Enhancing BERT Performance with Contextual Valence Shifters for Panic Detection in COVID-19 Tweets

Sandra Mitrović* †
sandra.mitrovic@idsia.ch
IDSIA-USI/SUPSI
Lugano, Switzerland

Vani Kanjirangat
vani.kanjirangat@idsia.ch
IDSIA-USI/SUPSI
Lugano, Switzerland

ABSTRACT

Panic phenomenon is one of the main challenges in the current pandemic time. In this work, we aim to explore the approaches to detect the panic-related COVID-19 tweets. Aligned to this, we propose an unsupervised clustering approach considering negation cues as an extracted feature input to the pre-trained model. This task cannot be done by simply applying state-of-the-art transformer models, since we observed that they occasionally fail in handling negations. Hence, we propose to utilize features based on Contextual Valence Shifters (CVS) along with the pre-trained BERT embeddings. We evaluate and compare the approaches in an unsupervised setup, using standard clustering metrics on a large set of COVID-19 tweets. The obtained results show that CVS effectively facilitates negation handling (positive/negative tweet discrimination).

CCS CONCEPTS

• **Computing methodologies** → **Information extraction**; **Cluster analysis**; **Transformers**.

KEYWORDS

Contextual Valence Shifters, BERT, Negation Handling, Panic Detection, COVID-19 Tweets

ACM Reference Format:
Sandra Mitrović and Vani Kanjirangat. 2018. Enhancing BERT Performance with Contextual Valence Shifters for Panic Detection in COVID-19 Tweets. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (6th International Conference on Natural Language Processing and Information Retrieval).* ACM, New York, NY, USA, 4 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

In the psychology and sociology literature, panic is defined as a state of sudden and overwhelming fear [1], that can be extremely dangerous and in some extreme cases even lead to life loss [2]. During turbulent periods, both individual and mass panic can be evidenced. It is, therefore, of high importance tracing panic during the COVID-19 pandemic.

Table 1: Outcome of BERT Sentiment Analysis pipeline on few examples containing negation.

Text	BERT SA	Validation
Do not spread rumours!	NEGATIVE	✗
Do not spread panic!	POSITIVE	✓
Panicking does not help	NEGATIVE	✗

Nevertheless, much of the current literature is focusing on other virus-related problems (e.g. studying related scientific research [3], socio-economic impacts [4]). Actually, even before this pandemic, detecting panic has mainly been performed on images (e.g. based on people facial expressions in crowded situations [5]). The only exception is the study of [6], that was assessing panic potential at textual level. However, this was done in a purely supervised manner (involving human annotators) using also non-semantic features e.g. number of hashtags, tweet length, punctuation. In this paper we focus on unsupervised detection of the panic-related content in the COVID-19 pandemic tweets using Contextual Valence Shifters (CVS) features. Since we operate in an unsupervised setup, we first aim at clustering the tweets to distinguish between panic related and non-related tweets. Given that panic is tightly linked with negative emotions such as anxiety and fear, one would expect that applying current state-of-the-art models such as BERT [7] to detect tweet sentiment would render this task easy, however, this is not always the case. As shown in Table 1, there are cases where BERT not only fails in properly handling negated sentences but also shows inconsistencies in judgement, which arise due to improper negation handling. Negation handling is a well-known but still very challenging issue in many NLP tasks. Negations, as lexical patterns, are very important especially in short texts like tweets. To improve the effectiveness of pre-trained models in distinguishing between negative and positive tweets, along with the pre-trained BERT embeddings, we propose using features based on Contextual Valence Shifters (CVS) [8], since existing studies report their positive impact in sentiment classification tasks [9–11].

We evaluate our approach[1][2] against baselines using standard clustering metrics, and demonstrate that the proposed approach exhibits better performances. We also present qualitative analysis to understand the explored approaches in a better way. The main focus of this paper is to improve the pre-trained models with CVS features in a completely unsupervised manner.

*Both authors contributed equally to this research.
†Corresponding author.

