EMOTION PREDICTIONS OF SENTIMENT DURING COVID PANDEMIC USING INTELLIGENT CHATBOTS

Venkata Duvvuri, Chetan Kulkarni, Sritha Gogineni

ABSTRACT

COVID-19 pandemic has created a major impact around the world. Governments and businesses small or big around the world are facing unprecedent decisions to either close up or reopen or drive other policies based on the sentiment of people. While, understanding this sentiment and accompanying emotions has been researched especially in social media channels like Twitter, we propose a novel way to capture sentiment and emotions using intelligent chatbots (EmoBot) that reduces the participants biases inherent in prior analysis. We devise Emotion Extraction Layers (EEL) based on latest deep learning techniques like BERT (Bidirectional Encoder Representations from Transformers) and compare these models with traditional machine learning models. We show for a variety of emotions that the new deep learning models predict 1-5% (Sad, Fearful & Angry) better than traditional machine learning techniques. Further, we showcase that leveraging retail sentiment data using transfer learning techniques can help cross the cold start chasm of having no chatbot data initially, and this technique achieves -8% closer in performance when compared to having enough COVID sentiment data.

KEYWORDS

COVID, sentiment analysis, chatbots, BERT, deep learning, transfer learning

1. INTRODUCTION

2020 has been a tumultuous year for several countries with the global outbreak of COVID-19 pandemic. As of early October, World health organization (WHO) reports that 34M+ has been infected with COVID-19 globally and 1M+ have died. During the early phases of the pandemic public sentiment was quite negative with skepticism. There were widespread public closures and lockdowns in various cities and counties as precautionary measures to help counter spread of the pandemic. However, as the economic toll started to threaten a pronounced long-term effect, these cities and counties started to consider various re-opening strategies as the pandemic progressed in late 2020. While, pandemic statistics serve as a guideline for such re-opening, these policies could have been better driven by an in-depth and accurate understanding of public sentiment on COVID-19. People in various countries took to various channels such as microblogging sites like twitter, reddit etc. to express themselves. There are several studies commenced to understand the sentiments in these channels [1], [2], but there were little efforts to substantiate this sentiment with other possible methods like chatbots [3], [4], [5] to collect pandemic data. Devising a data collection method and building a sentiment analysis on such novel data could potentially validate or repudiate the existing understanding of pandemic sentiment from such channels like Twitter etc. This is especially important where several billions of dollars are at stake during decisions made using such sentiments to drive public policies for schools, parks, stores alike and for businesses considering reopening strategies.

David C. Wyld et al. (Eds): SEC, DMA, AIAPP, CYBI, CoSIT, SIGL, NLPML – 2025
pp. 15-26, 2025. CS & IT - CSCP 2025 DOI: 10.5121/csit.2025.150802

In these moments of darkness, a little confirmation over a different method for sentiment analysis can go a long way. Secondly, accurate understanding and predictions of the individual sentiment during this pandemic can help businesses evaluate customer behaviors and strategize their products towards customer needs. On the other hand, due to the universal nature of the pandemic it is not possible to rely solely on traditional secondary research methods like tele-calling and/or online surveys. The purpose of this research is to elicit a novel method of data collection of COVID pandemic and understand sentiment and predicting corresponding emotions. Besides businesses, especially B2C (Business to Customers), and governments alike are equally impacted by COVID pandemic. If they knew how to measure the individual's or collective sentiment, then they can tailor their offerings or solutions to meet the current needs. Additionally, the sentiment needs to be detailed enough to help them make decisions. As such laying out the various emotions becomes important. The sentiment needs to be broken down along many such dimensions to help businesses figure out their strategies.

The main contributions of the study are in answering following questions:

- How can public institutions and private business alike collect customer sentiment during a pandemic in a different way other than what's expressed by customer via social media channels?
- What constitutes sentiment of an individual which can be identified by taking into account a range of emotions that are possible?
- How do latest sentiment models stack against traditional ones for COVID emotion's predictions?
- How can one leverage existing sentiment data captured in other ways like retail product reviews etc. to improve and jumpstart predictions?

The paper is organized as follows – Section 1 introduces the problem, Section 2 discusses the related work, Section 3 develops an intuition over the various machine learning methods used in the paper, Section 4 covers the experimental setup and data collection, Section 5 discusses the results and finally Section 6 ends with a conclusion.

