Detecting signs of Depression from Social Media: Examining the use of summarization methods as data augmentation for text classification

Filip Nilsson[1] and György Kovács[1]

***Abstract*— Depression is a common mental disorder that severely affects the quality of life, and can lead to suicide. When diagnosed in time, however, mild, moderate, and even severe depression can be treated. This is why it is vital to detect signs of depression in time. One possibility for this is the use of text classification models on social media posts. Transformers have achieved state-of-the-art performance on a variety of similar text classification tasks. One drawback, however, is that when the dataset is imbalanced, the performance of these models may be negatively affected. Because of this, in this paper, we examine the effect of balancing a depression detection dataset using data augmentation. In particular, we use abstractive summarization techniques for data augmentation. We examine the effect of this method on the LT-EDI-ACL2022 task. Our results show that when increasing the multiplicity of the minority classes to the right degree, this data augmentation method can in fact improve classification scores on the task.**

I. INTRODUCTION

The number of people suffering from depression has been steadily increasing since the 1990s [6], and depression is the leading cause of suicide-related deaths worldwide [1]. Because of this increase the importance of automatically detecting depression also increased. Over the past few years, transformers have taken over the field of natural language processing (NLP) and achieved state-of-the-art results on various problems [9]. Google uses BERT [2] at the core of its search feature to comprehend different web pages and to summarize web-pages to present to their users[1]. Yet the problem that remains unsolved is how to fine-tune transformers to perform well on domains with imbalanced data. Previous methods have studied using transformers to generate entierly new training data [8].

When fine-tuning a transformer to achieve good results on a specific domain it is important that there is sufficient annotated data for the transformer to learn in-domain knowledge. Another question is the class distribution in our dataset. In our case, for example, two classes are severely underrepresented. To alleviate this problem, we use an abstractive summarization technique to balance the dataset.

II. DATA

The dataset was provided by the organizers of LT-EDI-ACL2022 [3] and it contains social media posts from different users, categorized as severe depression, moderate depression and not depression. The categories have been annotated using [5] and [4]. These posts differ greatly from BERTs training data, moreover, the dataset is not very large and is highly imbalanced. Therefore a good dataset to study the effect of data augmentation.

We were provided a training partition, validation partition and a test partition. Labels for the test partition however, have not been published yet. Hence this paper is mostly based on the first two partitions.

Class	Training	Validation
Severe	901	360
Moderate	6019	2306
Not Depression	1971	1830
All	8891	4496

TABLE I

THE NUMBER OF LABELS FOR THE TWO PARTITIONS. [3]

III. METHODOLOGY

In this section we describe the pipeline we applied for the task. To ensure repeatability, our code is shared in a Github[2] repository.

A. *Preprocessing*

Minimal preprocessing was done on the dataset, URLs were removed and we used Huggingface base pre-trained BERT tokenizer trained on WordPiece.

B. *Summarization*

In NLP two major approaches have evolved for summarization. Namely, extractive and abstractive summarization. Extractive summarization is when the model has to select the sentences present in the input that best summarize the text. Whereas in abstractive summarization the model has to generate the summarization by itself.

In our work we use Google's T5 [7] a text-to-text transformer pre-trained on the c4 dataset then fine-tuned for abstractive summarization.

We chose a specific length by which t5-base should generate its summarizations. These lengths were between 10 and 50 tokens and went through all the training examples in the underrepresented classes. These summarizations were randomly sampled when balancing that class for the training partition.

As seen in Table II the summarizations that T5 produces do not always include the most important part of the sentiment in the social media post. In Table II you can see that in the 30 token version the summary includes suicide attempts

[1]F. Nilsson and G. Kovács is with EISLAB Machine Learning, Luleå University of Technology, 977 54 Luleå, Sweden

[1]https://blog.google/products/search/search-language-understanding-bert/

[2]https://github.com/flippe3/depression_detection_augmentation

which can be vital for the classification while the 10 token version misses that.

Type	Text
Original	I'm really struggling : So I don't know how to start things like this, So I'll start with basics. I'm 16yo, diagnosed depression at 14yo. Since then, my life is total mess. I've already been to two different psychologists, both of them said that i'm kind of unfixable, but that might be due to my young age.
30 token	i'm 16yo, diagnosed depression at 14yo. since start of 2019, my life got even worse. i've had 6 suicide attempts, all of them ended up with getting yelled at
10 token	i'm 16yo, diagnosed depression at 14yo. i'm

TABLE II

EXAMPLES OF SUMMARIZATIONS FROM THE SEVERE CLASS, THE ORIGINAL ONE IS 314 WORDS AND WAS TRUNCATED.

