Inferring Panic Attacks with Language Models

Dean Ninalga
justin.ninalga@mail.utoronto.ca

1 Key findings

The embeddings of certain language models appear to have developed a strong understanding of behaviors associated with certain mental disorders. Namely, taking the similarity between the embeddings from the NVDRS free text fields and PA is shown to significantly correlate with the presence of Anxiety disorder, Depression/dysthymia, and Bipolar disorder. Here, we will advance the recent call for a new DSM-5 PA transdiagnostic specifier to account for individuals who suffer from PA without a panic disorder diagnosis. We show that PA can be used as a proxy for *unknown* disorders, where the victim was noted as being treated for a mental health problem, but the actual diagnosis was not able to be determined.

2 Methodology

2.1 Motivation

PA often occurs across mental disorders and can cause severe outcomes with or without a panic disorder diagnosis. Roy-Byrne et al. (2000) found 21.6% of individuals who experience PA throughout their lifetime (independent of any other diagnosis) have a co-occurrence of major depression. In the same study, they found only 11.2% individuals with panic disorder have a co-occurrence of major depression. Moreover, nearly half of individuals with panic disorder likely have hypomanic or manic symptoms and other symptoms related to Bipolar disorder (Oh et al., 2021).

Recently Wideburg et al. (2024) advocated for the use of Panic Attacks (PA) / Fearful Spells (FS) as a distinct specifier in the DSM-5 (The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition). Here, we suggest PA as a variable that could advance youth mental health research and, strengthen the arguments presented in Wideburg et al. (2024).

2.2 Framework Summary

In broad terms, we generate a list of similarities between language model embeddings of the free text fields and the embedding of the prompt: "*panic attack behavior*". Then we compare the list to the presence of illness diagnosis found in*CME/LE_MentalHealthDiagnosis1/2* and present the most significant correlations. Notably, we use the descriptions found in *CME/LE_CircumstancesOtherText* since they encode information not captured by existing variables that may have contributed to the death.

2.3 Model Choice

We use `all-MiniLM-L6-v2` (Wang et al., 2020) an efficient language model performant on clustering and semantic search tasks without further finetuning.

2.4 Correlation Measure

The Point-biserial correlation coefficient r_{pb} is used to evaluate the similarity between the binary presence of the disorders and embedding similarities. For a binary variable Y and a continuous variable X the covariance can be measured as

$$r_{pb} = \sqrt{q_1 q_0}\frac{\mu_1 - \mu_0}{\sigma}$$

Note $\sigma = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(X_i - \bar{X})}$, μ_j is the mean of the X_i where $Y_i = j$ and, q_j is the proportion of $Y_i = j$ for $j = 0, 1$ At its core, r_{pb} measures the scaled distance between the population means $(\mu_1 - \mu_0)$ of the groups derived from Y.

2.5 Results

In Table 1 we show that the embedding similarities between the prompt (i.e. *panic attack behavior*) and circumstance descriptions from victims are shown to be correlated to the presence of coded disorders. We use these results as a baseline to

Table 1: Results (n=251) comparing the similarities produced by the CME-CircumstancesOtherText field with the occurrence of diagnoses found in CME-MentalHealthDiagnosis1. Correlations of diagnoses that are not included are not statistically significant (i.e. p-val < 0.9).

Disorder	q_1	r_{pb}	**p-val**
Depression/dysthymia	0.64	0.16	0.01
Anxiety disorder	0.03	0.20	<0.01
Bipolar disorder	0.06	0.11	0.09

Table 2: Results (n=73) comparing the similarities produced by the CME-CircumstancesOtherText field with the occurrence of diagnoses found in CME-MentalHealthDiagnosis1

Disorder	q_1	r_{pb}	**p-val**
Unknown	0.29	0.22	0.05
Anxiety disorder	0.23	0.30	0.01

validate our methodology as being capable of recovering/aligning supported connections between PA and other disorders. Here, we only include the statistically significant correlates due to space limitations. Table 2 shows the only significant correlation between the prompt and the 'Unknown' diagnosis code. Where this code is used for victims being treated for a mental health problem, but the actual diagnosis is not named in the documentation. These results may suggest that many of these cases are PA or anxiety disorder-related.

3 Code Snippets & Visualizations

```python
from sentence_transformers import SentenceTransformer


# 1. Load a pretrained model
model = SentenceTransformer("google/gemma-2-2b")

# 2. Feature extract
prompt = "panic attack"
prompt_embedding = model.encode([prompt])
doc_embeddings = model.encode(docs)

# 2. Cosine Similarity
similarities = model.similarity(doc_embeddings, prompt_embedding)
```

Figure 1: Generation of the document embedding and similarity generation pipeline using SentenceTransformers (Reimers and Gurevych, 2020)

```python
from scipy.stats import pointbiserialr


corr_obj = pointbiserialr(similarities[:, i], df[labels[j]])

corr_obj.correlation, corr_obj.pvalue # 2-tailed p-value
```

Figure 2: We use `scipy.stats.pointbiserialr` to compute the point biserial correlation coefficient and the associated p-value.