2. RELATED WORK

During the COVID-19 pandemic people have been expressing their sentiments through various channels, most notably the microblogging channel twitter. While, initial descriptive statistical studies like CoronaTracker [6] analyze the outbreaks, these mostly help with understanding the disease spread. Recent studies have involved understanding individual's sentiment using social media and microblogging web data [1], [2], [7] & [8]. Twitter, a digital channel, has been in the forefront for enabling individual to express sentiment during various previous outbreaks and calamities like Typhoon [9], Ebola [10], Floods [11], & Zika [12]. Most of the studies focus on positive and negative sentiment, but few delve into the variety of emotions experienced during these times. Moreover, fewer delver into detailed emotions as defined by Plutchik's eight basic emotions model [13]. Dubey in [3] utilizes the NRC Word-Emotion Association Lexicon to distinguish 2 primary sentiments - positive and negative tweets and also categorizes them into 8 additional emotions of fear, joy, anticipation, anger, disgust, sadness, surprise and trust. Most of the research methods deal with self-expressed (push model) opinions, and it may introduce bias when certain expressive individuals participate more than the general population. This introduces bias in such analysis.

Chatbots are widely used as an agent to assist in customer service, bookings, shopping assistance etc. Their use for emotional analysis is relatively new [14], [15] and [5]. We device a primary research mechanism for data collection using conversational agents or chatbots like [5], [3], [4] to collect sentiment data from diverse individuals. The individual's motivation to give information (pull model) is different from the push model laid out earlier. It is triggered by individual's expert judgement and opinions and this leads to openness while answering questions which is different from social motives with friends and family. This mechanism is also more structured than social and web channel's data collection and allows room for flexibility in expressiveness. We leverage such textual conversational agent or Chatbots that are driven by artificial intelligence can conduct a text-based communication with humans by exploiting several natural language processing techniques. As such our study contrasts from previous twitter and web studies for COVID-19 sentiment and the more recent studies using chatbots [5], specifically in the deep learning models we build and compare. All together we believe our research methods introduce less participant bias. More recent advances in building a context-aware chatbot [16] and injecting personality into the machine [17] are suggested for future work in this study. Our work focuses on simpler chatbots with no context awareness for now.

Secondly previous studies have been geared towards only an analysis of the sentiment [2] rather than building a predictive model for such sentiments in real time. Several studies are available in literature to study sentiments over opinions and feelings about products, services and company strategies [18], [19], [20], [21], [22] etc. A few focus on the predictions of sentiment or emotions and specifically in applying them in health care [14] & [15] and/or government reopening policies [23]. Previous studies in healthcare mostly leverage existing records like clinical discharge summaries [24] to analyze diseases. We focus on experimenting with the data generated by intelligent chatbots and contrast four different predictive models, some of them leveraging latest deep learning techniques called Bidirectional Encoder Representations from Transformers - BERT [25]. Finally, we study how learnings from retail product's sentiment carry-over in a COVID pandemic sentiment to help improve new pandemic sentiment predictions not undertaken before.

3. MACHINE LEARNING MODEL

Our approach is to leverage state of art language model - BERT (Bidirectional Encoder Representations from Transformers) [25] which is a recent language representation model that involves training deep bidirectional representation for textual languages. Leveraging pretrained embeddings or representations has been a hallmark achievement of several models like Elmo [26], [27] & OpenAI GPT [28]. In our approach we leverage pre-trained BERT model as it has been shown to outperform Elmo and GPT models in [26]. We condition the last output layer based on the labelled sentiment data during re-training (a.k.a fine tuning). Thus, we create a state of art language model for sentiment analysis without substantial modification of the inherent task-specific architecture and by fine tuning all previous pretrained parameters. We leverage BERT's inherent advantage [25] of avoiding heavily engineered task specific architectures. Since our architecture is identical to the original ideas proposed in [25], we omit a detailed description of the BERT transformer in detail, In summary our BERT deep learning approach leverages pre-training on unlabeled data over several different pre-training tasks [25] apriori by previous authors. We fine tune this model by first initializing it with prior pre-trained parameters, and then fine tune all parameters using labeled data from our collected or historical sentiment dataset. Specifically, we fine-tune with three datasets to arrive at three models and compare and contrast them:

a. Only COVID Chatbot (*EmoBot*) dataset (697 sentences)
b. Only retail product sentiment dataset of key amazon products from Kaggle (45,128 sentences)
c. Combo of COVID chatbot and retail product sentiment datasets (45,825 sentences)

