Periodicals of Engineering and Natural Sciences ISSN 2303-4521
Vol. 8, No. 4, December 2020, pp.2232-2241

A hybrid deep learning and NLP based system to predict the spread of Covid-19 and unexpected side effects on people

Mohamed Adel Al-Shaher
Computer Science - Information Technology, Computer Department, College of Computer Science and Mathematics; University of Thi-Qar, Nassiriyah, Iraq

ABSTRACT

The aim of this paper is to deeply analyze the unexpected side effects of people during the Covid-19 pandemic using the RNN based NLP sentiment analysis model. The normalized correlation values that is obtained by computing the cases values between the people behavior extracted and covid-19 reported case also has values close to 1 million by the end of June 2020 provided in dataset. In this research work, with more time, we would like to continue from the results we have achieved by training the RNN with NLP based sentiment analysis model for more extended periods of time for predicting the behavior of people during Covid-19 pandemic with 76.71% of accuracy which is high as compared to the CNN, such as days or weeks, in order to see how results can improve. The advancement in this field created an urge in me to research more on the techniques and methodologies developed for covid-19 extraction. During the outbreak of an epidemic, it is of immense interest to monitor the effects of containment measures and forecast of outbreak including epidemic peak affecting the behavior of people. To confront the change in behavior, a simple RNN based NLP sentiment analysis model is used to simulate the number of affected patients of Coronavirus disease. Our initial problem formulation involved investigating the ideal conditions and preprocessing for working with a specific NLP task: predicting the behavior during the specific time of May 20 – June 20 in 2020 for all four traits of common people during the Covid-19 pandemic.

Keywords: *Deep Learning, NLP, RNN, Covid-19, Sentiment Analysis, Side-Effects, Pandemic*.

Corresponding Author:

Mohamed Adel Al-Shaher
Computer Science - Information Technology
Computer Department, College of Computer Science and Mathematics, University of Thi- Qar
Nassiriyah; Iraq
Email: alshaher_comp82@sci.utq.edu.iq; alshaher2016@gmail.com; alshaher2006@yahoo.com

1. Introduction

One of the better-known applications of AI-powered NLP today is using it to correct grammar and spelling. The foremost implementation of this today is Grammarly, an ML-based grammar checking browser extension with over 10,000,000 downloads, also available for MS Office and as a standalone application [1]. Grammarly corrects grammar and spelling as the user types, using NLP to detect errors and provide corrections or suggestions [2]. As of today, Grammarly is only available in English, reflecting the priority status of the language in the field [3]. A similar tool for grammar and spell-checking based on ML-powered NLP techniques is Microsoft's Bing Spell Check API, which, again, is only available for English [4]. The primary problem is that the size and availability of datasets for English are far vaster than those for other languages. Using NLP methods on a language with few speakers or few written works would be next to impossible, as the quality and quantity of written output are so important. Additionally, the written output is only one aspect of the resources required for effective NLP. Tools for preparing datasets into a machine-readable format are needed in order for NNs to create full and accurate models of natural language. These tools are readily available for English, but scarce for other languages and especially smaller languages as mentioned in [5]. As an example, one of the most famous artificial intelligences (AI), IBM's Watson, only provides NLP tools for a limited number of languages, the smallest of which has 4.1 million speakers [6]. Even then, only English has full support. On the other hand, as long as a language has enough written output that using NLP becomes feasible, we feel that there is substantial value to be obtained from testing and implementing NNs working in these languages as mentioned in [7]. English is one of those languages, and as native speakers of English, we the authors feel that we have an opportunity to make a substantial contribution to the field. The utility of providing similar tools for other languages is obvious. The learning step for every language is different; otherwise, the same algorithms used for grammar and spell checking in English could be used for other languages [8]. Therefore, the onus is on ML researchers to develop and provide the algorithms and tools for working with natural languages in a wide variety of languages. Datasets for NLP are measured by the number of words they contain, with larger ones measuring in the billions. The largest dataset available for English, the Creative Commons-licensed English Corpus, contains one billion words extracted from a range of sources from 1950 to 2015 [9]. The NLP Corpus is however unsuited to our research, as we require the data to be as grammatically correct as possible. As the Corpus contains sources such as social media, this is not a guarantee as mentioned in [10]. Instead, we will use the smaller dataset, collected from the English newspaper of the same name. Every dataset is different, meaning that one algorithm cannot possibly cover them all. Different algorithms have different levels of accuracy when learning the same dataset as mentioned in [11]. Thus, our objective with this work is twofold: to perform experiments and benchmarks to first find the optimal hyper-parameters to use in working with the COVID-19 Open Research Dataset Challenge (CORD-19), and second, to find the best way the dataset can be preprocessed for training a grammar and spell-checking sentiment analysis based natural language processing model.