[1]Our code is available at: https://github.com/SandraMNE/CVS4PD
[2]Implementation details are available in Appendix B

6th International Conference on Natural Language Processing and Information Retrieval, December 16–18, 2022, Thailand, Bangkok

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

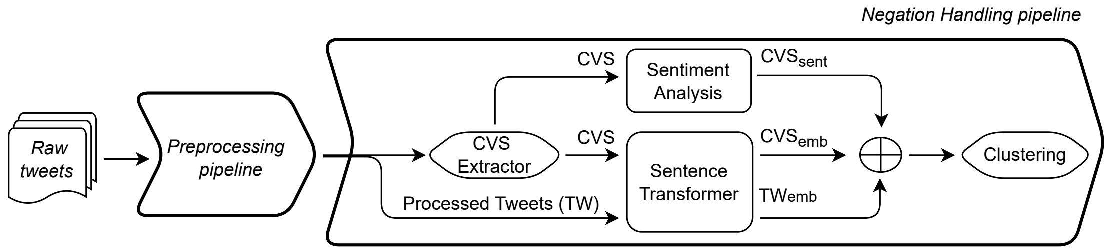

Figure 1: Workflow of the panic detection methodology with its main pipelines.

2 PANIC DETECTION

We design our methodology as a workflow (see Figure 1) consisting of two pipelines, explained in detail in the subsequent sections.

2.1 Preprocessing Pipeline

This pipeline focuses in filtering out the unnecessary tweets, containing positive sentiments or facts which are unrelated to panic. For this, we utilize the existing models to prune out tweets with positive sentiments and thus, to improve the model precision. More precisely, we first use the pre-trained BERT [12] sentiment analyzer pipeline to filter out tweets with positive sentiments. Next, we use Top2Vec [13] to eliminate the sentences related to the neutral/non-panic-related keywords extracted applying YAKE [14], an unsupervised automatic keyword extraction tool. To distinguish between the set of panic and non-panic related keywords, we relied on synonyms of word 'panic' from WordNet [15].

2.2 Negation Handling Pipeline using Contextual Valence Shifters (CVS)

Once the uninteresting tweets are removed, we pass to a *coarse-grained* level of differentiating between the panicking and non-panicking tweets (e.g. "Please do not stress, this too shall pass" as positive and "Omg I am really scared" as negative). To determine whether a sentence conveys positive or negative attitude, one might look at the words that constitute it. Among these, the lexical items named as Contextual Valence (CVS) can have different part-of-speech tags and, consequently, various functions in the sentence, which can alter the attitude, also known as valence of the sentence [8]. In this preliminary study, we focus on one of the main sentence based contextual valence shifters, termed *'negatives'*. For instance, the simple difference brought by a negation such as *not* is highly important in natural language understanding and hence in identifying a sentence as positive/negative. Other negatives include e.g. *never, none, nowhere, nothing, neither*. The additions of such negatives cause a valence flipping. We extract the CVS features using an explicit list of prevalent negatives defined in English language (details in Appendix A).

In Table 1, we demonstrate that the state-of-the-art pre-trained neural models lack the inherent capacity to identify such lexical features, which turns out to be quite important. Hence, we utilize the *'negatives'* as explicit features to improve the efficacy of such pre-trained models.

Concretely, we start by using BERT pre-trained model to generate the embeddings (TW_{emb}) of tweets obtained after the preprocessing pipeline. Next, we extract the negation-related CVS from each of these tweets and embed them using the BERT pre-trained model (CVS_{emb}). Additionally, we extract the sentiment of extracted CVS (CVS_{sent}) as a binary feature using BERT sentiment analysis pipeline. Finally, the concatenation of these three TW_{emb} + CVS_{emb} + CVS_{sent} is used as input for performing K-Means binary clustering.

3 EXPERIMENTAL SETUP

In this section we describe dataset, baselines and (clustering) evaluation metrics.

Dataset: We used publicly available Kaggle Coronavirus Dataset [16] containing tweets with COVID-19-related hashtags.

Negation Handling Baselines: However, we also measure the improvement made by adding CVS information (1. CVS embeddings and 2. related CVS sentiments) on the top of the raw text embeddings (denoted as $TW_{emb} + CVS_{emb}$ and $TW_{emb} + CVS_{sent}$ respectfully).

Evaluation Metrics: We perform evaluation based on three standard metrics for an unsupervised clustering setup. Silhouette Coefficient (SC) [17] measures the mean distance between a sample and all other points in the same class and all other points in the next nearest cluster. We consider two variations of SC, using Euclidean (denoted as SC_{Euc}) and cosine (denoted as SC_{Cos}) distances. Calinski-Harabasz Index (CHI) is based on cluster dispersion [18]. Davies-Bouldin Index (DBI) [19] signifies the average 'similarity' between clusters. Higher SC and CHI and lower DBI indicate good cluster separation.

4 RESULT ANALYSIS AND DISCUSSIONS

Steps described in the **Preprocessing pipeline** led to filtering only 21 576 tweets (about 0.91%), but this allowed us to preserve the true negatives. Further, after removing duplicates, we ended up with 2 182 410 tweets.