We train models with FastAI [29] which offers a high-level API and ready-to-use functions to train models in various applications. FastAI helps one to quickly apply pre-existing deep learning methods. Additionally, the main steps of our approach are:

I. Use BERT's Tokenizer & Vocab
II. Use Bert-Base pretrained model
III. Fine tune with FastAI APIs using a single learner class that integrates architecture, optimizer, and data automatically after choosing an appropriate loss function.
IV. Leverage discriminative learning [30] for different levels of learning rates and weight decays in different parts of the model architecture to achieve split learning of the model

4. DATA COLLECTION

COVID sentiment is widely collected and analyzed in social media microblogging web channels like twitter and reddit etc. These channels may suffer inherent selection bias due to differing participant propensities to voice their opinions in such a channel where privacy is a key concern. The purpose of this research is to validate and substantiate these findings using a novel data collection method using intelligent chatbots. This channel is presumably less susceptible to participant selection bias and additionally privacy is maintained where one can chat with an intelligent agent anywhere. As such the data collection differs in following dimensions a) Unstructured b) Unbiased c) Closer to the truth.

4.1. Methodology

The method involves launching an intelligent chatbot *Emobot* that can chat with a diverse group of individuals. The chatbot dynamically responds to the participant by conversing with them in a free-flowing Question and Answer formats listening to their concerns. This is pre-programmed intelligent agent to engage and gauge COVID perception. *EmotBot* is a simple chatbot that anyone can interact with to share their feelings and concerns about their current and future perspectives. All responses are analyzed by AI models. Using a shared Dashboard of the Analysis and predictions, the results from around the world are open for everyone to see. This sharing will give strength to people to know that they are not alone and they can get through this challenge together. There is no cost to interact with the *EmotBot* or to access the results.

4.2. Sampling

Following sampling technique has been used with other precautions to allow maximum accuracy and minimum bias:

- Non-probabilistic sampling method to allow for more response rates and enable comparisons with social media channels.
- Limited marketing relying on organic spread of the *EmoBot*, except initially for testing purposes, when we encouraged targeted individuals to engage with the Chatbot. While we undertook a few steps to encourage participation via our own social media handles, we did not allocate a significant marketing budget to buy participation.
- Ensured that we do not ask any Personally Identifiable Information, as that could have inserted bias into peoples' responses.
- We also did not put any limit in terms of the amount of text people can and must write. We wanted to allow users to feel at total liberty to share as much or as little as they wanted to share. We accepted a single word as well as large paragraphs.
- We also ensured that aggregate outcomes returned are allowed for anyone to access it free

of cost via a shared dashboard. We believe, this added a layer of transparency to our effort.

- Finally, we kept the duration of engagement very brief. We were conscious of the respondent fatigue, which could inject bias into their inputs.

By making it completely anonymous, voluntary, transparent, flexible size and brief, we believe, we were able to achieve our sampling goals.