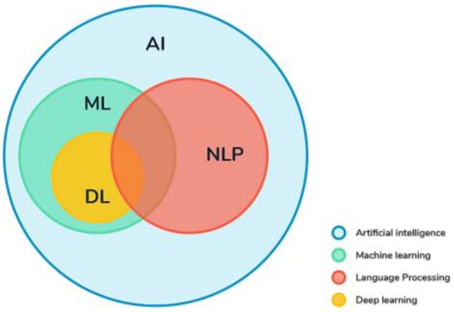

Figure 1. The natural language processing (NLP) is a subset of artificial intelligence (AI) and use the techniques of ML/DL [12]

1.1. Research problem

We would like to contribute to research in the field of NLP from the perspective of a non-English language, specifically Natural language processing. Using the COVID-19 Open Research Dataset Challenge (CORD-19) and NLP based Multi-Domain Sentiment Dataset as a dataset, we want to investigate how to achieve spelling and grammar correction adapted to the Natural language processing, by testing different inputs and ways to preprocess the dataset to best train a grammar and spell-checking machine learning model. The following section presents the primary and secondary questions that the research problems to solve. The questions are stated from the background to the research and covers both data acquisition and the use of NLP as a method.

- Is it possible to generate training data by using a COVID-19 Open Research Dataset and a measured ground truth of average people behavior percentage?
- Does the approach of natural language processing techniques as a way of automating the accurate behavior prediction of common people during the pandemic?
- Is a segmentation and sentiment analysis of the people behavior necessary to obtain good results?
- What aspects need to be considered if further work on the subject is to be performed?

1.2. Aim of study

Natural Language Processing is one of the most studied topics within AI today. Unfortunately, for many languages there are not enough substantial coid-19 data available, often because governments do not do enough to promote language studies and make available language resources for understanding of Covid-19. As a result, most research, implementations, and examples within Natural Language Processing are only available for the English language or have limited support for languages other than English. For that reason, we want to explore how to implement NLP in a non-English language, specifically, natural language processing. Our goal is to further the field of NLP by demonstrating how a model can be trained to detect Natural language processing - specific grammatical errors at the same time as it corrects spelling.

- Provide an overview of previous works and achievements on classification of corona virus and reported cases in contrast with the behavior people by analyzing the sentiments through natural language processing.
- Apply NLP based sentiment analysis method to the unified two open source datasets to understand the common people behavior with less data loss and more accuracy.
- Develop an advanced NLP based sentiment analysis technique based novel methodology with good results for covid-19 and predict the unexpected side-effects on people during the time of pandemic.

- Understanding people certain patterns and relations within the dataset through sentiment analysis of during the Covid-19 pandemic for forecasting the spread of infection.
- Provide an overview of previous works and achievements on utilization of NLP as a tool for the predicting the spread of virus during Covid-19 pandemic.

1.3. Literature review

Researchers propose a hybrid algorithm for spell-checking based on combining edit distance with n-grams, which group consecutive words in sets of two or three [13]. Though the report is not focused on spell-checking in the context of sentiment analysis based natural language processing's, the algorithm they design shows promise as it is demonstrably faster than using edit distance alone. In [14], researchers discuss the particular difficulties of natural language processing concerning NLP. Although their research is concerned with cross-language information retrieval, a separate field, it is still relevant as it focuses on a number of language-specific problems that we will also face in our research. Researchers have highlighted the importance of deep learning and sentiment analysis based natural language processing's for natural language processing in [15]. They argue that most of the problems found in NLP can be solved by using these machine learning techniques.

Researchers in [16] has created a natural language processing-based spell-checker using Python, Theano, and Keras, using a billion-word dataset that he injected with random noise [17]. Though he is mainly concerned with spelling rather than grammar, spelling is an important consideration in this study, and his work has very much in common with the context that we will be working in. Researchers use word2vec to assess both local and global coherence of a document, which is relevant to our research as an important feature of well-written texts is that they are locally (adjacent sentences) and globally (adjacent paragraphs or remote sentences) coherent [18]. This aids in understanding how a sentiment analysis based natural language processing can be trained to understand subtler aspects of grammar. Researchers in [19] is an example of a study concerning NLP powered by machine learning for a non-English language, in this case, Turkish. They use machine learning techniques to create word clusters using the n-gram model, in order to group similar words. Researchers in [20] propose a text corrector for Arabic using a seq2seq sentiment analysis based natural language processing language model, achieving a good result with a special biased-decoding method. Though most research on the topic is done on NLP for English, research work and the previous study show that there is both a need and an interest for research done on other languages as mentioned in [21].

For eliminating the noise and display better results, better threshold values can be selected. Also, in [22] the only depicts one covid-19 is extracted instead of both the covid-19s. This has resulted due to two possibilities. First is the area range provided is smaller which neglects the require covid-19 portion of the data. Second possibility is due to higher threshold value. Therefore, in this paper [23] which is presented on increasing the area range and decreasing the threshold, the region of interest can be extracted accurately. In this way unnecessary noise can be eliminated and correlation values that are tending towards one are displayed. However, in paper [24] author devised any of the data mining techniques, doing any changes may result in accurate results for some data and produce noisy output for some data. Also, after extracting the foreground and background portions of the data, researchers have determined how the cases in the covid-19 region are closely related to each other by computing the cross-correlation between target variable and covid-19 reported cases.