Table 2 shows the results of the K-Means clustering in terms of four evaluation measures. Performing 2-Means clustering using our proposed method TW_{emb}+CVS_{emb}+ CVS_{sent} for generating underlying feature vectors, outperforms competitors in terms of SC_{Euc}, SC_{Cos} and CHI and comes very close to TW_{emb}+CVS_{emb}, in terms of DBI.

Figure 3 demonstrates that our method clearly separates the two

Table 2: Results of K-Means (K=2) clustering applied on different competing methods. Best result per measure is marked in bold-face. Due to computational issues both SC measures are calculated using 500K sample size.

Method	SC_{Euc}	SC_{Cos}	CHI	DBI
TW_{emb}	0.0227	0.0472	56480.2737	6.1329
TW_{emb}+CVS_{sent}	0.0226	0.0473	56251.6503	6.1452
TW_{emb}+CVS_{emb}	0.3304	0.5526	576461.8613	**1.6827**
TW_{emb}+CVS_{emb}+CVS_{sent}	**0.3315**	**0.5535**	**577342.6443**	1.6830

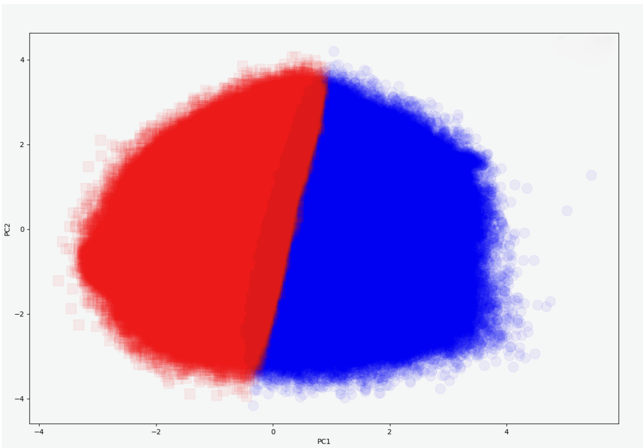

Figure 2: Cluster annotations on PCA projections for TW_{emb} method.

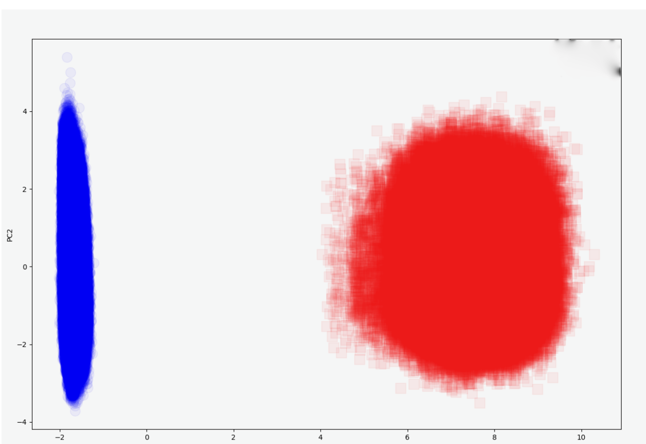

Figure 3: Cluster annotations on PCA projections for $TW_{emb} + CVS_{emb} + CVS_{sent}$ method.

clusters, thus eliminating the blurry stripe of panic (denoted by blue color) and non-panic related tweets (denoted by red color) visible in the middle of Figure 2.

4.1 Qualitative Analysis

For the qualitative analysis, we randomly selected seven tweets whose cluster assignments with respect to methods TW_{emb} and $TW_{emb} + CVS_{emb} + CVS_{sent}$ (preserving original PCA projections) are shown in Figure 4. Here, the color of tweet text represents the expected true label, the color of **marker left side** corresponds to the cluster according to TW_{emb}, while the color of **marker right side** corresponds to TW_{emb}+CVS_{emb}+CVS_{sent} cluster (in all cases, as before, blue=panic class, red=non-panic class). Agreement of colors indicates correct clustering (as e.g. in 'time to panic..' for both methods and in 'dont panic raab..' for $TW_{emb} + CVS_{emb} + CVS_{sent}$). We can see that TW_{emb} fails in five out of seven sample tweets, while $TW_{emb} + CVS_{emb} + CVS_{sent}$ fails in only two tweets, those containing terms 'stop' and 'over'. This is expected, given that we utilize only CVS 'negatives' as features (see Appendix A).