5. EXPERIMENTATION

5.1. Experimental Design

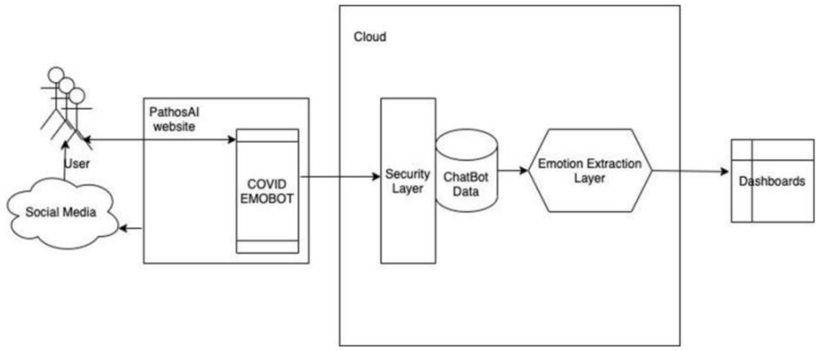

Figure 1. *Emobot* experimental design

Figure 1 shows the experimental design of the COVID emotion sentiment analysis. The COVID *EmoBot* is embedded into a website. The links to the bot is propagated through various social media channels (e.g. LinkedIn) in various diverse demographic groups. The participants are encouraged to express emotions and sentiments on COVID pandemic with the *EmoBot*. The data collected from the chatbot is stored securely in cloud behind a secure layer. The *Emobot* data is analyzed by an Emotional Extraction layer (EEL). The EEL is an asynchronous process that stores experimental models such as BERT, Random Forest (RF), Logistic Regression (LR) and Naïve Bayes (NB). The EEL results can be exported to be shown by visual analytics tools to help drive adoptions of the insights by businesses, people, government etc.

5.2. Data Processing

The *Emobot* converses with the users at times convenient to them and in places suitable to them. It guarantees privacy and security. We remove participant bias due to the fact that the collection mechanism is more of a free-flowing unstructured Q & A rather than a rigid inflexible interview or survey. Additionally, the channel of communication is not limited to particular social media outlets like twitter and thus avoids participant bias. Moreover, we have not limited to a particular social media channel for propagating the *EmoBot*. The *Emobot* is intelligent and handles the following example questions amongst many:

- How does the current situation make you feel?
- What are your biggest concerns right now?
- What are your views about the future?
- People are talking about making changes in their lives after the pandemic. What’s the one big change you will make in your lifestyle?
- There is a lot of discussion about certain services and industries changing forever. What is your view?

- Tell us which city you live in?
- Where do you work?
- When do you expect things to return to normal?

We experiment with two datasets a) *Emobot* COVID dataset b) Retail product sentiment dataset. The retail dataset is a list of over 34,000 consumer reviews for Amazon products like the Kindle, Fire TV Stick, and more from Datafiniti's Product Database in Kaggle's website [31] updated between September 2017 to October 2018 & February 2019 to April 2019.The collected dataset is free flowing unstructured text and we employ crowd sourcing to label the datasets to enable building supervised learning methods in EEL. A team of 20+ Northeastern students from diverse backgrounds were selected to review the dataset and rank the variety of emotions in each dataset in a scale of 0-10. The emotions considered are 4 positive ones (Surprised, Happy, Excited, Agreeing) and 4 negative ones (Sad, Frustrated, Fearful, Angry).

5.3. Data Analysis

Table 1 shows the size of the datasets and data collection times. The *Emobot* dataset has collected substantial data records greater than often suggested minimum of 30 needed per variable. Note, here the Retail Product review sentiment data is used to train a transfer learning model that is tested on Emobot data for accuracy in Section 5.4.3

Table 1: Datasets

Dataset	**Totals**	**Collection period**
Emobot	697	06/20/2020 – 07/20/2020
Retail Product Review	45,128	09/2017-10/2018; 02/2019 -04/2019
Emobot + Retail Product Review	45,825	Combo

Figure 2 & 3 shows the box plot distributions of retail product sentiment data and *Emobot* COVID data. It is clear that the interquartile range (IQR) for product sentiments are mostly positive and with sentiments of surprised, happy and excited all being 5+ in IQRs. The sentiment for COVID data is in contrast is more negative with sad, frustrated and fearful having more pronounced IQRs. This is also illustrated with an average of Likert like scaled (0-10) sentiment plotted for all emotions expressed in Figure 4 & 5.