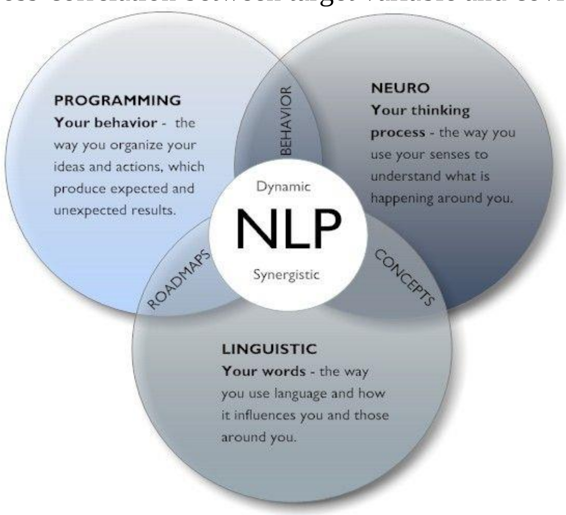

Figure 2. The basic components of natural language processing [25]

Prior to modelling, a number of initial data analysis steps must be discussed with background knowledge for the analysis study of covid-19 cases detection and reporting with NLP.

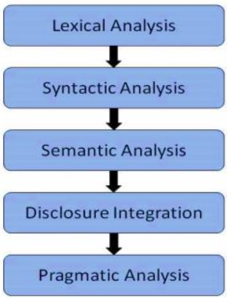

Figure 3. The typical flow of for AI-based natural language processing

Author in [26] has used the methods except manages to extract the covid-19 correctly. Also compared to the first output, we can see some noise in the output. Due to this, the normalized cross correlation values displayed are not tending towards. This is because of the inefficient threshold values used to segment the covid-19. This is because, there is hardly any noise seen in the data. However, this normalized cross-correlation technique is not applied on the covid-19 generated from the machine learning technique as mentioned in [27]. The reason is because in reported cases, researchers determined the covid-19 features and plot the regions based on the features extracted using the data mining techniques. Due to this, the region is not actually superimposed on the original data. The region is just highlighted to display the result. Hence it was not possible to calculated normalized cross-correlation for the covid-19 extracted using data mining technique. However, in [28] output of the covid-19 data mining is so clear that by just looking at the output, one can predict the accuracy.

2. Methodology

Our language model and accompanying scripts are written using the Python programming language, a popular language for ML tasks. There are several frameworks available for these tasks available, including Keras, PyTorch, and TensorFlow. Though any of these frameworks would serve us well in our research, we chose Google's TensorFlow, for its wide usage and the excellent availability of tutorials and quality documentation. We specifically use the GPU variant of TensorFlow to enable GPU acceleration and train the models faster.

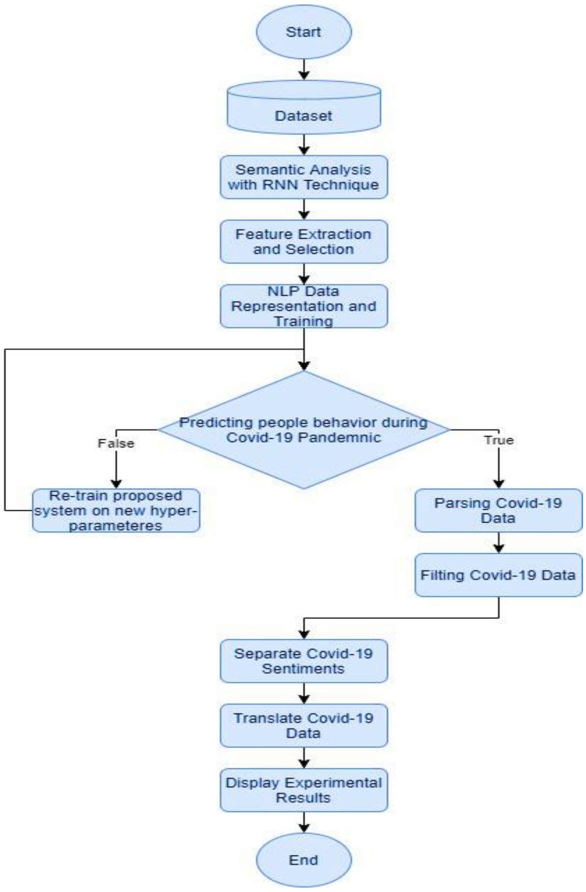

Figure 4. The flowchart represents the methodological approach being used in this research work

Our code is structured as a number of scripts responsible for different tasks, from downloading and preprocessing datasets to scripts detailing the sentiment analysis based natural language processing and scripts for training and running it. We used the code as a starting point. Table 1 shows the specifications of the computer used to train the model. This machine was the primary on which we trained the model, though we also used covid-19 for various testing purposes.