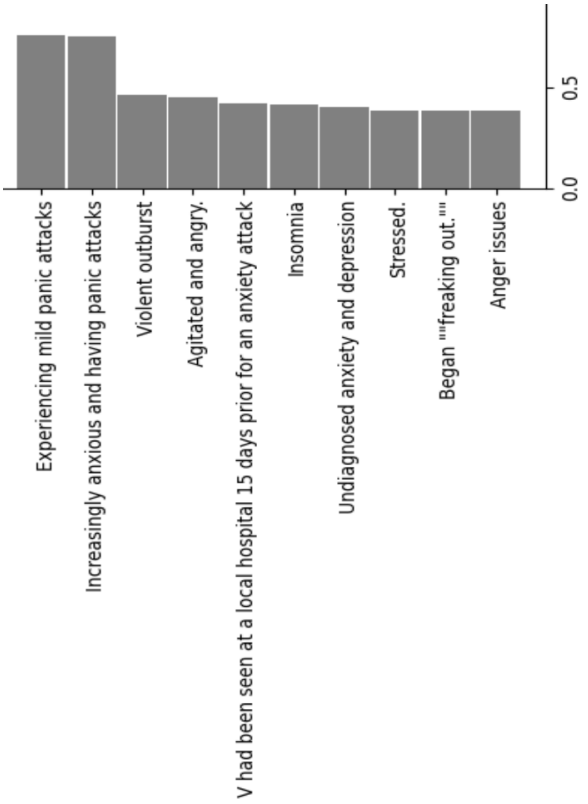

Figure 3: Top 10 descriptions (drawn from CME/LE_CircumstancesOtherText) ranked by embedding similarity with the prompt.

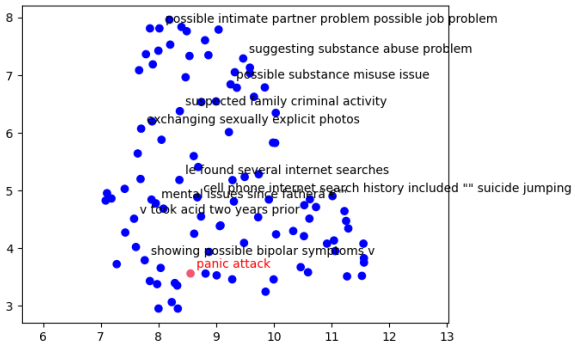

Figure 4: UMAP (McInnes et al., 2020) projection of the embeddings of the keyphrases drawn from CME/LE_CircumstancesOtherText. We extract keyphrases using RAKE (Rose et al., 2010). The embedding of PA is displayed in red and the remaining texts are displayed randomly for clarity.

References

Leland McInnes, John Healy, and James Melville. 2020. Umap: Uniform manifold approximation and projection for dimension reduction. *Preprint*, arXiv:1802.03426.

Daiyoung Oh, Doo-Heum Park, Seung-Ho Ryu, Jee Hyun Ha, and Hong Jun Jeon. 2021. Psychological predictors of bipolarity in panic disorder. *World Journal of Psychiatry*, 11:242 – 252.

Nils Reimers and Iryna Gurevych. 2020. Making monolingual sentence embeddings multilingual using knowledge distillation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Stuart Rose, Dave Engel, Nick Cramer, and Wendy Cowley. 2010. Automatic keyword extraction from individual documents. *Text mining: applications and theory*, pages 1–20.

Peter P. Roy-Byrne, Paul Stang, Hans-Ulrich Wittchen, Bedirhan Ustun, Ellen E. Walters, and Ronald C. Kessler. 2000. Lifetime panic–depression comorbidity in the national comorbidity survey: Association with symptoms, impairment, course and help-seeking. *British Journal of Psychiatry*, 176(3):229–235.

Wenhui Wang, Furu Wei, Li Dong, Hangbo Bao, Nan Yang, and Ming Zhou. 2020. Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers. *Preprint*, arXiv:2002.10957.

Sabrina Wideburg, Catharina Voss, Theresa M. Ollmann, Hanna Kische, Lars Pieper, and Katja Beesdo-Baum. 2024. The dsm-5 panic attack specifier as a severity indicator in mental disorders – findings from a cross-sectional epidemiological study among adolescents and young adults. *Journal of Affective Disorders Reports*, 15:100696.