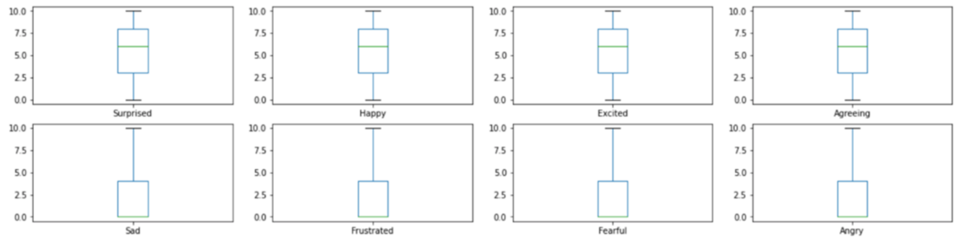

Figure 2. Product sentiment emotion boxplot distributions

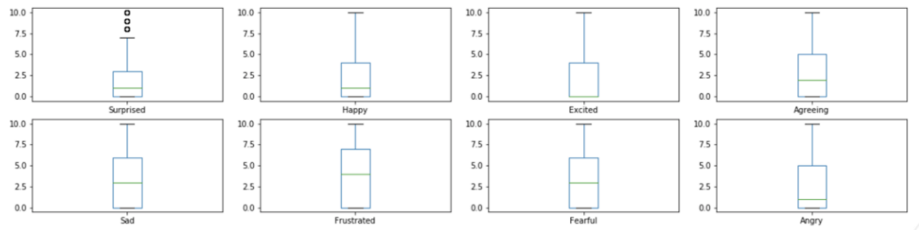

Figure 3. Emobot data emotion boxplot distributions

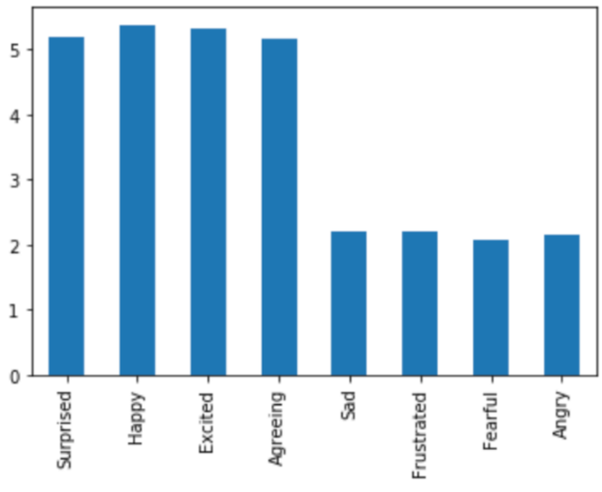

Figure 4. Average Product review sentiment

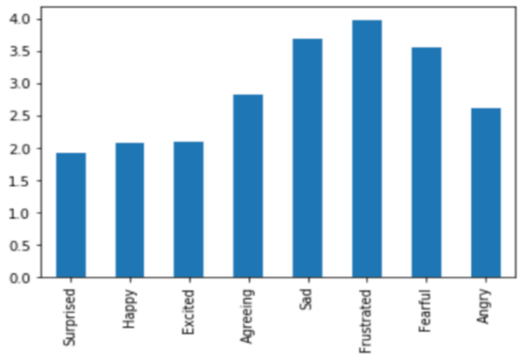

Figure 5. Average COVID *Emobot* sentiment

Finally, the analysis of the unstructured text expressed in *Emobot* transformed as a word cloud representation shows the similar insights that products labels like *tablet* and *batteries* along with positive emotions of *love* and *good* make it to the top for product data in Figure 6, while *uncertainty*, *anxious* and *future* bubble to the top sentiment in *Emobot* data as in Figure 7.

Figure 6. Product sentiment word cloud distributions

Figure 7. Emobot sentiment word cloud distribution

5.4. Results

5.4.1. Theoretical framework

We observe that initially the COVID pandemic cannot be predicted reliably with a chatbot as there might be no data collected initially. Therefore, we present and analyze a theoretical framework as in Figure 8 to collect retail product review sentiment data to an inferred (transfer learned) EEL models. We use transfer learning over this prior built historical model to start predicting sentiments on COVID emotions right from DAY 1 and thus address cold start problems. When enough data is available, we mix the *Emobot* COVID data with historical retail product data to asymptotically approach real COVID sentiment and emotions.