Table 1. Specifications of the computer used to train the model

OS	Ubuntu 18.04
RAM	32GB
Processor	Intel i7 8700K
GPU	GeForce GTX 2180

2.1. Training

While training, the script produces several logs and saves two different checkpoints for people who have covid-19 symptoms. The logs are used to give an overview of the training and provides data for our controlled experiments. The checkpoints save the state of the model. In case the program crashes or has to be paused, training can continue from the first checkpoint, which functions as a backup. The second checkpoint is intended to be used to run and test the model later. It is only saved once the validation loss of the model has reached a new lowest point. By running validation on the model using the validation dataset, which the model does not see during training, we can ensure that the model has not learned the training data too well. We save the checkpoint only when the model has reached a new lowest validation loss, rather than at the end of every epoch. This is called early stopping and is a proven technique for avoiding overfitting.

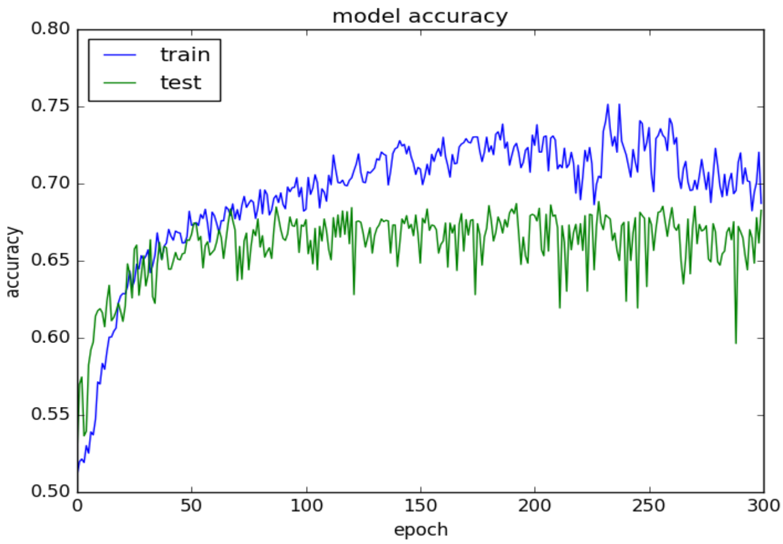

Figure 5. The RNN model for NLP based system being deployed at 300 epochs of training and testing

2.2. Semantic Analysis for Covid-19

The sentiment analysis based natural language processing has a number of hyper-parameters that can be changed in order to alter the effectiveness of training. If certain hyper-parameters are too low or too high, the sentiment analysis based natural language processing will not produce the desired result. The relevant hyper-parameters are number of epochs, batch size, number of hidden layers and size of the NLP, learning rate, and dropout. Though the number of epochs was set at 300 each time, we set a stop whereby the model would stop training if the model ran for ten epochs without achieving a new low in the validation loss. The full results contain the number of epochs the model ran before aborting. Loss is the output of the loss function, which calculates the error of a single training example. The loss we present in our results is the average loss for a whole epoch. Accuracy, on the other hand, is the percentage of testing sentences that were 100% correctly predicted. In order to measure how well the model performs with the different errors we introduced into the dataset, we calculate four different levels of accuracy: typo correction accuracy, Covid-19 (split compound words) correction accuracy, gender agreement correction accuracy, and average accuracy. Another consideration is the COVID-19 Open Research Dataset Challenge (CORD-19) dataset used. We use the same dataset to train the sentiment analysis based natural language processing each time; however, in different runs we preprocess the data differently. We did this in order to explore the best way that the dataset can be prepared in order for the network to learn the most common grammar and spelling mistakes in Swedish. Later runs are trained with a script that adds a maximum of one error per word, while earlier runs have no such restriction. Different type behavior of people is given by:

- Not concerned.
- Slightly change in behavior.
- Standard change in behavior.
- Extreme change in behavior.

2.3. RNN Technique with Semantic Analysis

The RNN technique is used for analyzing the behavior of people during the pandemic that is proposed to extract the Covid-19 cases as reported. RNN is an unsupervised classification technique where the initial set of neural links along with the layer centers needs to be initialized before segmenting using this RNN algorithm. Therefore, classifying the people behavior using RNN technique entirely depends on the value of neuron in different layers and the mean centers initialized. In the research paper, the value of neuron is considered as 4. That means, we will understand the behavior into 4 different categories each represented by a different order. To performs RNN classification, we first covert our covid-19 reported case to linear space using the python programming. This resultant covid-19 reported case is passed as an input to RNN network which generates the set of layers and neurons centroids. The covid-19 reported case in linear space is re-converted back to spatial domain and the individual classification results are sub plotted as a result in terms of people behavior. Based on the selected behavior during the pandemic, the people behavior is extracted and displayed as one of the class results for most of the covid-19 reported cases. Low dimensional models, with small number of compartments and having parameters which can be determined with the real data with good precision, are better to study and forecast the pandemic.