5 CONCLUSION

In this paper, we propose a pipelined approach that helps in detecting panic from COVID-19 tweets. This is being achieved by utilizing contextual valence shifters and their sentiments as features for negation handling combined with pre-trained BERT embeddings. This preliminary study shows that the proposed approach looks promising in a completely unsupervised setup. From the analysis of obtained results, it is obvious that the negation handling part could be further improved by using more advanced CVS (e.g. intensifiers, modals). We are aware that the correctness of clustering cannot be judged directly by the clustering metrics and that instead, manual analysis of the clusters gives better understanding of the positive impact of utilizing CVS features. Therefore, dataset annotation might be considered as it would allow the use of existing supervised approaches.

A CONTEXTUAL VALENCE SHIFTERS

The Contextual Valence Shifters are extracted based on the presence of the following negations in the input text:
["aint", "arent", "cannot", "cant", "couldnt", "darent", "didnt", "doesnt", "ain't", "aren't", "can't", "couldn't", "daren't", "didn't", "doesn't", "dont", "hadnt", "hasnt", "havent", "isnt", "mightnt", "mustnt", "neither", "don't", "hadn't", "hasn't", "haven't", "isn't", "mightn't", "mustn't", "neednt", "needn't", "never", "none", "nope", "nor", "not", "nothing", "nowhere", "oughtnt", "shant", "shouldnt", "uhuh", "wasnt", "werent", "oughtn't", "shan't", "shouldn't", "wasn't", "weren't", "without", "wont", "wouldnt", "won't", "wouldn't"]

B OTHER IMPLEMENTATION DETAILS

The code is completely developed in Python, using HuggingFace library for BERT-based computations. For Yake and Top2Vec author original implementations available on Github were used.
We used standard architecture and default parameters for BERT.

