

NHK STRL at WNUT-2020 Task 2:  
GATs with Syntactic Dependencies as  
Edges and CTC-based Loss for Text Classification

Yuki Yasuda, Taichi Ishiwatari, Taro Miyazaki, Jun Goto

NHK Science and Technology Research Laboratories

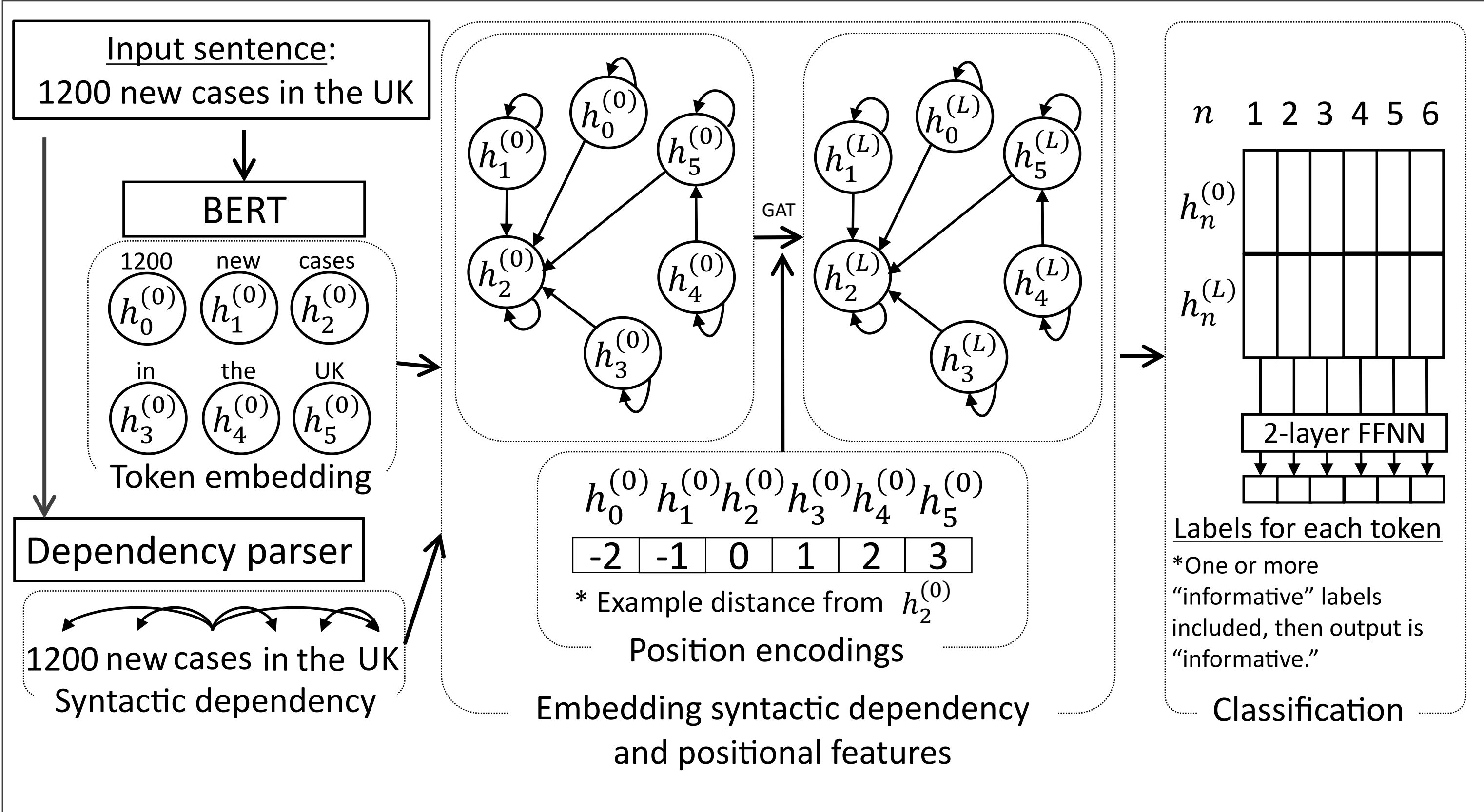


Introduction

- The outbreak of COVID-19
  - It’s important for everyone to understand the situation and grasp the latest information to avoid causing too much fear and panic.
- SNSs as real time and massive information sources
  - There is a lot of information regarding user’s personal events including that related to COVID-19 in real time.
  - Contain too much information on miscellaneous topics.
- Monitoring systems for COVID-19<sup>[1] [2]</sup>
  - Depend on manual work such as using cloud sourcing to extract informative posts from massive numbers of uninformative ones.
- We attempted to develop a method to extract important information.