Class	0	25	50	75	100
Severe	901	1505	3010	4515	6019
Moderate	6019	6019	6019	6019	6019
Not Depression	1971	1971	3010	4515	6019

TABLE III

THE DIFFERENT DEGREES OF BALANCING.

The data augmentation was done in five different degrees of how balanced the underrepresented classes were. This balancing can be seen in Table III.

IV. RESULTS

To evaluate the proposed data augmentation method, we applied it on the multi class classification task of the LT-EDI-ACL2022 challenge. Our results on the validation partition are outlined in Table IV. Although augmenting the data to the point of a completely balanced dataset improves the recall, it is at the cost of a lower precision. However, when selecting the degree of balancing carefully, one can improve the recall without a significant negative effect on precision. We did not observe the same improvement on the test partition (see Table V). Many factors can be behind this, including the test partition potentially having a markedly different distribution from the train and validation partitions. Unfortunately, due to the unavailability of test labels, we could not confirm this, and we also could not carry out further experiments on the test partition to more extensively examine the effect of our data augmentation.

Score	0	25	50	75	100
Macro F1	0.50	0.52	0.52	0.52	0.49
Macro Recall	0.50	0.52	0.54	0.51	0.52
Macro Precision	0.57	0.57	0.57	0.55	0.56

TABLE IV

RESULTS FROM FIVE DIFFERENT DEGREES OF DATA AUGMENTATION ON THE VALIDATION PARTITION.

Score	0	50	100
Macro F1	0.51	0.45	0.48
Macro Recall	0.53	0.51	0.54
Macro Precision	0.49	0.44	0.47

TABLE V

RESULTS FROM THREE DIFFERENT DEGREES OF DATA AUGMENTATION ON THE TEST PARTITION.

V. CONCLUSIONS AND FUTURE WORK

We examined a method of using an abstractive summarization model, T5 to do data augmentation. This was done before fine-tuning a BERT transformer on the dataset which was balanced to different degrees. We found that with the right degree of augmentation, the proposed method improved the performance of the BERT model on the task of detecting signs of depression on the validation partition, however this result was not evident in test data distribution. Future work for data augmentation for fine-tuning transformers could be done by comparing the result using an extractive summarization method. Our method was examined on one dataset, future work should use this technique of data augmentation to fine-tune for different domains on multiple datasets.

REFERENCES

[1] Silke Bachmann. Epidemiology of suicide and the psychiatric perspective. *Int. J. Environ. Res. Public Health*, 15(7), July 2018.

[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North*, 2019.

[3] Thenmozhi "Durairaj, Bharathi Raja Chakravarthi, Jerin Mahibha C, and Kayalvizhi" Sampath. Findings of the shared task on Detecting Signs of Depression from Social Media. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*. Association for Computational Linguistics, May 2022.

[4] S Evans-Lacko, S Aguilar-Gaxiola, A Al-Hamzawi, J Alonso, C Benjet, R Bruffaerts, W T Chiu, S Florescu, G de Girolamo, O Gureje, J M Haro, Y He, C Hu, E G Karam, N Kawakami, S Lee, C Lund, V Kovess-Masfety, D Levinson, F Navarro-Mateu, B E Pennell, N A Sampson, K M Scott, H Tachimori, M Ten Have, M C Viana, D R Williams, B J Wojtyniak, Z Zarkov, R C Kessler, S Chatterji, and G Thornicroft. Socio-economic variations in the mental health treatment gap for people with anxiety, mood, and substance use disorders: results from the WHO world mental health (WMH) surveys. *Psychol. Med.*, 48(9):1560–1571, July 2018.

[5] David Losada, Fabio Crestani, and Javier Parapar. erisk 2017: Clef lab on early risk prediction on the internet: Experimental foundations. pages 346–360, 08 2017.

[6] Institute of Health Metrics and Evaluation. Gbd results tool. `http://ghdx.healthdata.org/gbd-results-tool?params=gbd-api-2019-permalink\%2Fd780dffbe8a381b25e1416884959e88b`.

[7] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21(1), jan 2020.

[8] Sana Sabah Sabry, Tosin Adewumi, Nosheen Abid, György Kovacs, Foteini Liwicki, and Marcus Liwicki. Hat5: Hate language identification using text-to-text transfer transformer, 2022.

[9] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online, October 2020. Association for Computational Linguistics.