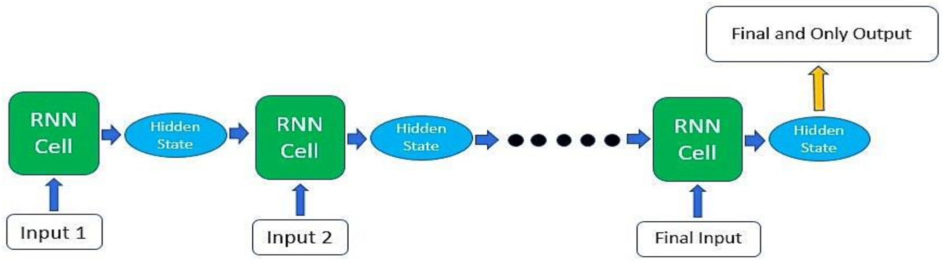

Figure 6. The basic architecture for RNN with NLP based input

3. Results

The RNN manages to show fairly accurate results for all of the covid-19 reported cases. The people behavior region is represented by a white portion for thresholding and covid-19 reported case segmentation. In covid-19 reported case segmentation, the people behavior is represented by a reported cases whereas for Covid-19 positive people behavior extracted is also evalated. For RNN, each of the layer is sub plotted from where we can understand the layer in which the people behavior is present. The normalized correlation values that is obtained by computing the cases values between the people behavior extracted and covid-19 reported case also has values close to 1 million by the end of June 2020. This is because, there is hardly any spike seen in the covid-19 reported case. However RNN with NLP sentiment analysis technique is applied on the people behavior generated from the two datasets acquired. The reason is because in RNN, we determine Covid-19 features and plot the regions based on the features extracted using the python programming functions. Due to this, the behavior of is not actually superimposed on the orginal covid-19 reported case. The region is just highlighted to display the result. Hence it was not possible to calculated normalized cross-correlation for the people behavior extracted using RNN technique with NLP sentiment analysis. However by looking at the output, RNN with NLP sentiment produces the best result as compared to other results.

Table 2. Various parameters of people's behavior during Covid-19 pandemic with difference and change percentage being evaluated by RNN with NLP

People Behavioral Parameters	May 20	June 20	Difference	Percent/Change
Washing your hands more often	46.2	76.5	30.3	65.7
Avoiding large public gatherings	26.2	75.4	49.2	188.1
Avoiding shaking hands	25.6	74.3	48.6	189.6
Avoiding restaurants	12.8	72.2	59.4	464.4
Avoiding movie theaters	15.6	65.7	50.0	319.8
Using hand sanitizer more often	33.9	61.8	27.9	82.2
Avoiding public transportation	17.7	58.7	41.0	232.1
Avoiding sporting events	12.1	53.7	41.6	344.0
Cancelling vacation travel	11.4	50.8	39.4	346.3
Avoiding Doctors' Offices/Hospital	13.4	43.6	30.2	224.4
Cancelling business trips	6.2	24.5	18.3	293.4
Wearing a mask in public	4.3	23.9	19.6	459.4
None of the above	27.3	2.8	-24.5	-89.6

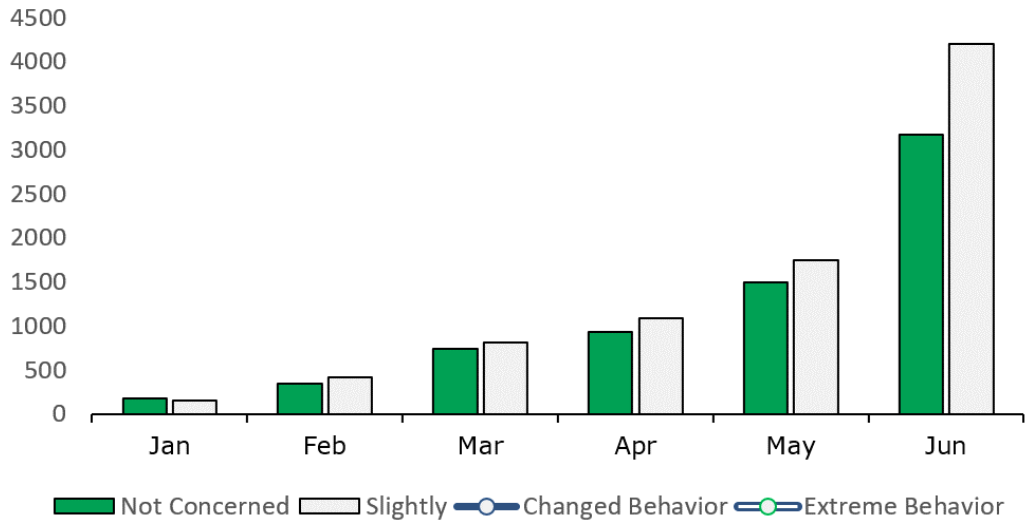

Figure 7. The apprehension or concern of people about the Covid-19 in first quarter of 2020 using the RNN with NLP sentiment analysis