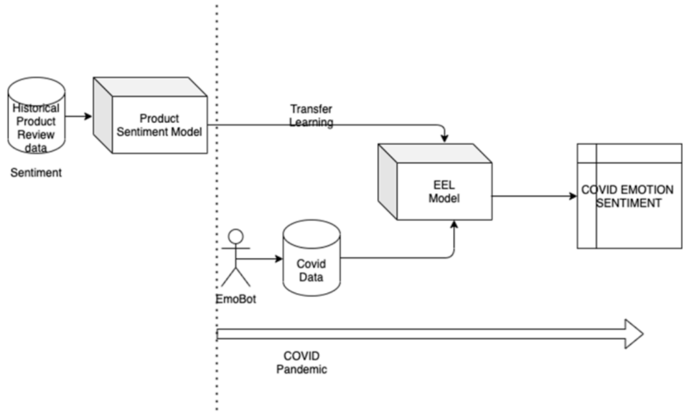

Figure 8. Transfer learning model for COVID pandemic

5.4.2. EEL Models Results

We split the COVID data into train, validation and test datasets. We shave manually labelled each dataset as per Section 5.1. The validation set is used to tune the model parameters and the testset is used to find accuracy of each emotion over unseen data. To allow for comparisons we have used the same test set of 219 records in all comparisons. We have used >= 7 in Likert scale of 0-10 as a threshold for identifying the strength of each emotion to be fed as a dichotomous classified label target variable in each EEL method. With the *Emobot* COVID data we compare following EEL models: Bidirectional Encoder Representations from Transformers (BERT), Random Forest (RF), Naïve Bayes (NB) and Logistic Regression (LR) with code in https://github.com/dvshekar/covid_sentiment_emobot_transfer_learning. Table 2 shows this scenario where we have enough *EmoBot* data collected after a few days or weeks of pandemic.

Table 2. Emobot only EEL model's accuracy comparison

	LR %	NB %	RF %	BERT %
SURPRISED	88.97	86.03	**91.91**	90.81
HAPPY	86.03	80.51	**87.87**	87.50
EXCITED	87.87	82.72	84.19	**88.24**
AGREEING	76.47	71.69	76.84	**77.94**
SAD	73.16	67.65	72.79	**75.37**
FRUSTRATED	70.59	58.09	**72.06**	67.65
FEARFUL	74.26	66.18	73.53	**74.63**
ANGRY	79.78	72.06	76.47	**81.25**

We find that for 5 out of 8 emotions the BERT model is the best and leads to an improvement of 1-5% in emotions predictions accuracy, especially over negative emotions (Sad, Frustrated & Angry).

5.4.3. Transfer Learning Results

As in Section 5.4.1 initially we may not have enough *Emobot* COVID data. In such scenarios we examine the effect of transfer learning (i.e. training with) from retail product reviews sentiment data onto the COVID emotions with the same train, validation and testsets. Thus, in the experiment we evaluate transition period's problem when we may have some but not sufficient *Emobot* data. We device and evaluate a scheme of mixing emotions from Retail Product sentiment data and/or *EmoBot* data and using the BERT model for predicting each emotion.

Table 3. BERT model's transfer learning accuracy comparisons

	COVID *Emobot* only %	Retail Product only %	Retail Product + *Emobot* Combo %
SURPRISED	90.81	80.51	86.76
HAPPY	87.50	58.09	83.82
EXCITED	88.24	70.22	84.19
AGREEING	77.94	70.96	81.25
SAD	75.37	74.63	71.69
FRUSTRATED	67.65	71.69	66.54
FEARFUL	74.63	74.63	68.75
ANGRY	81.25	81.62	75.37
		-7.63%	**-3.13%**

We notice that when we train with retail product review sentiment data and transfer it to *Emobot* COVID testset, we can predict COVID sentiment to within -8% accuracy in aggregate over all emotions. COVID *Emobot* data pretty much serves as the upper bound on such transfer learning. Secondly, we notice that negative emotions (Sad, Frustrated, Fearful and Angry) prediction's accuracy trained over such data is very close to their asymptotic limits of pure *Emobot* data, indicating that these negative emotions can be predicted very well by transfer learning (training with) from Product Retail Sentiment data. Secondly, we notice that accuracy for positive emotions (Surprised, Happy, Excited, Agreeing) predicted with Retail Product data can be substantially improved by adding *Emobot* data (combo) into the learning (training) process over a period to approach these emotion's asymptotic limits.