Method

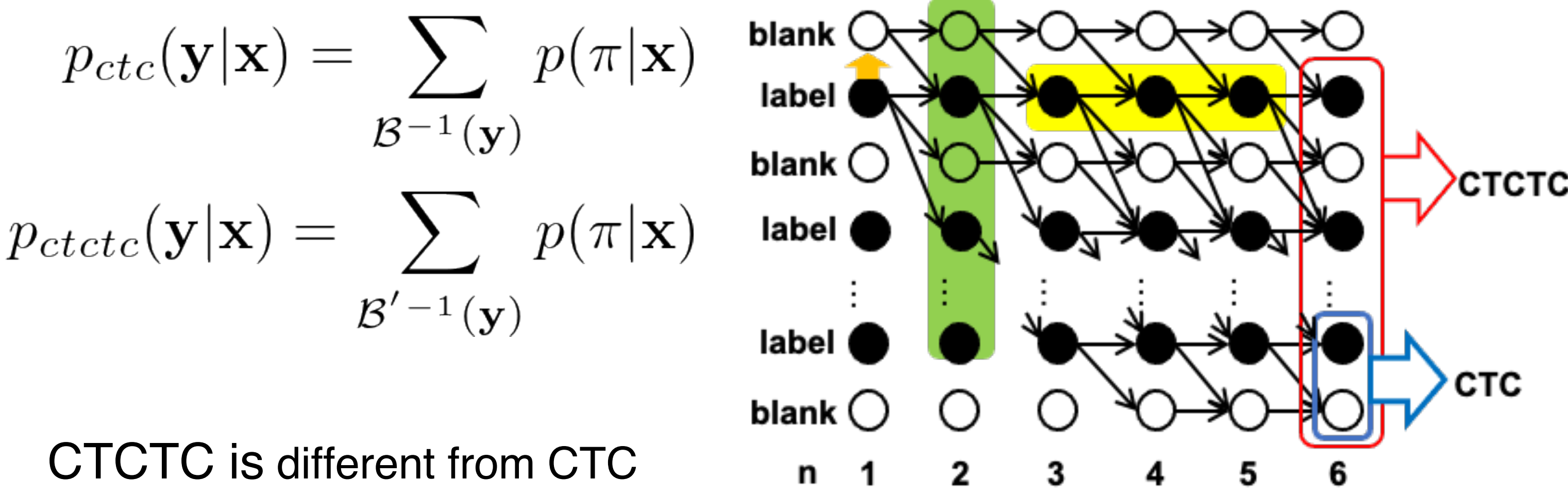
- The overview of our method



- Graph Attention Network (GATs)<sup>[3]</sup>
  - Nodes: Word embeddings fed from BERT<sup>[4]</sup>
  - Edges: Syntactic dependencies parsed by spaCy<sup>[5]</sup>
  - Position encodings: distances from each tokens to others<sup>[6]</sup>

$$\mathbf{h}_n^{(l)} = \sum_{m \in \mathcal{N}(n)} \alpha_{nm}^{(l-1)} \mathbf{W}^{(l-1)} \mathbf{h}_m^{(l-1)}$$
$$\alpha_{nm}^{(l-1)} = \text{softmax}_n \left( \text{LeakyReLU}(\mathbf{a}^T [\mathbf{W} \mathbf{h}_n || \mathbf{W} \mathbf{h}_m] + \text{PE}_{nm}) \right)$$

- CTC<sup>[7]</sup> for Text Classification (CTCTC)



CTCTC is different from CTC

in term of the inverse of the many-to-one map B of all possible labellings.