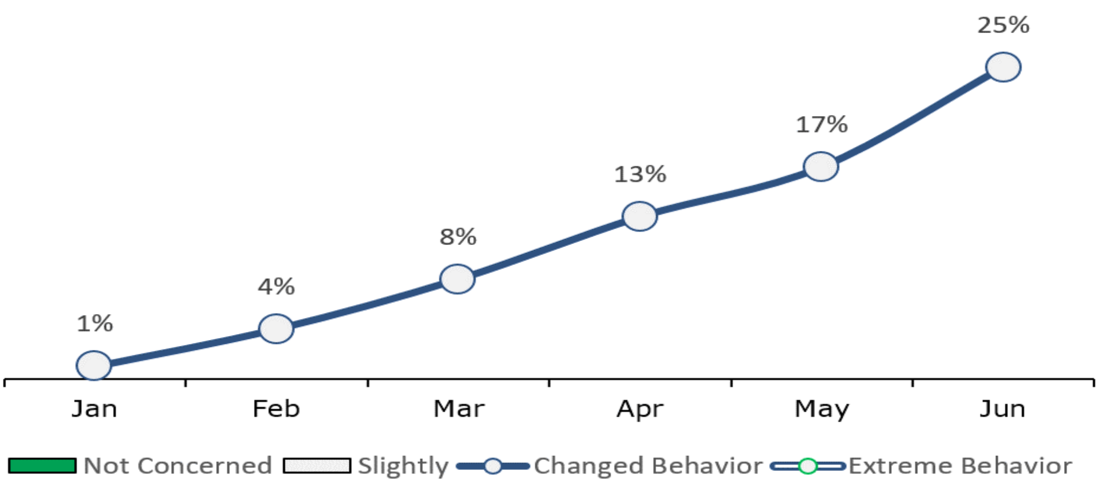

Figure 8. The line represents the standard change in behavior of people during the Covid-19 in first quarter of 2020 using the RNN with NLP sentiment analysis

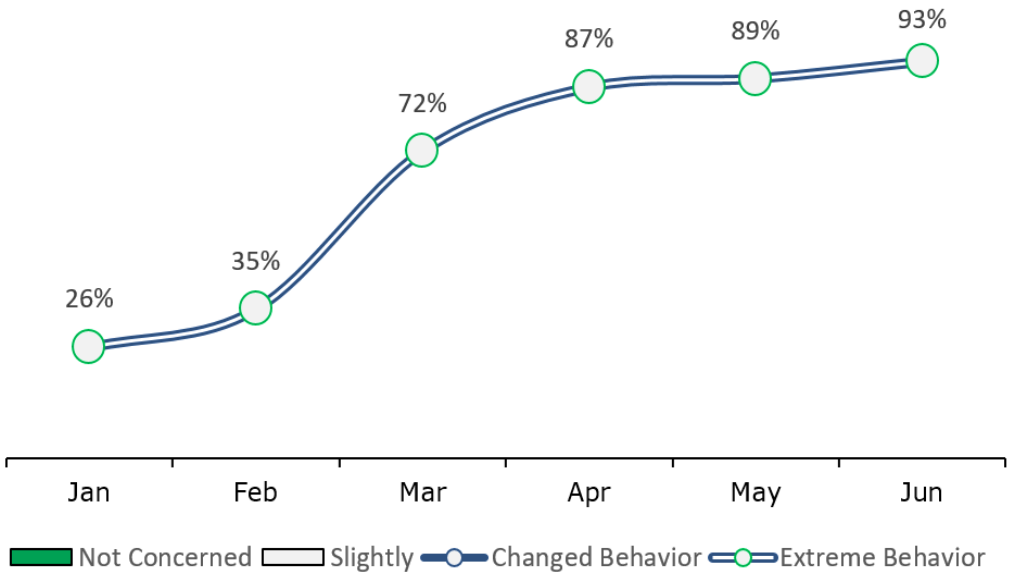

Figure 9. The line represents the extreme change in behavior of people during the Covid-19 in first quarter of 2020 using the RNN with NLP sentiment analysis

While our model shows good results for predicting the behavior of people and correctly used gendered experiments, accuracy scores for predicting the behavior were very high. The low loss achieved by our RNN model with NLP sentiment analysis compared to the high accuracy does show, however, that while the RNN model was good at predicting and correcting many different types of people behavior, it performed much better at predicting every single covid-19 based behavior in a sentence. Another consideration is that while sentences with either covid-19 or reported case generally many parameters, a sentence with covid-19 could have as many as one per parameters, increasing accuracy while loss stayed low.