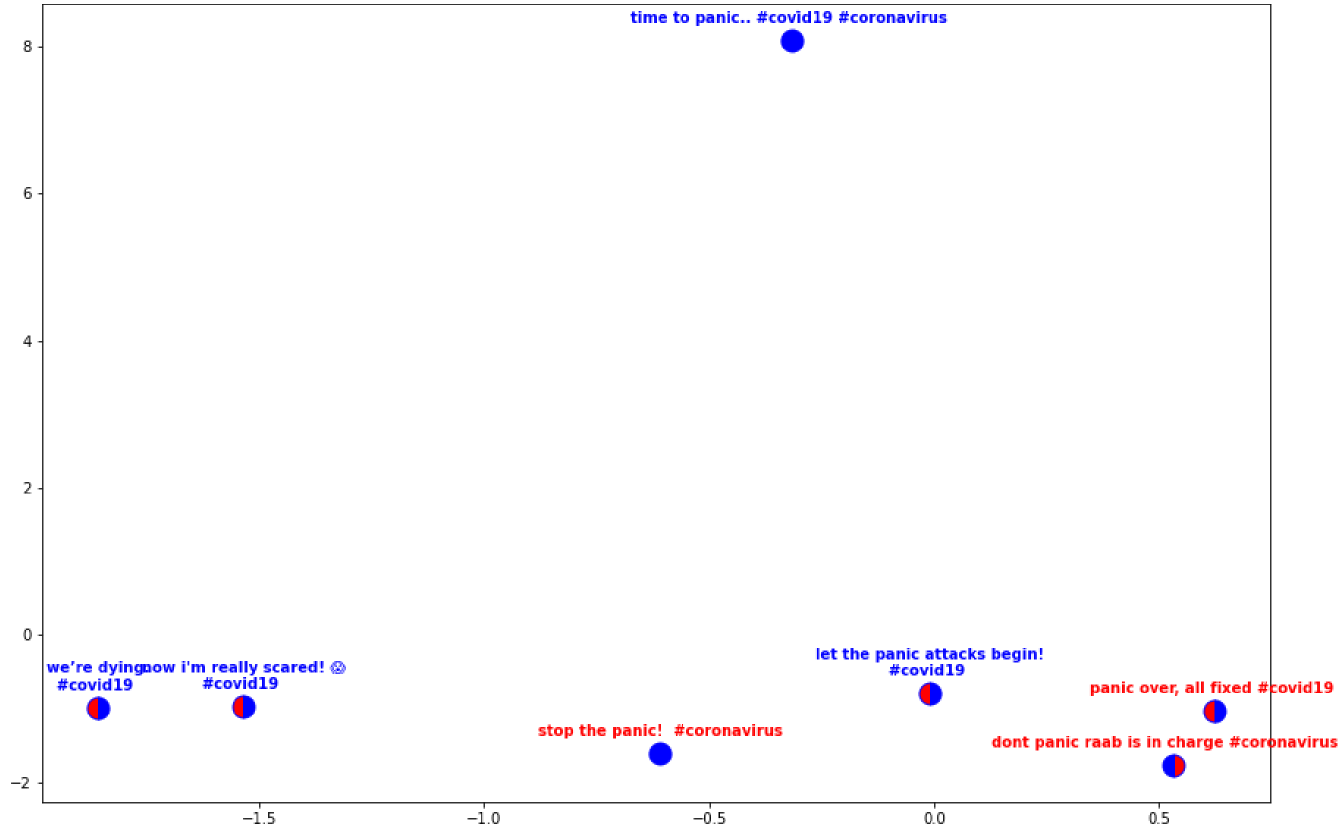

Figure 4: Cluster annotations of sample tweets for TW_{emb} (left side color of marker) and TW_{emb}+CVS_{emb}+CVS_{sent} (right side color of marker). Expected true label is expressed by the tweet text color.

Computations are performed on the HPC cluster with CPU nodes having following characteristics: 2 x Intel Xeon E5-2650 v3 @ 2.30GHz, 20 (2 x 10) cores and 128GB DDR4 @ 2133MHz.
Most demanding clustering jobs took approximately 8h on 10 nodes (with precalculated embeddings).

REFERENCES

[1] Floyd E Bloom, Nicholas C Spitzer, Fred Gage, and Tom Albright. *Encyclopedia of Neuroscience, Volume 1*, volume 1. Academic Press, 2009.
[2] D Jolene Kinley, Brian J Cox, Ian Clara, Renee D Goodwin, and Jitender Sareen. Panic attacks and their relation to psychological and physical functioning in canadians: results from a nationally representative sample. *The Canadian Journal of Psychiatry*, 54(2):113–122, 2009.
[3] Qingyu Chen, Alexis Allot, and Zhiyong Lu. Litcovid: an open database of covid-19 literature. *Nucleic Acids Research*, 2020.
[4] Maria Nicola, Zaid Alsafi, Catrin Sohrabi, Ahmed Kerwan, Ahmed Al-Jabir, Christos Iosifidis, Maliha Agha, and Riaz Agha. The socio-economic implications of the coronavirus and covid-19 pandemic: a review. *International journal of surgery*, 2020.
[5] Xuguang Zhang, Xiaohu Shu, and Zhen He. Crowd panic state detection using entropy of the distribution of enthalpy. *Physica A: Statistical Mechanics and its Applications*, 525:935–945, 2019.
[6] Anuja Hariharan, Verena Dorner, Christof Weinhardt, and Georg W Alpers. Detecting panic potential in social media tweets. 2017.
[7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, 2019.
[8] Livia Polanyi and Annie Zaenen. Contextual valence shifters. In *Computing attitude and affect in text: Theory and applications*, pages 1–10. Springer, 2006.
[9] Alistair Kennedy and Diana Inkpen. Sentiment classification of movie reviews using contextual valence shifters. *Computational intelligence*, 22(2):110–125, 2006.
[10] Vo Ngoc Phu and Phan Thi Tuoi. Sentiment classification using enhanced contextual valence shifters. In *2014 International Conference on Asian Language Processing (IALP)*, pages 224–229. IEEE, 2014.
[11] Sara A Morsy and Ahmed Rafea. Improving document-level sentiment classification using contextual valence shifters. In *International conference on application of natural language to information systems*, pages 253–258. Springer, 2012.
[12] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
[13] Dimo Angelov. Top2vec: Distributed representations of topics. *arXiv preprint arXiv:2008.09470*, 2020.
[14] Ricardo Campos, Vítor Mangaravite, Arian Pasquali, Alípio Mário Jorge, Célia Nunes, and Adam Jatowt. Yake! collection-independent automatic keyword extractor. In *European Conference on Information Retrieval*, pages 806–810. Springer, 2018.
[15] George A Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J Miller. Introduction to wordnet: An on-line lexical database. *International journal of lexicography*, 3(4):235–244, 1990.
[16] Kaggle. Coronavirus (covid19) tweets - early april. https://www.kaggle.com/smid80/coronavirus-covid19-tweets-early-april, 2020.
[17] Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.
[18] Tadeusz Caliński and Jerzy Harabasz. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27, 1974.
[19] David L Davies and Donald W Bouldin. A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, (2):224–227, 1979.