6. CONCLUSION

COVID pandemic has wreaked havoc all over the world. Businesses and governments alike would like to understand people's sentiment measured across various emotions to craft policies, strategies and re-opening plans. This was also important at the initial stages of the pandemic when limited pandemic sentiment data is available. We provide an accurate and bootstrapped way to leverage transfer learning from retail product sentiment data to predict COVID sentiment in such scenarios. We show that the latest deep learning models (BERT) are the best performing ones compared to traditional machine learning models in 5 out of 8 COVID emotions. Additionally, we show that the resultant BERT model leveraging retail product review data is about -8% less accurate than *Emobot* COVID only pandemic sentiment data. The negative emotions are closer in these asymptotic limits for retail product review data. Moreover, as the pandemic progresses, we can collect more sentiment data, and use this in combo with product review data to narrow the accuracy to being -3% less accurate in aggregate over all emotions. Thus, we propose ways to measure pandemic sentiment in terms of negative and positive emotions right from start to finish in a global pandemic. Secondly, we device a new way to collect sentiment using intelligent chatbots that are less biased due to pull-based model of information gathering that empowers the participant to be more expressive and be at ease and share with a laser focus on pandemic.

REFERENCES

[1] Drias HH, Drias Y. Mining Twitter Data on COVID-19 for Sentiment analysis and frequent patterns Discovery. medRxiv; 2020. DOI: 10.1101/2020.05.08.20090464.

[2] Dubey, A. (2020). Twitter Sentiment Analysis during COVID19 Outbreak. SSRN Electronic Journal. doi: 10.2139/ssrn.3572023

[3] Hao Fang, Hao Cheng, Maarten Sap, Elizabeth Clark, Ari Holtzman, Yejin Choi, Noah A Smith, and Mari Ostendorf. 2018. Sounding Board: A User-Centric and Content-Driven Social Chatbot. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations. 96–100.

[4] Braden Hancock, Antoine Bordes, Pierre-Emmanuel Mazare, and Jason Weston. 2019. Learning from Dialogue after Deployment: Feed Yourself, Chatbot! arXiv preprint arXiv:1901.05415 (2019).

[5] Ouerhani, N., Maalel, A., Ghézala, H. B., & Chouri, S. (2020). Smart Ubiquitous Chatbot for COVID-19 Assistance with Deep learning Sentiment Analysis Model during and after quarantine.

[6] Binti Hamzah FA, Lau C, Nazri H, Ligot DV, Lee G, Tan CL, et al. CoronaTracker: World- wide COVID-19 Outbreak Data Analysis and Prediction. [Preprint]. Bull World Health Organ. E-pub: 19 March 2020. doi: http://dx.doi.org/10.2471/BLT.20.255695

[7] Li, X., Zhou, M., Wu, J., Yuan, A., Wu, F., & Li, J. (2020). Analyzing COVID-19 on Online Social Media: Trends, Sentiments and Emotions. arXiv preprint arXiv:2005.14464.

[8] Kleinberg, B., van der Vegt, I., & Mozes, M. (2020). Measuring emotions in the covid-19 real world worry dataset. arXiv preprint arXiv:2004.04225.

[9] Soriano, Cheryll Ruth, Ma Divina Gracia Roldan, Charibeth Cheng, and Nathaniel Oco. "Social media and civic engagement during calamities: the case of Twitter use during typhoon Yolanda." Philippine Political Science Journal 37, no. 1 (2016): 6-25.

[10] Van Lent, Liza GG, Hande Sungur, Florian A. Kunneman, Bob Van De Velde, and Enny Das. "Too far to care? Measuring public attention and fear for Ebola using Twitter." Journal of medical Internet research 19, no. 6 (2017): e193.

[11] Nair, Meera R., G. R. Ramya, and P. Bagavathi Sivakumar. "Usage and analysis of Twitter during 2015 Chennai flood towards disaster management." Procedia computer science 115 (2017): 350-358.

[12] Fu, King-Wa, Hai Liang, Nitin Saroha, Zion Tsz Ho Tse, Patrick Ip, and Isaac Chun-Hai Fung. "How people react to Zika virus outbreaks on Twitter? A computational content analysis." American journal of infection control 44, no. 12 (2016): 1700-1702.

[13] Plutchik R. The Nature of Emotions. American Scientist. 2001;89(4):344.

[14] Pranav Rane, Kashyap Bhansali and Sindhu Nair. Article: Sentiment Analysis to Improve Emotional Health of User. International Journal of Computer Applications 120(1):21-24, June 2015.