4. Discussion

While our findings are very preliminary, and more work and time must be spent on improved results, our findings illustrate both some of the problems and possible solutions facing AI-powered natural language processing in predicting the change in behavior of common people infected with Covid-19 in first quarter of 2020. While our results show that today's datasets are large and complete enough to train a model, the improved accuracy of our RNN based NLP sentiment analysis model is indicative of Covid-19 problem. Ideally, as we have seen, we would start with a dataset of real sentences with a variety of grammatical problems, such as common people behavior, as well as the corresponding Covid-19 cases by a human. As this dataset exist, another alternative would be to programmatically create such a dataset such as we have done, but based on real, common people behavior and increasing cases in the world, rather than emulating cases. Our results show a wide gap in our RNN based NLP sentiment analysis-based model between the accuracy levels for covid-19 correction. While our model results are very accurate, we mention several reasons for this in the preceding methodology that is related to the dataset cleaning and feature extraction. The more interesting question is rather how closely our model mimics real human errors. Better datasets would not only result in a better-trained model, but also in a model that is more adept at correcting real errors that humans make, as opposed to random noise. The model had very high accuracy when predicting the change in behavior and correctly-used NLP techniques for sentiment analysis, which is very promising.

Table 3. Comparison of proposed work with the existing literature

ARTICLE	TECHNIQUE	ACCURACY
[29]	Support Vector Machine	46.00%
[30]	Convolutional Neural Network (CNN)	75.24%
Proposed	Recurrent Neural Network (RNN)	76.71%

5. Conclussion

The nature of our work prevented us from spending an excessive amount of time on any single training session, instead having to try a number of different setups and ways to preprocess data with 300 epochs of training. In this research work, with more time, we would like to continue from the results we have achieved by training the RNN with NLP based sentiment analysis model for more extended periods of time for predicting the behavior of people during Covid-19 pandemic with 76.71% of accuracy which is high as compared to the CNN, such as days or weeks, in order to see how results can improve. This would also necessitate using larger datasets, as the datasets that we used were very good and new for training as it finished in order to avoid overfitting. Our initial problem formulation involved investigating the ideal conditions and preprocessing for working with a specific NLP task: predicting the behavior during the specific time of May 20 – June 20 in 2020. Although the people suffered from the lack of Covid-19, the fact that we were able to overcome this problem and still achieve high accuracy levels for the two dataset that were acquired from an open source repository lead us to conclude that we have achieved overall success. We were able to highlight the specific problems facing the field of DL-powered NLP in predicting the people behavior, while showing that a RNN based NLP model can easily be trained to recognize and correct at least a subset of people unexpected side effects during the Covid-19 pandemic.

References

[1] D. Demner-Fushman, N. Elhadad. Aspiring to unintended consequences of natural language processing: A review of recent developments in clinical and consumer-generated text processing. Yearb. Med. Inform. 25, 224–233,2016.

[2] J. Zhang, K. Kowsari, J.H. Harrison, J.M. Lobo, L.E. Barnes. Patient2vec: A personalized interpretable deep representation of the longitudinal electronic health record. IEEE Access. 6, 65333–65346,2018.

[3] T. Hao, X. Chen, G. Li, J. Yan. A bibliometric analysis of text mining in medical research. Soft Comput. 22, 7875–7892,2018.

[4] X. Chen, Z. Liu, L. Wei, J. Yan, T. Hao, R. Ding. A comparative quantitative study of utilizing artificial intelligence on electronic health records in the USA and China during 2008–2017. BMC Med. Inform. Decis. Mak. 18, 117,2018.

[5] X. Chen, S. Wang, Y. Tang, T. Hao. A bibliometric analysis of event detection in social media. Online Inf. Rev. 43, 29–52,2019.

[6] M. R. Cowie, J.I. Blomster, L. H. Curtis, S. Duclaux, I. Ford, F. Fritz, S. Goldman, S. Janmohamed, J. Kreuzer, Leenay, M. Electronic health records to facilitate clinical research. Clin. Res. Cardiol. 106, 1–9,2017.

[7] H.T. Huhdanpaa, W.K. Tan, S. D. Rundell, P. Suri, F.H. Chokshi, B.A. Comstock, P.J. Heagerty, K.T. James, A.L. Avins, S.S. Nedeljkovic. Using Natural Language Processing of Free-Text Radiology Reports to Identify Type 1 Modic Endplate Changes. J. Digit. Imaging. 31, 84–90,2018.

[8] S. Sheikhalishahi, R. Miotto, J.T. Dudley, A. Lavelli, F. Rinaldi, V. Osmani. Natural language processing of clinical notes on chronic diseases: Systematic review. JMIR Med. Inform.7, e12239,2019.

[9] C. Dreisbach, T. A. Koleck, P.E. Bourne, S. A. Bakken. A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data. Int. J. Med. Inform. 125, 37–46,2019.

[10] G. Gonzalez-Hernandez, A. Sarker, K. O'Connor, G. Savova. Capturing the patient's perspective: A review of advances in natural language processing of health-related text. Yearb. Med. Inform. 26, 214–227,2017.

[11] Y. Luo, W.K. Thompson, T. M. Herr, Z. Zeng, M.A. Berendsen, S. R. Jonnalagadda, M.B. Carson, J. Starren. Natural language processing for EHR-based pharmacovigilance: A structured review. Drug Saf. 40, 1075–1089,2017.

