Anxiety and Depression Detection using NLP – Group 24

Alexander Korman, Netanel Sadeghi, Bailey Cislowski, Tom Yang, Yuan-Jung Chang

Abstract

Conventional machine learning struggles to detect concealed anxiety and depression cues due to societal stigma and avoidance behaviors. This creates a significant challenge as existing models rely on explicit data. Bridging this gap is crucial. Innovative methods are required to identify hidden mental health indicators effectively. Furthermore, we are interested in seeing the extensibility of models capable of predicting depression to the task of identifying anxiety, as these illnesses tend to be comorbid.

1 Introduction

1.1 Motivation

The realm of Natural Language Processing (NLP) has witnessed a surge in its application to mental health, particularly in detecting depression and anxiety. While traditional machine learning has been effective in various contexts, it encounters notable limitations in identifying subtle, concealed cues of mental health conditions like anxiety, especially when individuals, due to stigma or societal pressures, refrain from openly discussing their mental state. These hidden forms of anxiety are challenging to detect as they don't present explicit data signals typically relied upon by existing models.

1.2 Previous Research

This project is inspired by the study "Detecting Linguistic Traces of Depression in Topic-Restricted Text: Attending to Self-Stigmatized Depression with NLP" by Wolohan et al. (2018). Their research highlighted the challenges in detecting depression-related cues in language, especially when individuals avoid discussing their condition directly due to the stigma associated with mental health. They constructed a novel corpus from Reddit users and demonstrated significant differences in language used by depressed individuals when depression-related posts were withheld, suggesting that machine learning algorithms could still detect depression even in the absence of explicit discussions.

1.3 Our Project

Our project aims to extend this research by applying NLP techniques to not only detect hidden depression but also to identify signs of anxiety. Anxiety, often comorbid with depression, presents a unique challenge in mental health detection. Our project utilizes advanced NLP tools, such as sentiment analysis and emotion classification models, to identify subtle indicators of anxiety and depression without relying solely on explicit terms.

We have adopted a methodology similar to the one used by Wolohan et al. by collecting data from mental health-related subreddits and partitioning it into manageable segments for efficient machine learning applications. The aim is to explore whether a model trained on depression data can identify anxiety data, given their semantic relationship and commonality in expression.

One of the key contributions of our project is the exploration of extending models capable of predicting depression to the task of identifying anxiety. This involves addressing the challenges of detecting hidden cues in text data and identifying effective data sources outside traditional mental health forums to test the model's efficacy in distinguishing mental health-related text.

In summary, this project represents an important step in the application of NLP for mental health detection. By focusing on the nuanced detection of anxiety, in addition to depression, we aim to contribute to the development of more effective diagnostic tools and monitoring systems, providing valuable insights for healthcare professionals and researchers.

2 Methods

In our comprehensive study, we undertook a detailed exploration of various machine learning models to assess their capability in accurately identifying signs of depression and other negative emotions, such as anxiety, from textual data. Our initial approach focused on foundational models, where we employed the Support Vector Machine (SVM) and Perceptron algorithms. These models, known for their robustness in classification tasks, served as our baseline for comparison. To enhance our analysis and leverage recent advancements in natural language processing, we incorporated the more sophisticated Bidirectional Encoder Representations from Transformers (BERT) model. This state-of-the-art model, renowned for its deep learning capabilities and ability to understand the nuances of human language, offered a more complex and nuanced approach to our analysis.

2.1 Architecture & Training Procedure

For the Support Vector Machine (SVM), we opted for a linear kernel with an L1 penalized Support Vector Classifier. The optimal L1 penalty parameters were identified using KFold cross-validation. The final training parameters were set as follows: batch_size=16, K=5, C=64, iteration=1000, with the rest at default settings.

In the case of the Perceptron, given the binary nature of our classification task, a single-layer perceptron was deemed sufficient. The chosen parameters for this model were: batch_size=16, tolerance=0.001, iteration=1000, with other settings at default. Both methods employed TF-IDF for feature extraction.

For BERT, considering our limited data, we utilized a pre-trained transformer model from HuggingFace to enhance our outcomes. We employed the AutoTokenizer API for tokenizing our dataset, which automatically selects the suitable tokenizer based on the HuggingFace checkpoint. We used a multi-label classification variant of the BERT pre-trained model, which, despite our focus on binary classification (negative feelings or not), offers potential for future applications in distinguishing various negative emotions. The final parameters for this model were: batch_size=16, learning rate=0.00002, epoch=3, weight decay=0.01.