- Smoothing<sup>[8]</sup> (Green and Yellow Box)

$$l'(k) = (1 - \epsilon) \delta_{k,l} + \frac{\epsilon}{K}$$

- Leaking (Orange arrow)

$$p'_{i,blank} = (1 - \epsilon') p_{i,blank} + \epsilon' p_{i,\bar{y}}$$

Experiments

- Experimental settings
  - Dataset: COVID-19 English Tweet(W-NUT shared task)<sup>[9]</sup>
    - For the validation data based experiment, we used training data contains 7,000 tweets and validating data contains 1,000 tweets for training and validating, respectively.
    - For the testing data based experiment, we used 8,000 tweets mixed from the training and validating data for 4-fold cross validation. Then, an ensemble of the best model of each fold data were used for testing data.
  - Baseline methods
    - Only BERT CLS vector (use Cross Entropy loss)
    - BERT with GATs and Max-pool (use Cross Entropy loss)
    - All token vectors from BERT (use CTCTC loss)
- \*In addition, we’ve investigated whether Position encodings should have fixed parameters or a learning mechanism.

- Result

#	GATs parameters Use GATs	Positional feature	CTCTC parameters Use CTCTC	Label smoothing	Token smoothing	Leaking	F1 score
1							0.9154 ± 0.0041
2	✓						0.9134 ± 0.0015
3	✓	Fixed					0.9151 ± 0.0026
4	✓	Learned					0.9151 ± 0.0009
5			✓				0.0000 ± 0.0000
6			✓	✓			0.9128 ± 0.0026
7			✓	✓			0.9133 ± 0.0052
8			✓	✓	✓		0.9153 ± 0.0024
9	✓	Fixed	✓	✓	✓	✓	0.9172 ± 0.0027
10	✓	Learned	✓	✓	✓	✓	<b>0.9175 ± 0.0044</b>

Test dataset	Team / Method	F1 score
	Ours (#9 in Table 1)	0.8898
	Ours (#10 in Table 1)	0.8885

Discussion

- GATs without CTCTC
  - Can not make the best use of the improved vectors because they are mixed up vectors of tokens into one vector using max-pooling, so some of the details of the vectors are lost.
- CTCTC without GATs
  - We observed that the output vectors of each token in the sentence are almost the same. This means that token-level information is lost. By using GATs with CTCTC, we can avoid losing the information, which leads to good results.

Conclusion and Future work

- Conclusion
  - Proposed a GATs-based model that embeds token-to-token relations, and a loss function that can learn classes for each tokens.
  - Conducted evaluations using the identification of informative COVID-19 English Tweets dataset and confirmed that our proposed methods are effective.
- Future work
  - Determine whether CTCTC can work for other tasks especially for the classification into large amount of classes.
  - Exploit pre-trained models other than BERT. Especially for task specific models. Such as BERTweet<sup>[10]</sup> and CT-BERT<sup>[11]</sup>.

References

[1] <https://coronavirus.jhu.edu/map.html>  
[2] <https://www.covid19healthsystem.org>  
[3] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2018. Graph Attention Networks. International Conference on Learning Representations.  
[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. 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, Minneapolis, Minnesota. Association for Computational Linguistics.  
[5] Matthew Honnibal and Ines Montani. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear.  
[6] Taichi Ishiwatari, Yuki Yasuda, Taro Miyazaki, and Jun Goto. 2020. Relation-aware graph attention networks with relational position encodings for emotion recognition in conversations. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP 2020).  
[7] Alex Graves, Santiago Fernández, Faustino Gomez, and Jürgen Schmidhuber. 2006. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In Proceedings of the 23rd international conference on Machine learning, pages 369–376.  
[8] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2818–2826.  
[9] <http://noisy-text.github.io/2020/covid19tweet-task.html>  
[10] Dat Quoc Nguyen, Thanh Vu, and Anh Tuan Nguyen. 2020a. BERTweet: A pre-trained language model for English Tweets. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations.  
[11] Martin Müller, Marcel Salathé, and Per E Kummervold. 2020. Covid-twitter-bert: A natural language processing model to analyse covid-19 content on twitter. arXiv preprint arXiv:2005.07503.