[12] Available Online: https://devopedia.org/natural-language-processing

[13] T. A. Koleck, C. Dreisbach, P.E. Bourne, S. Bakken. Natural language processing of symptoms documented in free-text narratives of electronic health records: A systematic review. J. Am. Med. Inform. Assoc. 26, 364–379,2019.

[14] S. Velupillai, D. Mowery, B. R. South, M. Kvist, H. Dalianis. Recent advances in clinical natural language processing in support of semantic analysis. Yearb. Med. Inform. 24, 183–193,2015.

[15] M. Gutiérrez-Salcedo, M.A. Martínez, J. Moral-Munoz, E. Herrera-Viedma, M.J. Cobo. Some bibliometric procedures for analyzing and evaluating research fields. Appl. Intell.48, 1275–1287,2018.

[16] M. De la Flor-Martínez, P. Galindo-Moreno, E. Sánchez-Fernández, A. Piattelli, M.J. Cobo, E. Herrera-Viedma. H-classic: A new method to identify classic articles in Implant Dentistry, Periodontics, and Oral Surgery. Clin. Oral Implant. Res.27, 1317–1330,2016.

[17] C. De Maio, G. Fenza, V. Loia, M. Parente. Biomedical data integration and ontology-driven multi-facets visualization. In Proceedings of the 2015 International Joint Conference on Neural Networks (IJCNN); Killarney, Ireland, 12–16 July; pp. 1–8,2015.

[18] N.S. Houari, N. Taghezout. Integrating Agents into a Collaborative Knowledge-based System for Business Rules Consistency Management. Int. J. Interact. Multimed. Artif. Intell. 4, 61–72,2016.

[19] N. Settouti, M.E.A. Bechar, Chikh, M.A. Statistical comparisons of the top 10 algorithms in data mining for classification task. Int. J. Interact. Multimed. Artif. Intell. 2016, 4, 46–51,2016.

[20] M.J. Cobo, M. A. Martínez, M. Gutiérrez-Salcedo, H. Fujita, E. Herrera-Viedma. 25years at Knowledge-Based Systems: A bibliometric analysis. Knowl. Based Syst.80, 3–13,2015.

[21] C.A. Lester, M. Wang, V.V. Vydiswaran. Describing the patient experience from Yelp reviews of community pharmacies. J. Am. Pharm. Assoc.59, 349–355,2019.

[22] X. Chen, D. Zou, G. Cheng, H. Xie. Detecting latent topics and trends in educational technologies over four decades using structural topic modeling: A retrospective of all volumes of computer & education. Comput. Educ. 2020.

[23] X. Chen, D. Zou, H. Xie. Fifty years of British Journal of Educational Technology: A topic modeling based bibliometric perspective. Br. J. Educ. Technol. 2020.

[24] X. Chen, G. Yu, G. Cheng, T. Hao. Research topics, author profiles, and collaboration networks in the top-ranked journal on educational technology over the past 40 years: A bibliometric analysis. J. Comput. Educ.6, 563–585,2019.

[25] Available Online: https://www.pinterest.co.uk/pin/387802217894345149/

[26] H. Xie, H.-C. Chu, G.J. Hwang, C.-C. Wang. Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017. Comput. Educ.140, 103599,2019.

[27] Y. Song, X. Chen, T. Hao, Z. Liu, Z. Lan. Exploring two decades of research on classroom dialogue by using bibliometric analysis. Comput. Educ.137, 12–31,2019.

[28] X. Chen, H. Xie, F.L. Wang, Z. Liu, J. Xu, T. Hao. A bibliometric analysis of natural language processing in medical research. BMC Med. Inform. Decis. Mak. 2018, 18, 14,2018.

[29] S. Rashid, A. Ahmed, I. Al Barazanchi, A. Mhana, and H. Rasheed, "Lung cancer classification using data mining and supervised learning algorithms on multi-dimensional data set," *Period. Eng. Nat. Sci.*, vol. 7, no. 2, pp. 438–447, 2019.

[30] B. Alireza, M. Martin, S. David, M. Henrik, M. Orhun, V. Kartik. Natural Language Processing with Deep Learning for Medical Adverse Event Detection from Free-Text Medical Narratives: A Case Study of Detecting Total Hip Replacement Dislocation, 2020.

[31] I. Al Barazanchi, S. A. Hamid, R. A. Abdulrahman, and H. Rasheed, "Automated telemedicine and diagnosis system (ATDS) in diagnosing ailments and prescribing drugs," *Period. Eng. Nat. Sci.*, vol. 7, no. 2, pp. 888–894, 2019.

[32] I. Al Barazanchi, H. R. Abdulshaheed, S. A. Shawkat, and S. R. Binti, "Identification key scheme to enhance network performance in wireless body area network," *Period. Eng. Nat. Sci.*, vol. 7, no. 2, pp. 895–906, 2019.