2.2 Main Achievement/Novelty

Our results clearly demonstrate the capability of these models to detect depression from text, as shown in Table 1. Remarkably, they could also identify other negative emotions like anxiety, even without training specifically on anxiety data, as seen in Table 2. This indicates that there are underlying connections between different negative texts that our models can detect. Additionally, our findings suggest that advanced models like transformers yield better results, and that transfer learning is effective in detecting negative emotions from textual data.

3 Experiments

3.1 Datasets and Pre-processing

For our experiment we collected data from reddit posts using reddit api. As our experiment focused on anxiety and depression detection, the majority of our data came from posts found in the r/Depression and r/Anxiety subreddits. To further generalize our training data we also collected posts from r/nba, r/funny, r/gaming, r/NoStupidQuestions.

We have partitioned the raw Reddit data into more manageable segments. This step involved breaking down large blocks of text into smaller units, such as sentences or paragraphs, thereby making it more amenable to machine learning techniques. It also nearly doubles the amount of data available to us. This pre-processing is crucial for efficient feature extraction and will ultimately lead to more accurate results.

To create our training data we created a dataset made up of a mix of r/Depression posts and a collection of posts from the various non depression or anxiety subreddits. We also created two testing datasets. The first of these was meant to predict depression and in doing so included a mix of r/Depression and unrelated posts. The second training dataset was a mix of r/Anxiety and unrelated posts in order to predict anxiety.

3.2 Models

We initially focused on leveraging two fundamental machine learning models: the Support Vector Machine (SVM) and the Perceptron. These models laid the groundwork for our exploration into more complex algorithms. Building upon this foundation, we further enhanced our approach by integrating advanced techniques from the realm of natural language processing, specifically through the employment of Bidirectional Encoder Representations from Transformers (BERT).
To harness the full potential of BERT, we utilized a pre-trained model available from the Hugging Face library, a renowned repository for state-of-the-art machine learning models. The specific variant we adopted was the "bert-finetuned-sem_eval-english" model. This model represents a refined adaptation of the original "bert_base_uncased" architecture, tailored to excel in semantic evaluation tasks in the English language. For training we used a batch size of 16, a learning rate of 2e-5, 5 epochs, and a weight decay of 0.01 as our hyperparameters.

4 Results & Discussions

Our project has embarked on a comprehensive journey in the realm of data science, beginning with the meticulous task of data collection and preparation. We have now moved into a pivotal phase, characterized by the implementation of cutting-edge technologies, specifically harnessing the capabilities of Hugging Face's libraries. The cornerstone of this phase is the integration of a sophisticated, pre-trained machine learning model sourced from Hugging Face. This model stands out due to its ability to streamline and expedite the process of data classification. The utilization of this pre-existing model is not merely a matter of convenience; it brings with it a substantial benefit in terms of time efficiency. Moreover, it imbues our analytical procedures with a heightened level of dependability. This is largely attributed to the model's prior training on a vast and varied dataset, which equips it with a robust foundation for accurate analysis.

We have adopted a two-pronged approach by implementing a pair of baseline models: the support vector machine and the perceptron. These models have been trained using a dataset comprising 2,411 instances that blend cases of depression with assorted data. Subsequently, our testing procedures were carried out on a separate set containing 1,205 instances. In an effort to optimize the data for these models, we transformed it into a format known as TF-IDF vectors. This conversion was crucial for enhancing the models' capacity to process and interpret the data effectively. Following this, we embarked on a thorough analysis of the models' performance based on the test set data. This evaluation was further extended to data pertaining to anxiety, providing a broader scope for assessment.

Additionally, we implemented a BERT pre-trained model from Hugging Face. This is a model dedicated to helping people who don't have a lot of data to pre-train a model. This makes it better for more generalized tasks, such as our task where we want to train on depression data and detect anxiety snippets of text.

The results derived from these performance metrics were illuminating. They indicated that a model initially trained on data related to depression possesses the capability to identify and interpret data associated with anxiety, even though it had not been exposed to such data during its training phase. This finding is particularly intriguing as it suggests a potential semantic link in the expression patterns of individuals suffering from depression and anxiety. Such a discovery could have profound implications in understanding the underlying commonalities in these mental health conditions. Below you may find the tables detailing the exact results of our models when trained and evaluated.