[15] Aditya Joshi, Anoop Kunchukuttan, Pushpak Bhattacharyya, Mark J Carman. SarcasmBot: An open-source sarcasm-generation module for chatbots, WISDOM at KDD 2015, Sydney, Australia, August 2015.

[16] Alessandro Sordoni, Michel Galley, Michael Auli, Chris Brockett, Yangfeng Ji, Margaret Mitchell, Jian-Yun Nie, Jianfeng Gao, and Bill Dolan. 2015. A Neural Network Approach to Context-Sensitive Generation of Conversational Responses. In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, Denver, Colorado, 196–205. https://doi.org/10.3115/v1/N15-1020

[17] Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. 2018. Personalizing Dialogue Agents: I have a dog, do you have pets too? In Proceedings of the 56th Annual Meeting of the Association

[18] Federico Alberto Pozzi, Elisabetta Fersini, Enza Messina, Bing Liu, Sentiment Analysis in Social Media, Morgan Kaufmann, 2016

[19] Kyle S. Van Houtan, Tyler Gagne, Clinton N. Jenkins, Lucas Joppa, Sentiment Analysis of Conservation Studies Captures Successes of Species Reintroductions, Patterns 1, 100005, April 10, 2020, https://doi.org/10.1016/j.patter.2020.100005

[20] Liu B. (2010). Sentiment analysis and subjectivity. Handbook of Natural Language Processing, 2nd ed.

[21] Neviarouskaya, A., Prendinger, H., and Ishizuka, M. (2011). Sentiful: A lexicon for sentiment analysis. IEEE Transactions on Affective Computing, 2:22–36.

[22] Thelwall, M., Buckley, K., and Paltoglou, G. (2012). Sentiment strength detection for the social web. JASIST, 63(1):163–173.

[23] Ahmed, M. E., Rabin, M. R. I., & Chowdhury, F. N. (2020). COVID-19: Social Media Sentiment Analysis on Reopening. arXiv preprint arXiv:2006.00804.

[24] Yang, H., Spasic, I., Keane, J. A., & Nenadic, G. (2009). A text mining approach to the prediction of disease status from clinical discharge summaries. Journal of the American Medical Informatics Association, 16(4), 596-600.

[25] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[26] Matthew Peters, Waleed Ammar, Chandra Bhagavatula, and Russell Power. 2017. Semi-supervised sequence tagging with bidirectional language models. In ACL.

[27] Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018a. Deep contextualized word representations. In NAACL.

[28] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding with unsupervised learning. Technical report, OpenAI.

[29] Howard, Jeremy, and Sylvain Gugger. "Fastai: A layered API for deep learning." Information 11.2 (2020): 108.

[30] Howard, J.; Ruder, S. Fine-tuned Language Models for Text Classification. arXiv 2018, arXiv:1801.06146.

[31] Datafiniti (2019). Consumer Reviews of Amazon Products. A list of over 34,000 reviews of Amazon products like the Kindle, Fire TV, etc. Retrieved from https://www.kaggle.com/datafiniti/consumer-reviews-of-amazon-products?select=Datafiniti_Amazon_Consumer_Reviews_of_Amazon_Products_May19.

AUTHORS

Venkata C. Duvvuri is a doctoral student at Department of Technology Leadership and Innovation at Purdue University. Additionally, he is an Architect level Consulting Member of Technical Staff - Data Scientist at Oracle corporation in Redwood City, CA, USA. He loves teaching and is an adjunct faculty at Northeastern University. He has held several leadership positions in data science at various companies. He holds a Master's degree in computer science from University of Massachusetts – Amherst and an MBA from University of California - Davis.

Chetan Kulkarni is working as an Informatics Data Analyst at Genentech, Inc. He holds a Master's degree in Analytics from Northeastern University and Bachelor's degree in Information technology from Visvesvaraya Technological University.

Sritha V Gogegineni received a master's in professional studies, Business Analytics from Northeastern University. She has done her bachelor's in economics from the Indian Institute of Technology, Kanpur. She is currently working as Sr. data analyst, experimentation in Realtor.com. Her interest is in working on developing Machine learning projects and work on solving some of the interesting and most important real-life problems of Global warming.