814.

[8] A. G. Agirre, M. V. Montserrat, A. G. Fandiño, J. A. Estapé, M. Pàmies, J. L. Palao, J. S. Ocampo, C. P. Carrino, C. A. Oller, C. R. Penagos, Maria:: Modelos del lenguaje en español, Procesamiento del lenguaje natural (2022) 39–60.

[9] J. M. Pérez, D. A. Furman, L. A. Alemany, F. Luque, Robertuito: a pre-trained language model for social media text in spanish, CoRR abs/2111.09453 (2021). URL: https://arxiv.org/abs/2111.09453. arXiv:2111.09453.

[10] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of naacL-HLT, volume 1, 2019, p. 2.

[11] D. Vallejo-Aldana, A. P. López-Monroy, E. Villatoro-Tello, Leveraging events sub-categories for violent-events detection in social media, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022), CEUR Workshop Proceedings. CEUR-WS. org, 2022.

[12] P. Turón, N. Perez, A. Garcıa-Pablos, E. Zotova, M. Cuadros, Vicomtech at da-vincis: Detection of aggressive and violent incidents from social media in spanish, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022), CEUR Workshop Proceedings. CEUR-WS. org, 2022.

[13] A. M. Mármol-Romero, S. M. Jiménez-Zafra, F. M. Plaza-Del-Arco, M. D. Molina-González, M.-T. Martín-Valdivia, A. Montejo-Ráez, Sinai at erisk@clef 2022: Approaching early detection of gambling and eating disorders with natural language processing, in: CEUR Workshop Proceedings, volume 3180, CEUR-WS, 2022, pp. 961–971.

[14] G. Bebis, M. Georgiopoulos, Feed-forward neural networks, IEEE Potentials 13 (1994) 27–31. doi:10.1109/45.329294.

[15] N. Reimers, I. Gurevych, Sentence-bert: Sentence embeddings using siamese bert-networks, 2019. arXiv:1908.10084.

[16] M. Iyyer, V. Manjunatha, J. Boyd-Graber, H. Daumé III, Deep unordered composition rivals syntactic methods for text classification, in: Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on

natural language processing (volume 1: Long papers), 2015, pp. 1681–1691.

[17] D. Cer, Y. Yang, S. yi Kong, N. Hua, N. Limtiaco, R. S. John, N. Constant, M. Guajardo-Cespedes, S. Yuan, C. Tar, Y.-H. Sung, B. Strope, R. Kurzweil, Universal sentence encoder, 2018. arXiv:1803.11175.