Model	**Precision**	**Recall**	**F1**
SVM	95.75%	93.01%	94.36%
Perceptron	96.98%	96.83%	96.90%
BERT	96.27%	96.35%	96.31%

(Table 1: Model trained on depression data, predicting a mix of depression and random data)

Model	Precision	Recall	F1
SVM	100%	62.83%	77.17%
Perceptron	100%	73.34%	84.62%
BERT	90.49%	90.64%	90.56%

(Table 2: Model trained on depression data, predicting anxiety data)

The tables above give our results for the all the models we attempted to implement. Since the original depression paper did not publish its dataset it is hard to draw exact comparisons between our models. However we used similar tactics to create a dataset similar to that of the paper. With that said our F1 score was higher for all of our models compared to the model from the depression paper (F1-Score: 0.729). We are comparing this to the task 1 results in the paper as we did not conduct tests with the depression topic withheld.

To encapsulate our progress thus far, we have successfully navigated through several significant milestones. These include the acquisition of targeted data from specific channels on Reddit, the thorough pre-processing of this data, and the initial forays into the application of advanced machine learning algorithms. The use of Hugging Face's pre-trained models has been a pivotal aspect of our strategy for data classification, marking a crucial step forward in our research endeavors.

5 Ethical Considerations

During our time developing this project we took into consideration the many reasons why someone would choose not to disclose feelings of anxiety or depression. With that said we hope any application of our research always abides by the classical principle of non-maleficence in bioethics: do no harm.

6 Conclusions

In this project, we have successfully redeveloped a system for detecting depression using various natural language processing (NLP) methodologies. We've expanded our research to include the identification of other similar negative emotions, like anxiety, leveraging the capabilities of NLP. As we look to the future, our focus is shifting towards the realm of multi-label classification.

An exciting aspect of this next phase involves broadening our data collection methods. Instead of limiting ourselves to specific types of posts from certain subreddits on Reddit, we plan to engage a group of individuals to annotate a wide range of emotional states. This strategy will not only diversify our dataset but also allow us to include a broader array of the emotional labels that interest us. Moving beyond the constraints of specific post types, we can significantly expand our dataset, enhancing its size and the range of emotions it covers.

Adopting this approach opens the door to not just detecting negative emotions but also to precisely identifying the specific types of feelings present in a given text segment. This refined capability in emotion detection holds tremendous promise for mental health support and intervention. The system we envision will be more nuanced and sensitive to the varied emotional states individuals experience, providing a richer understanding of the emotional content in texts.

Such advancements in NLP technology could revolutionize how we approach mental health in digital spaces, offering more tailored and effective support. By recognizing and categorizing a spectrum of emotions, we can better understand the complexities of mental states conveyed through language. This has the potential to enhance mental health interventions, offering more personalized and responsive support based on a deeper understanding of an individual's emotional landscape as reflected in their language use.

Overall, this project stands at the forefront of merging technological innovation with empathetic understanding, aiming to create tools that not only understand human emotion in a more sophisticated way but also contribute meaningfully to mental health care and support systems.

References

JT Wolohan, Misato Hiraga, Atreyee Mukherjee, Zeeshan Ali Sayyed, and Matthew Millard. 2018. Detecting Linguistic Traces of Depression in Topic-Restricted Text: Attending to Self-Stigmatized Depression with NLP. In Proceedings of the First International Workshop on Language Cognition and Computational Models, pages 11–21, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Xinchen Yu, Eduardo Blanco, and Lingzi Hong. 2022. Hate Speech and Counter Speech Detection: Conversational Context Does Matter. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 5918–5930, Seattle, United States. Association for Computational Linguistics.

Glen Coppersmith, Mark Dredze, and Craig Harman. 2014. Quantifying mental health signals in twitter. In Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, pages 51–60.

Munmun De Choudhury and Sushovan De. 2014. s: Self-disclosure, social support, and anonymity. In ICWSM.

Haque, A., Reddi, V., and Giallanza, T. (2021). Deep learning for suicide and depression identification with unsupervised label correction. arXiv preprint .arXiv:2102.09427.